

# Statistical investigative questions

---

*An enquiry into posing and answering investigative questions  
from existing data.*

**Phillipa Mareea Arnold**

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy,  
The University of Auckland, 2013.*

## **Abstract**

*Statistical investigations is a major strand of the New Zealand statistics curriculum. A preliminary study identified the initial problematic situation that some year 10 students (ages 14–15) were unable to complete a statistical enquiry because their question, which the teacher had approved, was unsuitable for the multivariate data set they were investigating. An understanding about key concepts underpinning statistical questions seemed to be lacking. A subsequent review of the literature failed to find criteria for what makes a good statistical question. Hence the research topic for this thesis was established: what makes a good investigative question, and developing links between the investigative question and the analysis and conclusion.*

*Using a design research method, four teaching experiments were conducted over a period of five years. Each experiment involved one year 10 class and altogether 93 students and two teachers from two mid-decile multicultural schools participated. From the initial identified problematic situation, a teaching experiment was planned which involved identification of underpinning concepts and development of innovative prototypical instruction material; the teaching plan was then implemented. Data collected were student pre-and post-tests, interviews and class transcripts, which provided insights into student reasoning and which fed back into the next teaching experiment.*

*The main findings from the research were: (1) identification of the criteria for what makes a good question and for describing distributions; (2) explication of the conceptual infrastructure that students need for investigative questions, making a call, and distributions; and (3) promising indications that the implementation of the especially developed learning materials designed to build concepts such as variable, sample, population, sampling variability, and distribution, improved students' reasoning processes. Frameworks to describe concepts and to assess the level of student reasoning for investigative questions, making a call, and distributions were developed from the literature and data. The main themes that emerged from the research were the necessity for concept identification, concept development, particular learning and teaching approaches, and building knowledge about student learning, when attempting to engineer a new paradigm for enculturating students into new ways of thinking statistically. The implications of these findings are that teachers will need extended professional learning to meet the demands of the new curriculum.*

## Acknowledgements

This journey started in 1990 through a conversation with Andy Begg. Andy interviewed me regarding professional development as part of his data collection for his first doctoral thesis. Following the interview we talked for a further hour and a career path was planned for me: Master's then PhD. To Andy and to Sue Winters, my first head of department (mathematics), thank you both for seeing in a third-year teacher the potential to grow beyond her wildest dreams. From small acorns grow large oak trees.

Skip to 2005 and an invitation from Maxine Pfannkuch to attend the Fourth Statistical Reasoning, Thinking and Literacy Forum in Auckland, New Zealand. A good selection of who's who in statistics education was in attendance, although I was blissfully unaware of that at the time. The week of indulgence in the world of research and with those who were madly passionate about statistics education was the light needed to start the PhD fire. While it had been on the agenda, the will and the way had not yet been realised and it was this week, listening to and talking with Maxine, Chris Wild, Katie Makar, Dani Ben-Zvi, Bill Finzer, Jane Watson, Bob delMas, Andee Rubin and Chris Reading (I am sure I have missed people, my apologies) that inspired me to get started on my PhD in statistics education. The ongoing support from many of these wonderful people as I have travelled on my PhD journey has been nourishing and rewarding.

To Associate Professor Maxine Pfannkuch, no words can really express the deep respect and appreciation I have for and of you. The combination of working in teacher professional development, developing material for a new curriculum, supporting the development of new achievement standards and doing a PhD required careful negotiation of time, with the curriculum and assessment development usually taking priority. Maxine was always aware of the pressures on me during this time and was flexible and supportive in helping to adjust and align the work I was doing with the research. I have been blessed to have had the great fortune of Maxine as both my PhD supervisor and a good friend. Thanks also to Professor Chris Wild, my second supervisor, who provided clarity, drive and a second independent pair of eyes and ears when things were at a stalemate and in need of an injection of humour and direction. Chris's depth of knowledge, both from a statistics point of view as well as from a statistics education point of view, meant that I had the best of both worlds in terms of support from my supervision team. Thank you both for nurturing and challenging me over these last six-and-a-half years.

To the two teachers and their classes who let me into their space and tried everything with enthusiasm and blind faith, I cannot thank you enough – it was a privilege to be in your classrooms. And these two teachers continue to inspire me. To the Teaching and Learning Research Initiative project team – Maxine Pfannkuch, Chris Wild, Matt Regan, Sophie Wright, Jason Florence, Jeanette Saunders, Anna Martin, Marina McFarland, Jason Ellwood and Lindsay Smith – thank you all for your contribution to the project and the deep thinking we did to get where we did. The strength was in the collective; the sum was far greater than the parts.

I would like to thank both of my employers during this time for allowing me to work part time. I enjoyed the work I did too much to give it up completely to do my PhD full time, but I knew I could not do justice to a PhD working full time. To Team Solutions and especially to Cognition Education, thank you for supporting me in this endeavour. To Terry Bates, Phil Coogan and all of the team at Cognition Education, I have appreciated both your support and interest in my research and how it has progressed. Alongside this over the last two years has been the Te Toi Tupu mathematics team, an amazing group of educators who are passionate about mathematics and who supported me as I wrote up my research, not only by picking up various pieces of work when I was writing, but also for being strong and not sending me emails while I was writing – I have appreciated this more than you realise.

A huge thanks, also, to Liz Stone from Shepston Editing Services for her “final polish”.

Completing a PhD requires one to be quite selfish, and in order for one to be selfish, you need others who are prepared to support you. I have been truly fortunate in the team that have supported me on the home front. To Martine, thank you for just being there when I needed someone to talk to and understanding what I was talking about, and for walking the dog when I could not. Many thanks to Dad and Mary for the times I could escape to Napier to do some writing, and to Pat and Jim – without you two and the bolt hole in Waiomu, I don’t think I would have finished yet! And to all of my family and friends, it is not that I haven’t thought about you many times, I just have not had time to follow up on those thoughts; I do appreciate that when we do catch up that you will not hold this against me.

Lastly I want to thank Ian. Thanks, my best friend, for everything: for encouraging me to do my PhD, for loving and supporting me despite my unreasonableness at times, my being tired and grumpy, and my not coming to support your activities because I was writing, and for letting me escape to write and leaving you to fend (admirably) for yourself.

# Table of Contents

|                                                                 |             |
|-----------------------------------------------------------------|-------------|
| <b>Abstract.....</b>                                            | <b>ii</b>   |
| <b>Acknowledgements .....</b>                                   | <b>iii</b>  |
| <b>Table of Contents .....</b>                                  | <b>v</b>    |
| <b>List of Figures.....</b>                                     | <b>viii</b> |
| <b>List of Tables.....</b>                                      | <b>xi</b>   |
| <b>Chapter 1. Introduction.....</b>                             | <b>1</b>    |
| 1.1. Rationale for research .....                               | 3           |
| 1.2. Research questions.....                                    | 4           |
| 1.3. Scope of the research .....                                | 5           |
| 1.4. Outline of the thesis chapters .....                       | 6           |
| <b>Chapter 2. Literature Review .....</b>                       | <b>7</b>    |
| 2.1. Introduction.....                                          | 7           |
| 2.2. Assumptions.....                                           | 7           |
| 2.3. Statistical investigations.....                            | 9           |
| 2.4. Questions within the statistical investigation cycle ..... | 19          |
| 2.5. Question posing .....                                      | 22          |
| 2.6. Question asking.....                                       | 30          |
| 2.7. Investigative questions and nature of data .....           | 34          |
| 2.8. Summary .....                                              | 39          |
| <b>Chapter 3. Research Methods.....</b>                         | <b>41</b>   |
| 3.1. Introduction.....                                          | 41          |
| 3.2. Theoretical discussion.....                                | 41          |
| 3.3. Design research.....                                       | 43          |
| 3.4. Data collection methods.....                               | 52          |
| 3.5. Data analysis .....                                        | 60          |
| 3.6. Ethics .....                                               | 63          |
| 3.7. Triangulation.....                                         | 63          |
| 3.8. Validity and reliability .....                             | 64          |
| 3.9. Summary .....                                              | 65          |

|                                                                                                   |            |
|---------------------------------------------------------------------------------------------------|------------|
| <b>Chapter 4. The Teaching Experiments .....</b>                                                  | <b>66</b>  |
| 4.1. Introduction.....                                                                            | 66         |
| 4.2. Design research.....                                                                         | 66         |
| 4.3. Teaching experiment 1: Recognising the initial problematic situation.....                    | 71         |
| 4.4. Teaching experiment 2: What makes a good investigative question? .....                       | 73         |
| 4.5. Teaching experiment 3: Answering the investigative question – making the call .....          | 75         |
| 4.6. Teaching experiment 4: Answering the investigative question – describing distributions ..... | 78         |
| 4.7. Overview of the four teaching experiments .....                                              | 81         |
| 4.8. Summary .....                                                                                | 82         |
| <b>Chapter 5. Recognising the Initial Problematic Situation.....</b>                              | <b>83</b>  |
| 5.1. Introduction.....                                                                            | 83         |
| 5.2. Setting the scene .....                                                                      | 83         |
| 5.3. Student pre- and post-test analysis.....                                                     | 84         |
| 5.4. The problematic situation is realised .....                                                  | 85         |
| 5.5. Retrospective analysis of investigative questions .....                                      | 87         |
| 5.6. Discussion .....                                                                             | 93         |
| 5.7. Summary .....                                                                                | 95         |
| <b>Chapter 6. What Makes a Good Investigative Question .....</b>                                  | <b>96</b>  |
| 6.1. Introduction.....                                                                            | 96         |
| 6.2. Problematic situation .....                                                                  | 96         |
| 6.3. Planning and preparation .....                                                               | 97         |
| 6.4. Teaching experiments .....                                                                   | 102        |
| 6.5. Retrospective analysis.....                                                                  | 109        |
| 6.6. Quantitative analysis of test questions .....                                                | 125        |
| 6.7. Discussion .....                                                                             | 135        |
| 6.8. Practical implications of the research .....                                                 | 138        |
| <b>Chapter 7. Answering the Investigative Question: making the call .....</b>                     | <b>140</b> |
| 7.1. Introduction.....                                                                            | 140        |
| 7.2. Problematic situation .....                                                                  | 141        |
| 7.3. Literature review .....                                                                      | 144        |
| 7.4. Planning and preparation .....                                                               | 149        |
| 7.5. Teaching experiments .....                                                                   | 159        |
| 7.6. Retrospective analysis.....                                                                  | 165        |
| 7.7. Quantitative analysis of test questions .....                                                | 181        |
| 7.8. Discussion .....                                                                             | 187        |

|                                                                                                                                     |            |
|-------------------------------------------------------------------------------------------------------------------------------------|------------|
| 7.9. Practical implications of the research .....                                                                                   | 189        |
| <b>Chapter 8. Answering the Investigative Question: Describing Distributions.....</b>                                               | <b>190</b> |
| 8.1. Introduction.....                                                                                                              | 190        |
| 8.2. Problematic situation .....                                                                                                    | 190        |
| 8.3. Literature review .....                                                                                                        | 194        |
| 8.4. Planning and preparation .....                                                                                                 | 201        |
| 8.5. Teaching experiments .....                                                                                                     | 205        |
| 8.6. Retrospective Analysis.....                                                                                                    | 212        |
| 8.7. Quantitative analysis of test questions .....                                                                                  | 219        |
| 8.8. Discussion.....                                                                                                                | 229        |
| 8.9. Practical implications.....                                                                                                    | 232        |
| <b>Chapter 9. Theoretical Frameworks, Implications, and Where to from Here.....</b>                                                 | <b>233</b> |
| 9.1. Overarching themes and frameworks .....                                                                                        | 233        |
| 9.2. Implications for policy and practice.....                                                                                      | 248        |
| 9.3. Looking ahead – suggestions for further research .....                                                                         | 249        |
| 9.4. Limitations .....                                                                                                              | 251        |
| 9.5. Conclusion .....                                                                                                               | 251        |
| <b>Publications associated with this research.....</b>                                                                              | <b>253</b> |
| <b>Appendix A – Statistical investigations achievement objectives.....</b>                                                          | <b>254</b> |
| <b>Appendix B – Pre- and Post-Tests .....</b>                                                                                       | <b>256</b> |
| <b>Appendix C – Marking criteria for Teaching Experiment 1 pre- and post-tests.....</b>                                             | <b>302</b> |
| <b>Appendix D – Interview schedules for teaching experiments 3 and 4 – related interview<br/>questions to making the call .....</b> | <b>304</b> |
| <b>Appendix E – Investigative questions in assessment and curriculum materials .....</b>                                            | <b>310</b> |
| <b>Appendix F – Presentations on posing investigative questions.....</b>                                                            | <b>312</b> |
| <b>Appendix G – Teaching and Learning Research Initiative (TLRI) papers and presentations..</b>                                     | <b>314</b> |
| <b>Appendix H – Lessons on shape.....</b>                                                                                           | <b>319</b> |
| <b>References.....</b>                                                                                                              | <b>325</b> |

## List of Figures

|                                                                                                                         |         |
|-------------------------------------------------------------------------------------------------------------------------|---------|
| Figure 1-1. New Zealand secondary schooling years and associated curriculum and qualification levels.....               | 1       |
| Figure 1-2. The Statistical Enquiry Cycle.....                                                                          | 2       |
| Figure 2-1. Comparison between 1992 and 2007 level 5 (ages 13–15) statistical investigation achievement objectives..... | 11      |
| Figure 2-2. Comparison of PPDAC and PCAI models.....                                                                    | 17      |
| Figure 2-3. Question posing and question asking within the PPDAC cycle.....                                             | 20      |
| Figure 2-4. Questions within the statistical investigative cycle: primary data.....                                     | 21      |
| Figure 2-5. Questions within the statistical investigative cycle: secondary data .....                                  | 22      |
| Figure 2-6. (a) Flow of influence primary data (b) Flow of influence secondary data.....                                | 37      |
| Figure 2-7. Overall Structure Built During Teaching .....                                                               | 38      |
| Figure 3-1. Design research phases intertwined with hypothetical learning trajectory .....                              | 45      |
| Figure 3-2. Statistics teaching cycle .....                                                                             | 46      |
| Figure 3-3. AS91035: Investigate a given multivariate data set using the statistical enquiry cycle .....                | 62      |
| Figure 4-1. Overview of the four teaching experiments .....                                                             | 82      |
| Figure 5-1. Grade movement from pre- to post-test .....                                                                 | 85      |
| Figure 5-2. Examples of student-posed questions where the sample size was too small.....                                | 86      |
| Figure 5-3. Summary question categories.....                                                                            | 88      |
| Figure 5-4. Comparison question categories.....                                                                         | 89      |
| Figure 5-5. Category of question selected to complete the post-test .....                                               | 91      |
| Figure 6-1: Relationship question posed in class.....                                                                   | 102     |
| Figure 6-2. Examples of teacher modelled comparison investigative questions .....                                       | 109     |
| Figure 6-3. Criteria for posing investigative questions.....                                                            | 110–111 |
| Figure 6-4. Summary question categories.....                                                                            | 112     |
| Figure 6-5. Updated summary question categories.....                                                                    | 113     |
| Figure 6-6. Summary question examples.....                                                                              | 114–115 |
| Figure 6-7. Comparison question categories.....                                                                         | 117     |
| Figure 6-8. Updated comparison question categories .....                                                                | 118     |
| Figure 6-9. Comparison question examples.....                                                                           | 119–120 |
| Figure 6-10. Population descriptor categories .....                                                                     | 124     |
| Figure 6-11. (a) Summary investigative question matrix, (b) Comparison investigative question matrix .....              | 125     |
| Figure 6-12. Populations suggested to improve question 1 .....                                                          | 126     |
| Figure 6-13. Populations suggested to improve question 3 .....                                                          | 127     |



|                                                                                                                                                                                                                                                                        |         |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Figure 6-14. Decision regarding question 4 and supporting reasons .....                                                                                                                                                                                                | 128     |
| Figure 6-15. Pre-test mean score for comparison questions .....                                                                                                                                                                                                        | 129     |
| Figure 6-16. SOLO criteria for grading comparison investigative questions .....                                                                                                                                                                                        | 130     |
| Figure 6-17. (a) Scatterplot and (b) table of 2008 student pre- and post-test mean scores for<br>comparison investigative questions.....                                                                                                                               | 131     |
| Figure 6-18. Graph of difference between post-test mean score and pre-test mean score.....                                                                                                                                                                             | 131     |
| Figure 6-19. (a) Scatterplot and (b) table of 2011 student pre- and post-test mean scores for<br>comparison investigative questions.....                                                                                                                               | 132     |
| Figure 6-20. Graph of difference between post-test mean score and pre-test mean score.....                                                                                                                                                                             | 132     |
| Figure 6-21. 2008 question type for all post-test comparison investigative questions .....                                                                                                                                                                             | 133     |
| Figure 6-22. 2011 question type for all post-test comparison questions.....                                                                                                                                                                                            | 134     |
| Figure 6-23. Population categories for all post-test comparison questions: (a) 2008, and (b) 2011...                                                                                                                                                                   | 134     |
| Figure 6-24. Link between investigative questions, making the call and describing distributions.....                                                                                                                                                                   | 138     |
| Figure 7-1. (a) dot plot of hair length, (b) associated sample statistics (c) box plot of hair length ....                                                                                                                                                             | 141     |
| Figure 7-2. Framework for thinking about statistical inference and sampling reasoning.....                                                                                                                                                                             | 148     |
| Figure 7-3. Karakare College population bag with data cards.....                                                                                                                                                                                                       | 152     |
| Figure 7-4. Example of data cards .....                                                                                                                                                                                                                                | 153     |
| Figure 7-5. Collated student graphs on classroom wall .....                                                                                                                                                                                                            | 155     |
| Figure 7-6. Students in class raising hands.....                                                                                                                                                                                                                       | 156     |
| Figure 7-7. How to make a call at curriculum level 5 and 6.....                                                                                                                                                                                                        | 157     |
| Figure 7-8. (a) hands close together (b) hands apart.....                                                                                                                                                                                                              | 162     |
| Figure 7-9. Structure of conclusion in 2009 worksheet .....                                                                                                                                                                                                            | 163     |
| Figure 7-10. Framework for thinking about statistical inference and sampling reasoning.....                                                                                                                                                                            | 165     |
| Figure 7-11. Random sample of year eight New Zealand boys and year eight New Zealand girls<br>(right foot lengths in cm).....                                                                                                                                          | 172     |
| Figure 7-12. Graph of repeated sample: (a) 2011 S8, pre-interview, and (b) 2009 S1, pre-interview                                                                                                                                                                      | 172     |
| Figure 7-13. Graph showing boys tending to have bigger right feet than girls<br>(2009 S1, pre-interview).....                                                                                                                                                          | 173     |
| Figure 7-14. Graph of repeated sample: 2011 S9, post-interview .....                                                                                                                                                                                                   | 174     |
| Figure 7-15. Abbreviated versions of pre- and post-test questions.....                                                                                                                                                                                                 | 177     |
| Figure 7-16. SOLO criteria for grading making the call.....                                                                                                                                                                                                            | 181–183 |
| Figure 7-17. Pre- and post-test responses for situation where no call could be made: (a) scatterplot<br>showing pre- and post-test grade, (b) summary table showing pre- and post-test grade,<br>and (c) dot plot showing difference in pre- and post-test grade ..... | 184     |
| Figure 7-18. Mean pre- and post-test scores for making a call: (a) scatterplot showing mean pre-<br>and post-test scores, (b) summary table showing mean pre- and post-test scores, and<br>(c) dot plot showing difference in mean pre- and post-test grades.....      | 185     |

|                                                                                                                                                                                                                                                                      |         |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|
| Figure 8-1. What Sina sees .....                                                                                                                                                                                                                                     | 191     |
| Figure 8-2. What Rita sees .....                                                                                                                                                                                                                                     | 192     |
| Figure 8-3. Connection between posing an investigative question, the population data bag and<br>the image of the distribution for a given variable.....                                                                                                              | 193     |
| Figure 8-4. Between data and distribution .....                                                                                                                                                                                                                      | 198     |
| Figure 8-5. Suggested theoretical framework of IIR .....                                                                                                                                                                                                             | 199     |
| Figure 8-6. Mental Processes Framework.....                                                                                                                                                                                                                          | 200     |
| Figure 8-7. Lessons 2–4 in teaching experiment .....                                                                                                                                                                                                                 | 202     |
| Figure 8-8. Four shape groups with graphs and additional ideas.....                                                                                                                                                                                                  | 206     |
| Figure 8-9. “Solutions” to description of shape from teacher’s activity sheet.....                                                                                                                                                                                   | 206     |
| Figure 8-10. Final collation of shapes into four groups with modality distinction.....                                                                                                                                                                               | 207     |
| Figure 8-11. (a) Reaction times graph, (b) Final description for reaction times graph.....                                                                                                                                                                           | 210     |
| Figure 8-12. Two examples of student descriptions that have been “actively reflected” on .....                                                                                                                                                                       | 211     |
| Figure 8-13. Example of symmetrical contexts from 2011 student class exercise book .....                                                                                                                                                                             | 212     |
| Figure 8-14. Components of shape descriptors.....                                                                                                                                                                                                                    | 214     |
| Figure 8-15. From lesson notes re: descriptions .....                                                                                                                                                                                                                | 215     |
| Figure 8-16. Distribution Description Framework for curriculum level 5 (ages 13–15).....                                                                                                                                                                             | 217     |
| Figure 8-17. (a) skeleton of a dog (b) showing muscular structure (c) the whole dog .....                                                                                                                                                                            | 218     |
| Figure 8-18. (a) drawing the shape (b) filling in the features .....                                                                                                                                                                                                 | 219     |
| Figure 8-19. (a) All Blacks’ (NZ rugby team) scores in test matches 2005–2010, (b) heights of NZ<br>Year 5–10 students, (c) heights of Tokoeka Kiwis (NZ native bird) .....                                                                                          | 220     |
| Figure 8-20. Heights of Tokoeka kiwis: (a) provided graph (b) example of sketched shape approx.<br>symmetrical and bimodal (c) example of sketched shape slightly right skewed and<br>unimodal (d) example of sketched shape slightly right skewed and bimodal ..... | 223     |
| Figure 8-21. Criteria for a good description of distributions .....                                                                                                                                                                                                  | 224–225 |
| Figure 8-22. (a) Pre- and post-test results for task C (b) Median grade movement from pre- to post-test<br>.....                                                                                                                                                     | 226     |
| Figure 8-23. Student response to task E, year 10 students (2011 student, pre-test).....                                                                                                                                                                              | 227     |
| Figure 8-24. Sketch of predicted distributions for age (a) year 10 class, (b) year 5–13 students C@S, ..<br>(c) teachers at NZ Maths conference (2011 student, post-test) .....                                                                                      | 228     |
| Figure 9-1. The four focus areas arising from the problematic situations. ....                                                                                                                                                                                       | 234     |
| Figure 9-2. Research questions within the four focus areas.....                                                                                                                                                                                                      | 234     |
| Figure 9-3. Framework for overarching themes .....                                                                                                                                                                                                                   | 235     |
| Figure 9-4. Conceptual framework for statistical investigations thread, curriculum level 5<br>of the New Zealand curriculum.....                                                                                                                                     | 236     |
| Figure 9-5. Thinking prompts for the four focus areas with examples.....                                                                                                                                                                                             | 241     |
| Figure 9-6. Connecting research outcomes to the statistics teaching cycle .....                                                                                                                                                                                      | 247     |

## List of Tables

|                                                                                                                 |         |
|-----------------------------------------------------------------------------------------------------------------|---------|
| Table 2-1. A Critical Orientation towards Statistics .....                                                      | 29      |
| Table 2-2. Matching graphs to purpose .....                                                                     | 31      |
| Table 2-3. The link between the type of investigative question and the nature of the data,<br>school level..... | 37      |
| Table 4-1. Summary information from teaching experiments with year 10 (ages 14–15) classes .....                | 68      |
| Table 4-2. Lesson detail teaching experiment one.....                                                           | 71–72   |
| Table 4-3. Lesson detail teaching experiment 2 .....                                                            | 74      |
| Table 4-4. Lesson detail teaching experiment 3 .....                                                            | 76–78   |
| Table 4-5. Lesson detail teaching experiment 4 .....                                                            | 79–81   |
| Table 5-1. Student pre- and post-test marks .....                                                               | 84      |
| Table 5-2. Pre- and post-test summary question categories.....                                                  | 90      |
| Table 5-3. Post-test comparison question categories .....                                                       | 91      |
| Table 6-1. Student-posed investigative questions highlighting aspects of language .....                         | 105     |
| Table 6-2. Population descriptors in 2008 post-test responses for questions about “being taller” .....          | 122     |
| Table 7-1. Student responses to support for the answer in 2008 post-test.....                                   | 142     |
| Table 7-2. Teaching and learning sequence for 2009 .....                                                        | 150–151 |
| Table 9-1. Concepts: instructional materials and activities .....                                               | 238–239 |

## Chapter 1. Introduction

In 2007 New Zealand introduced a new school curriculum for all years of schooling (years 1–13; ages 5–18) (Ministry of Education, 2007). The curriculum was organised into eight learning areas, one of which was titled *mathematics and statistics*. For the first time statistics was recognised as a separate, but connected, discipline to mathematics. The separateness and connectedness is articulated in the mathematics and statistics essence statement:

Mathematics is the exploration and use of patterns and relationships in quantities, space, and time. Statistics is the exploration and use of patterns and relationships in data. These two disciplines are related, but they use different ways of thinking and solving problems. Both equip students with effective means for investigating, interpreting, explaining, and making sense of the world in which they live.

Mathematicians and statisticians use symbols, graphs, and diagrams to help them find and communicate patterns and relationships, and they create models to represent both real-life and hypothetical situations. These situations are drawn from a wide range of social, cultural, scientific, technological, health, environmental, and economic contexts. (Ministry of Education, 2007, p. 26)

At the time the new curriculum was introduced I was working in a teacher professional development role supporting mathematics and statistics teachers in secondary schools. Two key focuses of my work were: (1) supporting mathematics and statistics teachers as they implemented the national qualification for years 11–13 (ages 15–18), the National Certificate of Educational Achievement (NCEA); and (2) supporting mathematics and statistics teaching and learning at all secondary levels, from years 9–13 (ages 13–18). Figure 1-1 summarises the connections between the year of schooling, age, curriculum level and NCEA level.

|                                     |       |       |        |        |        |
|-------------------------------------|-------|-------|--------|--------|--------|
| Year of secondary schooling         | 9     | 10    | 11     | 12     | 13     |
| Ages                                | 13–14 | 14–15 | 15–16  | 16–17  | 17–18  |
| New Zealand Curriculum level        | 4–5   | 5     | 6      | 7      | 8      |
| NCEA (national qualification) level |       |       | NCEA 1 | NCEA 2 | NCEA 3 |

**Figure 1-1. New Zealand secondary schooling years and associated curriculum and qualification levels**

In addition I had been involved in writing the mathematics and statistics curriculum, working primarily with the team that developed the statistics strand to better reflect the needs of our students in the 21st century. It was this collision of roles that became the impetus for this

thesis. From my perspective there needed to be a real purpose for undertaking research and it had to ultimately support student learning and make a difference for both students and teachers. As I was already working in teacher professional development and given my depth of involvement in the development of the statistics strand, it seemed logical to base my research in this area.

The 2007 curriculum saw major changes to the statistics strand. The statistical enquiry cycle (Figure 1-2) became a fundamental organising principle for framing the teaching and learning of statistics. Students needed to become data detectives, using exploratory data analysis and relevant contextual knowledge to inform each of the stages of the enquiry cycle.

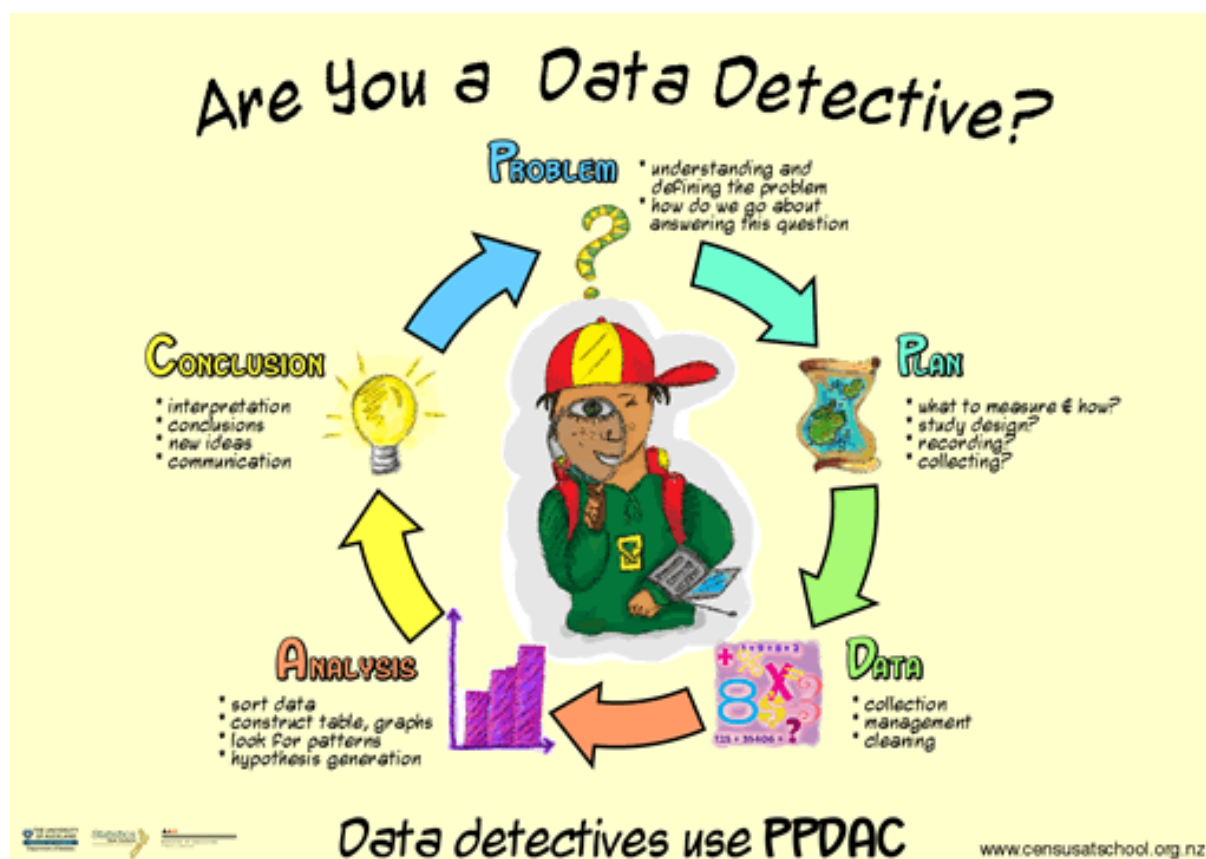


**Figure 1-2. The Statistical Enquiry Cycle**

*Note:* Reprinted from “Data Detective Poster,” by CensusAtSchool New Zealand, 2008, retrieved from <http://new.censusatschool.org.nz/resource/data-detective-poster/>. Copyright 2008 by CensusAtSchool New Zealand. Reprinted with permission.

In addition more emphasis was placed on reasoning from plots and statistics rather than focusing on the construction of plots or the calculation of statistics, and the mean and median were reconceived as properties of distributions rather than stand-alone statistics. Finally

conceptual understanding needed to be built across all levels and starting earlier than in previous curricula (Ministry of Education, 2010b).

### **1.1. Rationale for research**

In order to support teachers as they up-skilled themselves to meet the requirements of the new curriculum I decided to research teacher statistical content knowledge. The focus was at curriculum level 5 (years 9–10, ages 13–15) as this was the foundational curriculum level for national qualifications (years 11–13, ages 15–18). Initially the research involved exploring the new curriculum with teachers to ascertain their statistical content knowledge needs, followed by observing one of the teachers in the classroom. It was at the point of reflection on the students' post-tests and the realisation that there was a large gap in the statistics knowledge base that the focus for the research changed from a focus on teacher statistical content knowledge to a focus on posing investigative questions. Posing investigative questions is a key activity that students need to be able to show evidence of in order to achieve some of the statistics achievement standards in the national qualification (NCEA), in particular achievement standard *AS91035: Investigate a given multivariate data set using the statistical enquiry cycle* (New Zealand Qualifications Authority, 2010).

The major role that posing investigative questions plays in assessment for qualifications and the realisation that teachers were not aware that the investigative questions that students were posing were potentially flawed resulted in the crystallisation of the initial problematic situation, which is described more fully in chapter 5. Teachers and assessors need to know what makes a good investigative question, the components and concepts underpinning a good investigative question, and the learning that students need to be immersed in to support their posing of good investigative questions. Therefore, research was needed in the area of investigative questions to build a new knowledge base.

It could be argued that posing an investigative question is not necessary for statistical investigation, that it is possible to use a poorly constructed or conceived question to undertake an investigation. Expert statisticians and practitioners can cope with poorly constructed or conceived questions. The focus in this research, however, is on novices, students beginning their statistical journey, and it is important – I would even argue, critical – that their statistical foundations are solid. The expert statistician can cope, for example, with the idea of investigating the relationship between height and gender. The expert statistician

will undertake statistical analysis and use this analysis to infer about a wider universe. This expert statistician understands that they use a sample to make inferences about a wider population, that the poorly formed question is actually about a wider population and specific variables (measures). The novice, however, has yet to develop an understanding of the underlying concepts for posing investigative questions and subsequently how to make links between these concepts, statistical analysis and conclusions – reaffirming their need to build a knowledge base for posing investigative questions.

## **1.2. Research questions**

For my working hypothesis I conjectured that if students could learn to pose good investigative questions and if they really *understood* their investigative question then they would conduct better analyses and draw better conclusions. Based on four exploratory teaching experiments, the final research questions fell into four broad areas from the statistical enquiry cycle.

### ***Investigative question***

- What makes a good investigative question?
- What are the underpinning concepts that are needed to support teaching and learning around posing investigative questions?
- What level of comparative investigative questions are year 10 (ages 14–15) students posing?

### ***Prediction/ hypothesis***

- What distributional shapes and graphs do year 10 (ages 14–15) students predict when given the context?

### ***Analysis***

- What evidence do students use to “make the call” at curriculum level 5 (ages 13–15) given suitable learning experiences for developing criteria to make a call?
- What descriptors do year 10 (ages 14–15) students intuitively use for distributional shape?
- What makes a good distribution description at level 5 (ages 13–15) in the New Zealand curriculum?

## **Conclusion**

- What underpinning concepts do students need to support them to make a call at curriculum level 5 (ages 13–15)?
- Can year 10 (ages 14–15) students consistently and coherently make a statistical inference?

### **1.3. Scope of the research**

The participants in the research were four year 10 classes (students aged 14–15) and their teachers. The students and teachers came from two different multicultural secondary schools. One school was coeducational and the other single-sex (girls only). Both teachers were within their first ten years of teaching and had a passion for statistics. The research took place over five years, with four in-class teaching experiments in 2007, 2008, 2009 and 2011. The research method used was *design research* (Bakker, 2004a; Brown, 1992; Cobb, 2000a, 2000b; Hjalmarson & Lesh, 2008; Kelly & Lesh, 2000; Schwartz, Chang, & Martin, 2008), with an emphasis on innovative practice realised through an intervention, researching the intervention, and developing theories based on findings from the intervention. The repeated cycles of preparation and design, teaching experiment, retrospective analysis and identification of new problematic situations was eminently suited to the exploratory nature of the research. Qualitative and quantitative data were collected over the four teaching experiments and included pre- and post-tests, interviews, observations and field notes.

The curriculum focus was statistical investigations at level 5 (ages 13–15), with a specific focus on working with given (secondary) multivariate data sets. The four areas – investigative questions, prediction/hypothesis, analysis and conclusion – were all considered as part of the teaching and learning that was developed over the four teaching experiments. These areas reflect the curriculum at level 5 (ages 13–15) and map to the statistical enquiry cycle (Figure 1-2, page 2) well. The overall unit plans were developed collaboratively by the teacher and the researcher to follow the statistical enquiry cycle and reflected the change in emphasis in the new curriculum and acknowledged the necessity to prepare for the national curriculum and qualification (NCEA) in years 11–13 (ages 15–18).



#### **1.4. Outline of the thesis chapters**

This introductory chapter sets the scene for the research focus on posing investigative questions and the links between and across the stages of the statistical enquiry cycle in light of the new curriculum that was published in 2007. Chapter 2 reviews literature and research in the area of posing statistical questions and suggests that research in the area of what makes a good investigative question is needed. Chapter 3 presents the research methods driving the study, and the data collection and data analysis methods for both qualitative and quantitative data. Chapter 4 describes the research method within the context – the participants, the preparation and design phase, the teaching experiment, and the retrospective analysis in relationship to the four teaching experiments.

Results for the research are developed over chapters 5, 6, 7 and 8, starting with defining the problematic situation in chapter 5. Chapter 6 explores posing investigative questions and presents criteria for what makes a good investigative question and a possible teaching and grading framework. Chapter 7 looks at making links between the investigative question and the analysis and conclusion – “making the call” in comparative situations. Chapter 8 looks at making the link between the investigative question, the prediction/hypothesis and the analysis, i.e. describing distributions. In all four results chapters the retrospective analysis and the results of the quantitative and qualitative data are given. Finally, chapter 9 makes links across all of the results chapters discussed previously, identifying overarching themes, and suggesting implications for research and practice. Limitations of the current research are also discussed in chapter 9.

## **Chapter 2. Literature Review**

### **2.1. Introduction**

Posing questions was identified as a potential area for investigation during the first teaching experiment in this research when students were unable to complete an assessment task. The students had selected an unsuitable question to complete the investigation and as a result were unable to complete the analysis and therefore the investigation. This initial problematic situation and how it was realised is described in detail in chapter 5. After it was identified that students were finding it difficult to pose questions, the literature was initially reviewed with the goal of discovering what makes a good statistical question.

The focus in chapter 2 is on statistical questions within the statistical enquiry cycle (see Figure 1-2, page 2) as realised in the New Zealand curriculum (Ministry of Education, 2007). In this literature review the scene is set by discussing some of the assumptions that this research is based on and clarifying why particular perspectives are used. A theory on statistical questions is argued based on a review of the statistical enquiry cycle in its various guises across countries, authors and the literature. Related language and definitions are discussed to complete the picture of the perceived issues around posing questions.

### **2.2. Assumptions**

This section outlines some of the assumptions that have been used within this research. The difference between the novice and expert statistician is discussed as background on why some elements of the statistician's craft are highlighted and others are not. Secondary data sources are used extensively in the teaching and learning of statistics in the New Zealand setting and a short discussion on these and how secondary and primary data sources play out differently within the statistical enquiry cycle is given.

#### **2.2.1. Novice and expert statisticians**

The difference between novice and expert statisticians needs to be discussed from the outset. The novice statistician in this case is the student in school who is beginning their statistical journey and learning to think statistically. Novices come to the class with little or no previous experience, and any previous experience they may have is likely to be based on fallacy or

small-sample inference; for example, my brother had a bad experience with X, therefore X is bad (Watson, 2006). It is likely that novices have had limited interaction with specialised statistical terms, and where they are familiar with a term, it is often in the everyday use of the word rather than in the statistical sense. The novice is the clay to be moulded through experience into an “expert” statistician, one for whom statistical thinking is common place.

The expert statistician has many years of experience behind them; they have learnt through experience what works and what does not. “Statistical thinking is the touchstone at the core of the [expert] statistician’s art” (Wild & Pfannkuch, 1999, p. 223). Expert statisticians can work with vague and often ill-defined problems and seek solutions, knowing when they can infer about a wider universe and when they cannot. Elements of the statistical enquiry cycle that are meticulously adhered to when working with novice statisticians are not seen in the same way by expert statisticians or necessarily with the same importance.

This difference between novice and expert is important as statistical processes that are used with the novice are aligned to pedagogy, the teaching of statistical concepts, thinking and processes, and may not directly reflect the practice of the expert statistician. Pedagogical considerations are at the forefront of the work with novice statisticians (students) and therefore influence the decisions made around teaching and learning experiences for the novice. Novice statisticians are practising the craft of the expert statistician, but not necessarily applying the practice of the expert statistician.

### 2.2.2. Primary and secondary data sources

The statistical enquiry cycle that the New Zealand curriculum is based on is also known as the PPDAC (said as P-P-DAC) cycle (MacKay & Oldford, 1994; Wild & Pfannkuch, 1999), with the five stages **P**roblem, **P**lan, **D**ata, **A**nalysis and **C**onclusion forming the mnemonic. The statistical enquiry cycle given in Figure 1-2 (page 2) shows the investigation as one where the student collects the data (*primary data*) for subsequent analysis. This situation, although forming part of the student’s statistical journey, is not the primary situation in statistics education at the secondary school level in New Zealand. Often students are working with *secondary data*, data that has been collected for a different purpose, and they are now using this secondary data to do a secondary investigation. The terms primary data and secondary data are not common place and more often the data is referred to as either data that is collected by the student (primary) or data that is given to the student (secondary).

In New Zealand many teachers and their students use the CensusAtSchool databases for their statistical enquiry and most of the material used in this research is based on the CensusAtSchool databases with a few other secondary data sets used. CensusAtSchool, [www.censusatschool.org.nz](http://www.censusatschool.org.nz), is a biennial online survey that New Zealand students in years 4–13 (ages 9–18) complete. The survey usually has about 30 questions and generates a mixture of categorical and numerical data. Using secondary data, like the CensusAtSchool data, requires students to become familiar with the meta-data and to ask questions such as: What was the original survey question asked to generate a particular variable? How was the variable measured? How was the data collected? Why was the data collected? and What were the investigators' original questions? This last question is pertinent because the investigators' original questions are not included with the CensusAtSchool data – the main purpose of the survey being to provide useful and relevant data for students to explore as they develop their statistician's craft.

One point to consider regarding secondary data is the richness of the data set. While CensusAtSchool has been specifically designed to provide a rich multivariate data set from which interesting secondary investigations can be carried out, this may not always be true of other secondary data sets.

When students use secondary data the statistical enquiry cycle is altered and students usually start with the **Data**, familiarising themselves with the meta-data associated with the particular data set. This familiarisation includes exploring the **Planning** stage; for example, by exploring what survey questions were asked, how the data was collected, and what was measured. Once this meta-data has been explored, the **Problem** can be established, followed by the **Analysis** and then **Conclusion** stages of the enquiry cycle.

### **2.3. Statistical investigations**

Statistical investigations is a key thread in the New Zealand curriculum (the other two threads being statistical literacy and probability). Statistical investigations are explored in this section of the chapter, initially looking at the New Zealand situation and then comparing this with perspectives from the United Kingdom (Graham, 2006) and the United States of America (Franklin et al., 2005). As will be shown, statistical questions are an integral part of statistical investigations across the different countries: in New Zealand, the new curriculum (Ministry of Education, 2007) shows a greater emphasis on statistical thinking and reasoning; in the

United States, the GAISE (Guidelines for Assessment and Instruction in Statistics Education) report provides a “conceptual *Framework* for K–12 statistics education” (Franklin et al., 2005, p. 5); and Graham’s book *Developing Thinking in Statistics* (Graham, 2006) provides a comprehensive look at statistical thinking in the United Kingdom.

### 2.3.1. Statistical investigations thread in the New Zealand curriculum

Statistical investigations was one of the main threads of the statistics strand in the 1992 New Zealand mathematics curriculum (Ministry of Education, 1992), along with exploring probability and interpreting statistical reports. The previous syllabus for schools, *Mathematics: Forms 1 to 4* (ages 11–15) (Department of Education, 1987) had general objectives in relation to statistics. For example, students should “develop the ability to collect, order, display, analyse, and interpret data” (Department of Education, 1987, p. 5), which would seem to be a precursor to the statistical enquiry cycle. From these beginnings, statistics within the mathematics curriculum has developed to the extent that statistics is now recognised as having different ways of thinking and solving problems from mathematics. In the New Zealand curriculum published in 2007 (Ministry of Education, 2007) there is a name change from *mathematics* to *mathematics and statistics* and the essence statement articulates the difference between mathematics and statistics – two connected but different ways of thinking.

Mathematics is the exploration and use of patterns and relationships in quantities, space, and time. Statistics is the exploration and use of patterns and relationships in data. These two disciplines are related, but they use different ways of thinking and solving problems. Both equip students with effective means for investigating, interpreting, explaining, and making sense of the world in which they live. (Ministry of Education, 2007, p. 26)

In the 2007 New Zealand curriculum the three main threads that were in the 1992 mathematics curriculum are still evident: statistical investigation is couched explicitly within the statistical enquiry cycle; interpreting statistical reports is rebranded as statistical literacy with a clearer focus on understanding and critically evaluating other people’s reports; and exploring probability is concentrated more on linking empirical and theoretical outcomes. To illustrate the change from the 1992 curriculum to the 2007 curriculum, the level 5 (ages 13–15) achievement objectives for the statistical investigations thread are given in Figure 2-1 (next page). Appendix A has a full list of statistical investigation objectives for the 2007 curriculum from levels 1 to 8 (ages 5–18).

| <b>Comparison of level 5 (ages 13–15) Statistical Investigation Objectives</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>1992 Curriculum</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | <b>2007 Curriculum</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| <p><i>Within a range of meaningful contexts students should be able to:</i></p> <p>Plan and conduct statistical investigations of variables associated with different categories within a data set, or variations of variables over time; consider the variables of interest, identify the one(s) to be studied, and select and justify samples for collection; find, and authenticate by reference to appropriate displays, data measures such as mean, median, mode, inter-quartile range, and range; discuss discrete and continuous numeric data presented in quality displays; collect and display comparative samples in appropriate displays such as back-to-back stem-and-leaf, box-and-whisker, and composite bar graphs.</p> <p>(Ministry of Education, 1992, p. 186)</p> | <p><i>In a range of meaningful contexts, students will be engaged in thinking mathematically and statistically. They will solve problems and model situations that require them to:</i></p> <p>Plan and conduct surveys and experiments using the statistical enquiry cycle:</p> <ul style="list-style-type: none"> <li>• determining appropriate variables and measures;</li> <li>• considering sources of variation; gathering and cleaning data;</li> <li>• using multiple displays, and re-categorising data to find patterns, variations, relationships, and trends in multivariate data sets;</li> <li>• comparing sample distributions visually, using measures of centre, spread, and proportion;</li> <li>• presenting a report of findings.</li> </ul> <p>(Ministry of Education, 2007, p. 55)</p> |

**Figure 2-1. Comparison between 1992 and 2007 level 5 (ages 13–15) statistical investigation achievement objectives**

In the 2007 curriculum three particular changes can be noted:

1. There is an increased emphasis on the statistical enquiry cycle within the statistical investigations thread; therefore, teachers and students need to become familiar with all aspects of this cycle.
2. The mention of sample distributions confirms the intention that students are expected to use samples to make inferences about populations, i.e. informal *inferential* reasoning and *descriptive* reasoning are required during data analysis.
3. The objectives refer to thinking progressions at a generic level rather than specifying the content to be learnt.

Wild and Pfannkuch's (1999) description of the model for the investigative cycle (the statistical enquiry cycle) forms the basis of the statistical investigation strand in New Zealand schools. Curricular support material, such as the items found on the CensusAtSchool website ([www.censusatschool.org.nz](http://www.censusatschool.org.nz)), have been developed with the investigative cycle explicit.

Internationally, the necessity to emphasise statistical enquiry is borne out by the statistics education literature published since 1992 (e.g. Bishop & Talbot, 2001; Franklin & Garfield, 2006; Franklin et al., 2005; Friel & Bright, 1998; Glencross & Mji, 2001; Graham, 2006; Konold & Higgins, 2003; Stuart, 1995; Wild & Pfannkuch, 1999; Zayac, 1991). The different authors present the statistical investigative cycle with either four or five phases. The commonalities amongst them are greater than the differences. Without exception they are cyclical, have at their beginning a problem to be solved or a question to be answered, and at their end a conclusion.

This literature review explores the place of questions and questioning within the statistical investigative (or enquiry) cycle. Three frameworks for statistical enquiry (Franklin et al., 2005; Graham, 2006; Wild & Pfannkuch, 1999) are now discussed separately, and then compared with one another. The place of questions and questioning as appropriate is identified.

### **2.3.2. The four-dimensional framework for statistical thinking in empirical enquiry**

Wild and Pfannkuch (1999) developed a framework for statistical thinking in empirical enquiry based on reading of literature, their own experiences, and interviews with statistics students and practising statisticians. Their framework has four dimensions: the first of these is the investigative cycle, and the other three dimensions are the types of thinking, the interrogative cycle, and the dispositions. All elements of this framework need to be considered as “the thinker operates in all four dimensions at once” (Wild & Pfannkuch, 1999, p. 225). This indicates that to understand students’ reasoning within the investigative cycle, consideration should also be given to the other three dimensions because they impact on the investigative cycle.

In their interviewing, Wild and Pfannkuch (1999) noted that the statisticians gave prominence to the early stages of the investigative cycle, in particular to “grasping the dynamics of a system, problem formulation, and planning and measurement issues” (p. 225).

#### ***The Investigative Cycle***

This first dimension of the framework is concerned with what one thinks about and the way in which one acts during a statistical investigation. Wild and Pfannkuch (1999) worked with the PPDAC model (MacKay & Oldford, 1994) of the statistical investigative cycle:

- The **problem** stage deals with grasping a particular system's dynamics and understanding and defining the problem.
- The **planning** stage involves deciding what to measure and how these will be measured, how the sample will be taken, the design of the study, and how the data will be managed, including the recording and collecting of data. It also includes piloting the investigation and planning the analysis.
- The **data** stage is concerned with collecting, managing and cleaning the data.
- The **analysis** involves sorting the data, constructing tables and graphs as appropriate, exploring the data, looking for patterns, planned and unplanned analysis, and generating hypotheses.
- The final stage of the cycle involves interpreting, generating **conclusions**, new ideas and communicating findings.

In this first dimension, questions and questioning arise in all areas. Questions are posed in both the problem and planning stages, in particular, although questions are asked in all stages. Definitions and clarification of the purposes of these questions are discussed in detail in section 2.4 (page 19).

### ***Types of Thinking***

In the second dimension Wild and Pfannkuch (1999) split the types of thinking into two distinct categories: thinking that is inherently statistical, and general types of thinking.

#### TYPES FUNDAMENTAL TO STATISTICAL THINKING (Foundations)

- Recognition of need for data
- Transnumeration (changing representations to engender understanding)
  - capturing “measures” from real system
  - changing data representations
  - communicating messages in data
- Consideration of variation
  - noticing and acknowledging
  - measuring and modelling for the purposes of prediction, explanation, or control
  - explaining and dealing with
  - investigative strategies
- Reasoning with statistical models
- Integrating the statistical and contextual
  - information, knowledge, conceptions

#### GENERAL TYPES

- Strategic



- planning, anticipating problems
  - awareness of practical constraints
  - Seeking explanations
- Modelling
  - construction followed by use
- Applying techniques
  - following precedents
  - recognition and use of archetypes
  - use of problem solving tools (Wild & Pfannkuch, 1999, p. 226)

General thinking involves asking questions before the investigation commences and at each stage in the cycle. As the investigation progresses the thinking involves looking forwards and anticipating problems, and looking backwards and adjusting questions, data collection methods, and analysis.

Statistical thinking occurs when statistical processes within the cycle are questioned; for example: What type of data is needed to answer the question posed? Which graph is best for this type of data? What are the statistics, graph or table inferring about the data? and What are the statistics, graph or table inferring about the population? When considering the five fundamental types of thinking, the rationale and purpose behind using each of them depends on the question that has been posed. Thus the posed question is an integral and key part of consequent thoughts and actions during statistical enquiry.

### ***The Interrogative Cycle***

The interrogative thinking process is in constant use when people are statistically problem solving. The cycle has five stages to it:

- **Generate:** thinking about plans of attack, brainstorming, looking for possible explanations and models, what information is needed.
- **Seek:** recalling or seeking information, digging into memories for information that may be relevant, finding information from other sources, reading relevant literature, querying the data at hand.
- **Interpret:** “taking and processing the results of our seeking.  
Read/see/hear → translate → internally summarise → compare → connect” (Wild & Pfannkuch, 1999, p. 232). There is a need for the interconnection of ideas within this stage, connecting new ideas to existing ideas and growing one’s mental models.

- **Criticise:** this involves checking incoming information against reference points; asking questions such as “Is this right?” “Does this make sense?” (Wild & Pfannkuch, 1999, p. 232). At this stage one’s own thinking should be monitored. Is one being unduly influenced by his or her belief system or emotional response to a situation?
- **Judge:** this is the end point with a decision to be made. Decide what to believe, to continue to think about, what to discard.

The interrogative cycle seems to be the key dimension when questioning all aspects of the enquiry cycle. For such questioning to become operational, one may need some subject matter knowledge, experience and a certain type of disposition. For the learner, however, some thinking tools to instigate such questions may be needed (Wild & Pfannkuch, 1999). It may be important to deliberately develop interrogative questioning within student instruction. Some of these deliberate questions could be the following:

- **Generate:** questioning at the planning stage – What questions need to be asked to collect the information required? Will this question work? Is the question clear and unambiguous?
- **Seek:** interrogating the data – Where did the data come from? What is the background context? What questions were asked to get the data?
- **Interpret:** Should I be updating my questions? How do I update my questions? How do these ideas connect with what I already know?
- **Criticise:** challenging all phases of the cycle, asking questions such as: Do these results make sense with what I know?
- **Judge:** finalising the problem, deciding on the nature of the data, making the final decision about how to proceed.

### ***Dispositions***

Wild and Pfannkuch (1999) look at the personal qualities a person needs to initiate or effect entry into a thinking mode. The qualities are curiosity and awareness, engagement, imagination, scepticism, being logical, perseverance, openness and a propensity to seek deeper meaning.

These qualities also play a role in questions and questioning. Without curiosity, awareness, and a desire to seek deeper meaning, there would be no questions or problems to investigate. Imagination comes into play when developing questions, and when questioning or

interrogating different stages of the cycle one must have a healthy dose of scepticism, perseverance, openness and a desire to seek deeper meaning.

### **Summary**

All four dimensions are key for this study in the statistics education domain. This research with its focus on questions and questioning touches on all four dimensions of the framework for statistical thinking in empirical enquiry. The overarching dimension is the investigative cycle.

### **2.3.3. The PCAI framework**

*Developing Thinking in Statistics* (Graham, 2006) is a book written for primary and secondary mathematics teachers in the United Kingdom. It is a practical book with tasks and pedagogic ideas for the reader to work through, and is based on research and effective practice. *Developing Thinking in Statistics* was sourced with a view to the United Kingdom perspective on statistics teaching.

Graham (2006) believes that an investigation is good way to teach statistics in a motivational way that develops useful skills, but cautions that it is important to place “statistical work in meaningful and purposeful contexts” (p. 28). He says, “the defining characteristic of an investigation is that it is based on a *question* that learners want to answer”. Graham (2006) uses a statistical investigation cycle, which closely matches the investigative cycle of Wild and Pfannkuch (1999). The name for Graham’s framework, PCAI, is (like PPDAC) an acronym for the stages in his investigative cycle.

#### ***P: Pose the question***

According to Graham (2006), posing the question is critical as it is the key to everything that follows. He says that students will not find this stage particularly easy as generally they have had little or no previous experience. He suggests five considerations for coming up with a good question (see pages 28-29). Graham stresses that while posing the question, students should be going through in their minds how the whole investigation might work. For example, problems that are foreseen may be able to be fixed if it is a problem with the wording of the question.

### ***C: Collect the data***

Graham (2006) considers the two sources of data that are most likely to be used: a primary source, i.e. data collected by students, or secondary sources, i.e. data already collected by someone else. If primary sources are being considered, then decisions on how the data will be collected need to be made. The investigation could be an experiment or a questionnaire, and sampling issues need to be considered as well as choices of measure.

### ***A: Analyse the data***

This stage involves collating the data into tables, calculating summary statistics, and/or drawing graphs as appropriate. Considering whether the data are single or paired, or discrete or continuous, helps decide what statistical elements would be best to use in the analysis stage. Students need to know “what insights graphs and summaries can provide about the data and how these can be used to inform their enquiry” (Graham, 2006, p. 216).

### ***I: Interpret the results***

The final step in the cycle is where students relate their analysis to the initial question. At this point they must consider seriously whether or not their question has been answered. If it has not, how might they now proceed?

#### **2.3.4. Linking PPDAC with PCAI**

The two investigative cycles proposed by Wild and Pfannkuch (1999) and Graham (2006) have basically the same components. The only real difference is that Wild and Pfannkuch separate out the planning and data stages whereas Graham has these included in the same stage. This is shown in Figure 2-2.

| <b>PPDAC (Wild &amp; Pfannkuch, 1999)</b>                          | <b>PCAI (Graham, 2006)</b> |
|--------------------------------------------------------------------|----------------------------|
| Problem – understanding and defining                               | Pose the question          |
| Plan – what to measure and how                                     | Collect relevant data      |
| Data – collection, management, cleaning                            |                            |
| Analysis – sort, look for patterns, generate hypotheses            | Analyse the data           |
| Conclusion – interpretation, conclusions, new ideas, communication | Interpret the results      |

**Figure 2-2. Comparison of PPDAC and PCAI models**

### 2.3.5. Statistical problem solving

The *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report* (Franklin et al., 2005) provides a “conceptual *Framework* for K–12 statistics education” (p. 5) in the United States of America. The framework has four process components, which reflect an investigative process. The processes are formulate questions, collect data, analyse data, and interpret results, and are similar to the PPDAC and PCAI models.

- ***Formulating the questions*** involves clarification of the problem at hand and developing questions that can be answered with the data.
- The ***data collection*** component refers to the planning phase and the collection of the data. As Graham (2006) did with the PCAI cycle, the GAISE report does not separate out the plan and the data as two different components.
- ***Analysing the data*** involves appropriate selection of methods to analyse the data, including graphical and numerical.
- The final component of ***interpreting results*** includes relating the interpretation of the analysis to the original question.

### 2.3.6. Summary

The three models discussed are all cyclical. As Bishop and Talbot (2001) state, “The model is cyclic because, at each stage, one must look forward as well as backward and the conclusions relating to one problem lead on to a new problem” (p. 215).

The statistical investigation cycle has posing questions at its heart, and as one delves deeper into each of the stages, questioning is a key facet of all stages. The literature on the enquiry cycle gives prominence to the importance of formulating the statistical question, and in the teaching setting, the question needs to be one that students want to answer.

## 2.4. Questions within the statistical investigation cycle

The previous section discussed the different stages or phases of a statistical investigation cycle. In this section the role of questions and their purpose within the PPDAC statistical investigation cycle is discussed.

I argue that statistics education researchers have not defined the types of questions or differentiated between the purposes of questions used at different points in statistical enquiry. The consequence is a lack of clarity in the design of their studies and their findings.

I will, first, present my theory including two frameworks on statistical questioning and then I will use this theory to illustrate how researchers lack clarity in this area.

### 2.4.1. Question posing and question asking – a theory

The motivating question for this literature review was: What makes a good statistical question? A review of statistics education research found a number of studies where forming statistical questions was part of the researched process (e.g. Burgess, 2007; Hancock, Kaput, & Goldsmith, 1992; Lehrer & Romberg, 1996; Pfannkuch & Horring, 2005; Russell, 2006) and a number of papers or books where forming statistical questions were reported as part of an overview of the current status of statistics education (e.g. Graham, 2006; Konold & Higgins, 2002; Whittin, 2006). As a result of the literature review, and considering the statistical investigative cycle, the picture of what makes a good statistical question was still unclear. There were mixed messages about the purpose of statistical questions, about whether they were used for an investigation or to collect data from people. From the literature and from my own experiences I concluded that within statistical investigations there are two types of questions: those that are formally posed and those that are spontaneously asked throughout the investigative process.

My theory is that there is *question posing* and *question asking*. Question posing results in a question being formally structured, whereas question asking is a continual spontaneous interrogative process. Question posing arises as a result of having a problem that needs to be addressed using a statistical investigation. Posed questions may be asked for investigative or survey purposes: *investigative* questions are those to be answered using data (the problem), while *survey* questions are those asked to get the data (the plan). Question asking also has two purposes, both of which involve an interrogation element: *interrogative* questions are those asked as checks within the cycle (the problem, the plan, the data, the analysis, the

conclusion), while *analysis* questions are those asked about the statistics, graphs and tables in order to develop a description of and an inference about what is noticed (the analysis). Figure 2-3 summarises these question purposes across the PPDAC cycle.

| PPDAC cycle     | Problem       | Plan          | Data          | Analysis               | Conclusion    |
|-----------------|---------------|---------------|---------------|------------------------|---------------|
| Question posing | Investigative | Survey        |               |                        |               |
| Question asking | Interrogative | Interrogative | Interrogative | Interrogative Analysis | Interrogative |

**Figure 2-3. Question posing and question asking within the PPDAC cycle**

The role of questions can be further clarified by placing the four different question purposes into the context of the statistical investigation cycle. Figure 2-4 (next page) suggests how questioning fits within the statistical investigative cycle when the investigation involves collecting the data, i.e. using primary data.

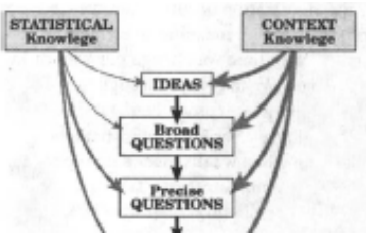
| The PPDAC cycle in situations where the student collects the data (primary data).                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>PROBLEM</b><br>                                                                                                                                                                                                                                                                                                                         | <p>Motivating situation/question/idea.</p> <p>Could be either for a survey or experiment situation.</p> <p>POSE <i>investigative</i> question.</p> <p><i>Interrogate</i> investigative question.</p>                                |
| <b>PLAN</b><br>                                                                                                                                                                                                                                                                                                                            | <p>Develop measurement instruments and data collection procedures.</p> <p>POSE <i>survey</i> questions.</p> <p><i>Interrogate</i> plan.</p> <p>Update investigative question as appropriate and as necessary based on planning.</p> |
| <b>DATA</b><br><p>Collect data.</p> <p>Revisit the investigative question updating if necessary (maybe pose new investigative questions that have become apparent after collecting the data).</p>                                                                                                                                                                                                                           |                                                                                                                                                                                                                                     |
| <b>ANALYSIS</b><br><p>Calculate statistics, draw graphs, create tables as appropriate.</p> <p>ASK <i>analysis</i> questions about the statistics, graphs and tables and describe what is noticed and what is inferred.</p> <p><i>Interrogate</i> analysis.</p> <p>Revisit the investigative question updating if necessary (maybe pose new investigative questions that have become apparent while analysing the data).</p> |                                                                                                                                                                                                                                     |
| <b>CONCLUSION</b><br><p>Write the “conclusion” answering the investigative question(s).</p> <ul style="list-style-type: none"> <li>• Use supporting evidence from the analysis.</li> <li>• Make inferences about the population.</li> </ul> <p><i>Interrogate</i> conclusion.</p> <p>POSE further <i>investigative</i> questions as a result of “conclusion”.</p>                                                           |                                                                                                                                                                                                                                     |

**Figure 2-4. Questions within the statistical investigative cycle: primary data**

*Note:* Statistical and context knowledge diagram reprinted from “Statistical thinking in empirical enquiry,” by C. Wild and M. Pfannkuch, 1999, International Statistical Review, 67(3), p. 228. Copyright 1999 by the International Statistical Review. Reprinted with permission.



Figure 2-5 shows the same for an investigation that is based on using a given multivariate data set, i.e. using secondary data.

| The PPDAC cycle in situations where the data is given to the students (secondary data).                                                                                                                                                                                                                                                                                                                                                                                       |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p><b>Original Investigators DATA/PLAN</b><br/>Data set is given.</p> <p><i>Interrogate</i> the background to the data.<br/>Understand the original <i>investigative</i> question(s) and data collection plan including how the data was collected and who the data was collected from. Find out the <i>survey</i> questions used. Identify the variables of interest and what they measure, and identify the population of interest.</p> <p><i>Interrogate</i> the data.</p> |  |
| <p><b>PROBLEM</b><br/>Motivating situation/question/idea of what “I” will investigate using this provided data.</p> <p>POSE new <i>investigative</i> question.</p> <p><i>Interrogate</i> investigative question.</p>                                                                                                                                                                                                                                                          |  |
| <p><b>ANALYSIS</b><br/>Calculate statistics, draw graphs, build tables as appropriate.</p> <p>ASK <i>analysis</i> questions about the statistics, graphs and tables and describe what is noticed and what is inferred.</p> <p><i>Interrogate</i> analysis.</p> <p>Revisit the investigative question, updating if necessary (maybe pose new investigative questions that have become apparent after analysing the data).</p>                                                  |  |
| <p><b>CONCLUSION</b><br/>Write the “conclusion” answering the new investigative question(s).</p> <ul style="list-style-type: none"> <li>• Use supporting evidence from the analysis.</li> <li>• Make inferences about the population.</li> </ul> <p><i>Interrogate</i> conclusion.</p> <p>POSE further investigative questions as a result of “conclusion”.</p>                                                                                                               |  |

Figure 2-5. Questions within the statistical investigative cycle: secondary data

## 2.5. Question posing

Posing questions has been noted and identified as a potential problem area for students and teachers by Burgess (2007), Pfannkuch and Horring (2005), and Rubick (2000). Research on posing questions appears to be limited.

Different authors and researchers refer to question posing in different ways. For example, discussions include answering “authentic questions” (Hancock et al., 1992), formulating “statistical questions” (Konold & Higgins, 2003; Russell, 2006), answering “motivating

questions” (Pfannkuch & Horring, 2005), posing “productive questions” and “questions of great interest” (Lehrer & Romberg, 1996), posing “appropriate questions or hypotheses” (Burgess, 2007), and formulating “realistic questions” that can be investigated (Graham, 2006).

### **2.5.1. Making the distinction between investigative and survey questions**

When referring to posing questions, there is a lack of clarity amongst researchers about the two different purposes of questions, i.e. investigative and survey. Some specify the purpose of the question that is being posed (e.g. Burgess, 2007; Graham, 2006; Pfannkuch & Horring, 2005; Rubick, 2000), while others appear to make the distinction but do not specify the distinction (e.g. Hancock et al., 1992; Russell, 2006). There are also those who appear to make no distinction between the two quite different purposes of question posing (e.g. Konold & Higgins, 2002, 2003; Lehrer & Romberg, 1996; Whittin, 2006). These three distinctly different situations that exist in the literature will now be elaborated upon.

#### ***Explicit distinction***

Burgess (2007), in his research on teacher knowledge, clearly makes the distinction between the two purposes for posing questions. He researched specialised content knowledge, common content knowledge, knowledge of content and teaching, and knowledge of content and students, and related these to the aspects of Wild and Pfannkuch’s (1999) four-dimensional framework for statistical thinking in empirical enquiry. Within his research, posing questions for investigation was mentioned as a key area of knowledge needed by teachers when teaching statistics. Burgess showed that his teachers, while cognisant of the need for questions to be posed, found some challenges in this area when teaching. He also reported that posing questions for investigation was difficult at times for the students in the teachers’ classes. In addition, he reported on examples of teacher knowledge around data collection questions, which he identified as being distinct from questions posed for investigation.

Because so many statistical concepts were covered in the investigative process (from the posing of question for investigation, consideration of data collection questions, analysis through sorting and other transnumerative processes, and concluding statements), the examples given are a small sample covering a wider variety of statistical concepts.  
(Burgess, 2007, p. 166)

Pfannkuch and Horring (2005) report on the first year of a three-year project around developing a statistics curriculum for 15-year-old students. The project involved working partly with seven teachers in a secondary school and more specifically with one of the teachers to identify problematic areas in the 2002/2003 curricula. They detailed problematic areas that they identified and changes that were made to teaching approaches, a process which was repeated over a couple of years. In 2002 one of the key problem areas they reported on was posing questions. The issues around posing questions noted by Pfannkuch and Horring (2005, p. 208) include:

- students forgetting to go back and answer the question – “looked at the data and just talked about the data instead of going back to the questions”, and
- teachers’ ability to pose questions was limited – “teachers tended to pose narrowly framed statistical questions according to a template”.

In this instance the purpose of posing the question was to investigate a given data set.

### ***Implicit distinction***

In her article about data investigations, Russell (2006) concludes that there are two central issues: “What do we want to know and how can we formulate a way of collecting data that furnish good information? What do the data we collected and represented tell us about what we wanted to know?” (p. 29). In the first question Russell refers implicitly to the two purposes of posing questions: investigative (“What do we want to know?”) and survey (“How we formulate a way of collecting data”).

In their work on data modelling, Hancock, Kaput and Goldsmith (1992) identify two parts that make the whole: data creation and data analysis. They state that data creation has been neglected and that it is a critical part of the whole enquiry process. Hancock et al. implicitly mention the two different purposes of posing questions: they refer to a “problem of interest and concern to them ... devise a plan to collect data that will solve the problem” (p. 338). Hancock et al., however, did caution that their data-based inquiry projects started “without a clear question, and ended without clear answers” (p. 359), suggesting the importance of the investigative question being part of the process and being appropriate. Graham (2006) concurs with this idea of the investigative question being key: “... without a clear question, there are no criteria for deciding what to do next” (p. 26).

In the previous examples the distinction between the investigative question and the survey question was made by researchers, either explicitly or implicitly. The following examples give situations where the distinction is either not made, or is not clear.

### ***No clear distinction***

Konold and Higgins (2003) in their review of statistics education research devote some discussion to the forming of a statistical question. They report on the challenge for students and teachers of taking a general question and transforming this into a question that can be answered with data, i.e. a statistical question. Clearly they are discussing the investigative question. However, later on in their discussion there is a slight confusion as to which of the two purposes they are referring to, investigative or survey:

Elementary school students can learn a lot about data as they grapple with issues that arise in formulating statistical questions, especially when they anticipate conducting surveys with the questions they design. By thinking about how they would answer a proposed question, students quickly discover not only the range of different responses but also that multiple interpretations of a question are possible and the wording of the question matters. (Konold and Higgins, 2003, p. 195)

This paragraph begins with reference to the investigative question (“formulating the statistical questions”) but ends up talking about the survey question (“range of different responses but also multiple interpretations of a question are possible”).

Lehrer and Romberg (1996) report extensively on students posing questions in a project on data modelling. They reported that students were involved in posing questions about their own lives, which involved brainstorming, categorising and clarifying questions for investigation. A number of steps were involved in creating these questions, which were used as part of a survey. Therefore the initial part of the project is dealing with survey questions. Once the data were collected, the students were then asked to produce questions they could ask of the data (the investigative questions). Lehrer and Romberg (1996) identified a need for teachers to scaffold the task by giving seed questions to help students pose questions for this purpose. The students posed questions such as: “What is your favourite movie? Who watches more TV, boys or girls? What’s the most favourite subject?” (Lehrer & Romberg, 1996, p. 81). By my definition these three questions all have different purposes. The first question (“What is your favourite movie?”) is a survey question, the second question (“Who watches more TV, boys or girls?”) is an investigative question, and the third question (“What’s the

most favourite subject?”) is a question that would be *asked* during the analysis stage to answer an investigative question around favourite subjects.

My conjecture is that clarity around the purpose of posing questions may be one of the reasons students find difficulty with posing and understanding questions in statistical investigations. This difficulty may also influence the students’ ability to analyse data and draw conclusions. The next two sections specifically consider issues related to the two different purposes, the investigative question and the survey question.

### **2.5.2. The investigative question**

In the big picture of statistical enquiry the investigative question is the statistical question or problem that needs answering or solving. In most instances the investigative question starts from an “inkling” and is developed into a precise question. The process of developing or creating the investigative question is iterative and requires considerable work to get it right (delMas, 2004; Franklin et al., 2005; Hancock et al., 1992; Russell, 2006; Wild & Pfannkuch, 1999). There is also a need when developing the investigative question to have “an understanding of the difference between a question that anticipates a deterministic answer and a question that anticipates an answer based on data that vary” (Franklin & Garfield, 2006, p. 350).

Posing investigative questions has been identified as a problem area for students. One of the identified problems for students in posing investigative questions is related to the idea of asking questions of the data. Pfannkuch and Horring (2005) note that students lack understanding of what a question is, and include the idea that one can pose a problem by asking questions of data: “Maybe students haven’t yet formed that understanding of what a question is – how you can ask a question in a set of data” (p. 208). Lehrer and Romberg (1996) also found that students initially had problems with asking questions of data: “students believed that questions cannot be asked of data, only of people” (p. 80). Burgess (2007) noted that students found posing investigative questions a problem but did not specify the particular issue that arose.

Other issues related to investigative questions include teachers needing to model posing investigative questions, initially as seed or starter ideas (Burgess, 2007; Lehrer & Romberg, 1996; Pfannkuch & Horring, 2005; Russell, 2006), but also to start students thinking about,

for example, “typicalness” and data as an aggregate rather than individual cases (Bakker & Gravemeijer, 2004; Konold & Higgins, 2002, 2003).

### **2.5.3. The survey question**

There are a number of research studies where the generation and/or use of survey questions has been explored (Burgess, 2007; Hancock et al., 1992; Konold & Higgins, 2002, 2003; Lehrer & Romberg, 1996; Russell, 2006). The survey question is used to collect data to use in a statistical investigation. Formulating a good survey question is as important as formulating a good investigative question. A good survey question provides sensible and meaningful data (e.g. Russell, 2006).

Russell’s (2006) research primarily focuses on the survey question “How can we formulate a way of collecting data that furnish good information?” (p. 29). Russell found that students as young as second graders can define “their questions in a way that will be clear to those they survey and will provide information they can interpret accurately” (p. 19). The students, in her study, understood that the survey question needed to provide them with sensible data and needed to avoid multiple interpretations by the people who were surveyed. She did caution, however, against the danger “that a focus on creating a clear question can overshadow the focus on collecting meaningful data that are of interest” (p. 19).

Investigative and survey questions, therefore, serve quite different purposes and both have their own challenges for teachers and students.

### **2.5.4. Interrogating the questions posed**

In order to get precise investigative questions or survey questions that can be correctly interpreted and that yield useful information, an *interrogative* process, which involves asking questions of the investigative and survey questions, is necessary (Burgess, 2007; Graham, 2006; Hancock et al., 1992; Konold & Higgins, 2002, 2003; Russell, 2006; Whittin, 2006). Specific research, however, on how students develop the ability to interrogate data and the enquiry process has not been conducted. The following provide examples of interrogating investigative and survey questions that will be useful to consider when thinking about what makes a good investigative question.

Russell (2006) mentions briefly a situation where the teachers she was working with wanted to investigate the amount of reading that was done at home by students in their school. This is

an example of developing the investigative question. The article reports on, according to my definition, some of the interrogative questions that were asked to help form the investigative question, questions such as: “Which students should they ask? Over how long a period should they monitor reading? What should they count – number of books or number of hours spent reading?” Russell referred to this process as the teachers facing the same sorts of issues that a statistician faces: “How do you turn a question about events into a statistical question, that is, a question that can be answered with data?” (p. 17). The teachers’ starter question was “How much reading do students in our school do at home?”, leading to a final question of “How many hours of reading do first-grade and fifth-grade student do each day in a two-week period, as reported on the recording sheet and signed by a parent?”

Burgess (2007) acknowledges that some of the specialised content knowledge a teacher needs relates to their ability to be able to decide if a question posed by their students is suitable, unsuitable, or whether changes can be made to make the question suitable. Such decisions are equally applicable to the investigative and the survey question. He notes in particular that teachers need to ask whether the students will find the investigative question interesting. Whether an investigative question is interesting or productive is another aspect of interrogating the question. Hancock et al. (1992) also noted “the cola unit benefited from students’ strong interest in the outcome” (p. 359). In this situation the investigative question or problem and how it was determined led to the students’ interest.

Graham (2006) provides five useful considerations for forming a good investigative question. These considerations pick up a number of different aspects of interrogating the investigative question. The considerations are whether the question is:

- actually a question, rather than simply an area for investigation – an investigation based on a question is more likely to draw on statistical and mathematical skills and provide greater focus and clearer direction;
- personally interesting to you – not only will this bring greater motivation, but also your common-sense knowledge about the context should help to ensure that the investigation proceeds along sensible lines;
- likely to draw on data that will be available within the time frame of the investigation – for example, do not investigate the growth of flowers or plants during the winter months;
- specific, so that it is answerable from data – questions that are too vague and general are harder to answer;

- measurable – think through in advance what you will measure and whether it will help to answer the question. (p. 88)

Another approach to interrogation is thinking like a critic and developing a critical perspective about statistics. Whittin (2006) gives a number of suggestions about how to question the question. Some of his suggestions relating to posing questions are given in Table 2-1. Using his critical orientation towards statistics, Whittin reports on situations where both the teacher and students are acting as “critical consumers of data” (p. 33) and give examples of how each of these orientations look in a classroom setting. The habits of a critic can be applied to situations where the teacher and students are acting as data analysts and where they are acting as data consumers. Therefore this critical perspective is useful when students are interrogating the questions they have posed, both investigative and survey.

**Table 2-1. A Critical Orientation towards Statistics**

| <b>Dimension of the Process</b>             | <b>A Critic’s Perspective</b>                                          | <b>Important Questions for the Teacher to Ask</b>                                                                                                                                               |
|---------------------------------------------|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| The Motive                                  | The intentions of the surveyor influence all aspects of the process.   | Why did you decide to gather data about this topic?<br>What did you hope to find out?                                                                                                           |
| The Question                                | The way a question is posed influences the kind of responses received. | How did you ask your question?<br>Why did you ask it in that way?<br>How might this language have influenced the responses that were received?<br>How else might the question have been worded? |
| The Categories (could also be part of data) | Data can be aggregated or disaggregated to serve different purposes.   | How were the categories decided on?<br>What happened to responses that did not fit into these categories?<br>What other categories could have been created?                                     |
| The Definitions                             | Broad or narrow definitions determine what gets counted.               | How did you define this word?<br>Why did you define it this way rather than another way?                                                                                                        |

*Note:* Reprinted from “Learning to talk back to a statistic,” by D. Whittin, 2006, *Thinking and Reasoning with data and chance*, p. 32. Copyright 2006 by the National Council of Teachers of Mathematics. Reprinted with permission.

Other aspects of interrogating the investigative question is deciding on the population of interest (Konold & Higgins, 2003; Watson, 2008), and whether or not the investigative question relates to the nature of the data to be explored. This is dealt with in section 2.7 (page 34).



## 2.6. Question asking

The argument has been made that there are two purposes for questions related to the investigative cycle – questions may be posed for investigation or survey, which has been discussed in section 2.5, and asked for interrogation or analysis, which is discussed in this section. Question asking is a continual spontaneous interrogative process and has two purposes: (1) *analysis* questions ask about the statistics, graphs and tables in order to develop a description of and an inference about what is noticed (the analysis); and (2) *interrogative* questions are asked as checks within the cycle (the problem, the plan, the data (given data sets), the analysis, the conclusion). While question asking was not a focus of this research, clarifying what it is completes the argument about the two purposes for questions.

### 2.6.1. Analysis questions

As part of the statistical enquiry cycle, students, when doing the analysis, should be asking questions about their statistics, graphs and tables. Research in the area of what questions students ask about their statistics, graphs and tables is limited (Dunkels, 1991). The majority of the research available focuses primarily on questions teachers ask students about their statistics, graphs and tables (Friel, Curcio, & Bright, 2001; Graham, 2006), but these questions give some insight as to possible analysis questions students can ask themselves as they describe their graphs, tables and statistics.

Graph comprehension, according to Friel et al. (2001), is more than just reading and interpreting graphs; it also includes graph choice and construction or even invention. In their discussion they interpret graph comprehension in terms of literacy: *translation* equates to extracting data from a graph; *interpretation* equates to finding relationships and interpolating; and *extrapolating* equates to analysing the relationships and extrapolating from the data. Questioning is regarded as an important aspect of comprehension and Friel et al. (2001) suggest that “teachers need to develop a framework within which to think about which questions to ask. Such a framework for question-asking is relevant for considering comprehension of graphs” (pp. 129–130). Friel et al. (2001) classify three levels of asking questions that they extracted from a review of research: elementary, intermediate and overall questions. The three levels of graph comprehension are also referred to as “reading the data, reading between the data and reading beyond the data” (Friel et al., 2001, p. 130).

Graham (2006) suggests some other questions that teachers can ask students to help them to think more deeply about what a representation might reveal. These questions that teachers can ask are given in Table 2-2 and have been classified using Friel et al.'s (2001) classification – elementary (E) or intermediate (I).

**Table 2-2. Matching graphs to purpose**

| <b>Representation</b>  | <b>Possible teacher questions</b>                                                                                                                                                                                                              |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Bar chart or pie chart | What is the largest/smallest bar (or sector)? E<br>How do the bars (or sectors) compare? I<br>Taking all the sectors together, what does the complete pie represent? I<br>What is measured on the vertical scale of the bar chart? E           |
| Box plot               | What is the largest/smallest value? E<br>How widely spread are the values? I<br>Are the values mostly bunched in the middle? I<br>What does this box plot's shape tell you? I                                                                  |
| Table                  | How do the rows compare? I<br>How do the columns compare? I<br>What are the row and column totals? I<br>Would it be helpful to include percentages? I                                                                                          |
| Scatter plot           | Is there an underlying pattern to the points? I<br>Is the pattern a straight line (linear) or something else? I<br>Is the trend increasing or decreasing? I<br>How closely do the points lie to the line of fit and what does that tell you? I |

*Note:* Reprinted from *Developing thinking in Statistics* by A. Graham, 2006, p. 216 & 264. Copyright 2006 by A. Graham. Reprinted with permission.

I would argue that these questions that are asked by the teacher need to be asked by the students so that they progress from answering questions about graphs and other representations to asking and answering their own questions. Question asking is a critical aspect of graph comprehension. The level of the asking (and answering) questions relates to a level of comprehension and analysis. If students can ask (and answer) questions that read between and beyond the data then their level of statistical thinking and reasoning is high.

There is limited research available on students asking their own questions during empirical enquiry. In particular, Dunkels (1991) reports on his work with primary students who posed questions of data displays. Initially the teacher was suggesting the questions that could be asked and then the children were encouraged to ask questions themselves. The quality of the

questions ranged from reading the data (“Are there any pupils who are the same length?”) to reading beyond the data (“When the shortest will be 30 cm over 1 m, how much will the tallest be over 1 m?”) (p. 134). Therefore, it would seem that it is possible, through instruction, to instigate students to ask and answer their own questions of data displays.

### **2.6.2. Interrogative questions**

The purpose of interrogative questions is to act as a check within the PPDAC enquiry cycle in order to ensure that all available information is taken into account before proceeding to take an action.

#### ***Interrogating the question posed***

This was discussed in section 2.5.4, page 27.

#### ***Interrogating the plan – primary data***

Many of the issues around interrogating the plan are covered with interrogating the survey question. However, other data collection methods, defining the variable, and the types of measures to be used also need to be considered at this point. For example, questioning whether a sample was obtained by taking a random sample or whether the treatments were randomly assigned within the experiment context (Scheaffer, 2000) are aspects of interrogating the plan.

#### ***Interrogating the meta-data – primary and secondary data***

In a situation whereby the data are given– for example, when working with CensusAtSchool data – it is necessary to interrogate what is behind the data. Graham (2006), delMas (2004), Pfannkuch (2006), and Konold and Higgins (2003) suggest that interrogating the meta-data is necessary to get a really good feel about where it came from, how it was produced and recorded, what questions were asked to get the data, and who the data are about. Such questioning plays an important role in understanding the context and therefore the investigative question.

A lack of background information about the data can be an issue. This lack means that speculation about rather than exploration of the meta-data happens. When starting to explore data, particularly secondary data, background information is necessary as the context is a “factor in determining whether confounding variables are present, and for determining

whether there are alternative explanations for the findings” (Pfannkuch, 2006, p. 42). To support students’ understanding of the process of data collection, delMas (2004) argues that students should experience first-hand data collection processes as these help students to become familiar with the different processes. In addition he suggests, “these experiences should include the opportunity to ask why and how data is produced” (delMas, 2004, p. 92).

### ***Interrogating the analysis – primary and secondary data***

The analysis phase of the enquiry cycle by its nature involves much question asking. Questions such as those suggested by Friel et al. (2001) and Graham (2006) are some of the many analysis questions that could be asked. In addition, interrogative questions should be asked of the analysis, such as “Is this a large outcome? or Is this a surprising result?” (Moore, 1990, p. 134) or “Why did you decide to show your information this way? How else could you have displayed your data? What information is concealed or revealed by representing it in this way, for example, how does a pie chart differ from a bar graph in representing the same data set?” (Whittin, 2006, p. 32).

### ***Interrogating the conclusion – primary and secondary data***

Whittin (2006), from his critic’s perspective, suggests the following questions as a starting point for interrogating the conclusion. “What do the data not show? ... How are your results different from your conclusions? How did your attitudes about this topic influence the decisions you made during this whole process?” (p. 32). Watson (2008) refers to a difficult question that was discussed by her students: “Which ‘group’ was the data we collected true for?” (p. 70). This question is relevant to interrogating the conclusion, and in writing the conclusion.

Without a doubt, when interrogating the conclusion one question that should be asked is: Does the conclusion answer the investigative question? In fact the link to the investigative question and the real world context is paramount in all stages of the cycle. As Konold and Higgins (2002) said:

During all phases of data analysis it is critical that students not lose sight of the questions they are pursuing and of the real world events from which the data come. These connections are easier to maintain when students work with data from familiar contexts and use representations they understand. (p. 195)

## **2.7. Investigative questions and nature of data**

The investigative question has an influence on the nature of data to be collected, especially in a primary data situation. The nature of data collected has an influence on the investigative question when it is a secondary data situation. The investigative question also defines the way the data will be analysed, i.e. what statistics, graphs and tables will be used and how they will be interpreted.

### **2.7.1. The nature of data**

Defining the nature of data is fraught with problems due to different definitions and classifications. Data can also be reclassified and in the process change its nature. Language is also an issue, especially when particular words, such as relationship, have a specialist meaning and also a more general meaning in statistics, and this can be unhelpful when learning concepts for the first time. Classifications with respect to the nature of data include the terms such as: univariate, bivariate and multivariate; categorical, numerical and measurement; and discrete and continuous. Since the nature of the data and the investigative question are intimately linked, definitions and classifications of these terms are discussed. The consequent impact these definitions have on the investigative question will then be explicated.

A simple classification would start with defining discrete and continuous data. Discrete data is “data which can take only certain values, such as the number of legs on animals” and continuous data is “data which could, in principle, assume any other value between any two given values. It is usually data collected by measurements, such as length” (Ministry of Education, 1992, pp. 211–212). These definitions of discrete and continuous data are similar to other references in New Zealand, United Kingdom and Australia (Barton, 2006; Department of Education, 1986; Graham, 2006; Lowe, 1991; Turner & Nightingale, 1997). McGillivray (2007) in Australia, however, splits variables into continuous and discrete and then defines discrete variables as categorical and count data. Wild and Seber (2000) in New Zealand, on the other hand, differentiate between quantitative variables (measurement and counts), which are split into continuous and discrete data, and qualitative variables, which are separated into categorical and ordinal data. In the United States, Franklin et al. (2005) refer to categorical and numeric data, where data that are non-numerical categories are positioned as categorical data and situations where the data are found by measurements being taken or objects counted are positioned as numerical data. By definition the numeric counted data

would appear to be the same as discrete data and numeric measured data would appear to be the same as continuous data.

Another classification method for data is based on “measurement” type. Graham (2006) describes Stevens’ taxonomy of measurement scales:

- A nominal (or naming) scale of measurement is used for named categories such as race, national origin, gender, surname, and so on.
- An ordinal (or ordered) scale of measurement involves data which can be ranked in order (first, second, third, and so on), but for which the numbers cannot be used for further calculation.
- An interval scale of measurement is a rather subtle concept – numbers are used for measurement of the amount of something, but the scale is such that the zero is arbitrary.
- A ratio scale of measurement has all the properties of an interval scale but, additionally, the operations of multiplication and division can be used meaningfully. (pp. 8–9)

Further reading of the definitions would suggest that some interval scale measurements could also be classified as discrete data and some interval and all ratio scale measurements could be classified as continuous data. Nominal and ordinal scale measurements are categorical data and these tend to be treated the same way as discrete data.

Lehrer and Romberg (1996), with their work on children’s data modelling, also classified and defined data types. The children, when given the opportunity to group the different types of answers to their survey questions, sorted them into yes and no (Boolean), numbers (integers and real), and categories. Lehrer and Romberg introduced another type of answer: string (for example, a person’s name). These classifications of data types are linked to *Tabletop*, the software package the children used for analysis. Again classifying these types into discrete and continuous, it would seem that Boolean, categories and string are categorical data, and the integer number is discrete while the real number is continuous. On reflection, it is noticeable that there is no common language across countries, authors and software when classifying types of data.

Furthermore, data can be classified according to the number of variables involved. “Data falls into three categories, depending on the number of variables involved in the information being collected – univariate, bivariate, and multivariate” (Department of Education, 1986, p. 94).

Traditionally statistics education focused on working with univariate or bivariate data sets (Pfannkuch & Watson, 2003; Ridgway, McCusker, & Nicholson, 2005). More recently the move has been towards using multivariate data sets (Pfannkuch & Horring, 2005; Pfannkuch & Watson, 2003; Ridgway et al., 2005). This allows the opportunity for exploration with multiple variables (Ridgway et al., 2005) and students to be selecting variables from a multivariate situation and posing questions about them. Bivariate situations occur when two variables are involved, such as when data are paired for an individual and a relationship is sought or when comparisons are made between two variables such as gender and height. Univariate data are singular and are usually summarised and described.

If discrete and continuous data are taken as single or paired there are four distinct combinations: single and discrete, single and continuous, paired and discrete, and paired and continuous (Graham, 2006). Graham suggests that these distinctions are useful for students when choosing appropriate graphs or statistical tools for their analyses as well as the type of question being investigated. However, by disregarding categorical variables, his classification is incomplete, particularly when considering the type of data inherent in investigative questions that are posed.

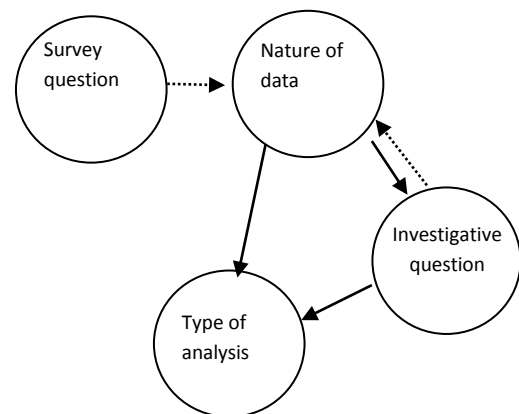
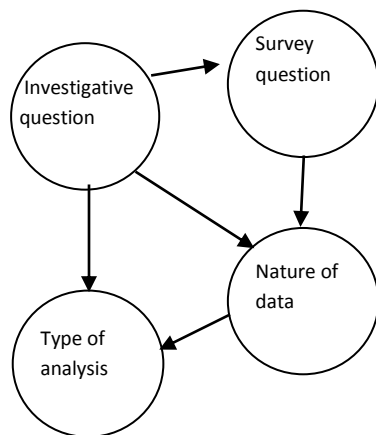
### **2.7.2. Types of investigative questions**

At the school level there are three basic types of investigative questions that are posed: “Typical questions that may inspire the need for data collection may be of the form, ‘How big is A?’ or ‘Is A bigger than B (and by how much)?’ or ‘How is X related to Y?’” (Graham, 2006, p. 4). These types of investigative questions are called summary, comparison and relationship questions, respectively. Summary investigative questions are posed when a description of the data is needed and are usually about a single data set; comparison investigative questions are posed for comparing two (or more) subsets of data – for example, male and female, young and old – across a common variable such as reaction time; and relationship investigative questions are posed for looking at the interrelationship between two paired variables. (Graham, 2006; Pfannkuch & Horring, 2005). Table 2-3 (next page) summarises the ideas discussed so far.

**Table 2-3. The link between the type of investigative question and the nature of the data, school level**

| Summary     |          |            | Comparison                         | Relationship       |                 |                   |
|-------------|----------|------------|------------------------------------|--------------------|-----------------|-------------------|
| Univariate  |          |            | Bivariate                          | Bivariate          |                 |                   |
| Categorical | Discrete | Continuous | Combination of any two, not paired | Paired categorical | Paired discrete | Paired continuous |

In a statistical investigation using primary data, the type of investigative question influences the nature of the data collected (Figure 2-6a). However, when a statistical investigation uses secondary data, the nature of the data influences the type of investigative question posed (Figure 2-6b). Both the type of investigative question posed and the nature of the data have an impact on subsequent analysis, especially the types of graphs drawn and the statistics calculated.

**Figure 2-6. (a) Flow of influence primary data****(b) Flow of influence secondary data**

*Note:* Solid arrows indicates direct influence; dotted arrows indicate possible influence.

Pfannkuch and Horing (2005), in their research on problematic areas in the statistics curriculum, reported on the overall structure built during teaching (Figure 2-7, next page). This included reference to the questions asked and how these were linked to different graphs and statistics.



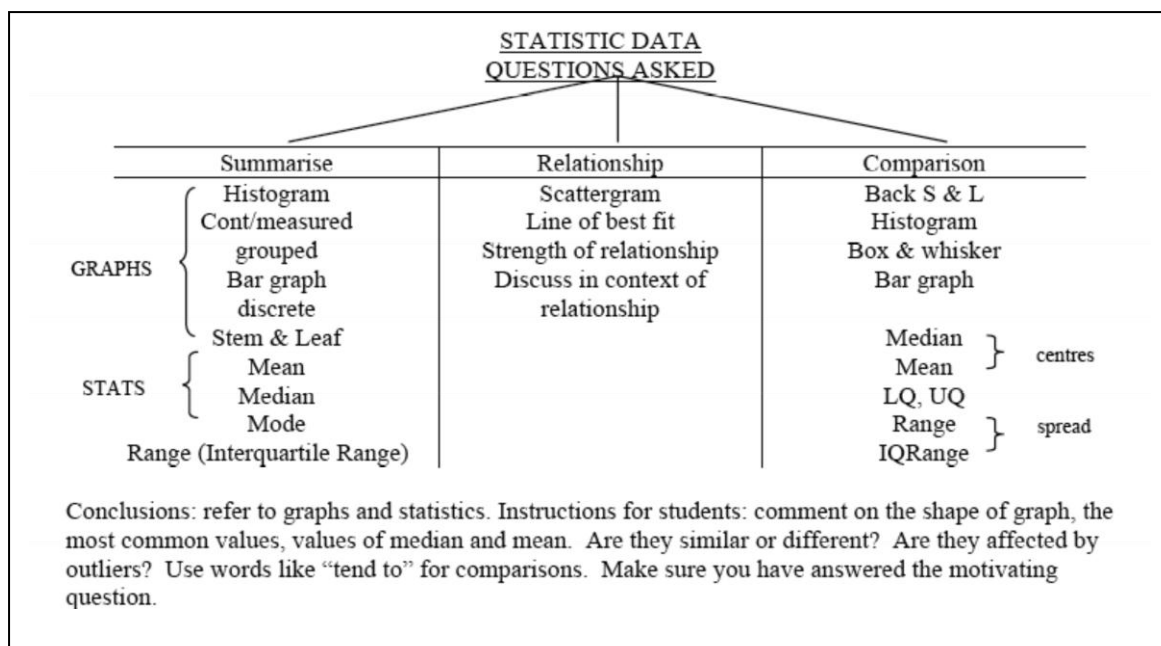


**Figure 2-7. Overall structure built during teaching**

*Note:* Reprinted from “Developing statistical thinking in a secondary school: A collaborative curriculum development,” by M. Pfannkuch and J. Horring, 2005, *Curricular development in statistics education: International Association for Statistical Education 2004 Round Table*, p. 210. Copyright 2005 by the International Statistical Institute. Reprinted with permission.

Graham (2006) also offers some advice as to the type of analysis that would be expected for each of the types of investigations – summary, comparison and relationship.

Investigations of the describing type typically involve collecting a single data set and summarising it in some way (in words, numbers or graphically), in order to discover its main features. Comparing investigations usually involve collecting two data sets of the same type and then comparing them, either graphically or by means of summary statistics. With an interrelating type of investigation, learners are looking at the relationship between two variables that measure different things. The data are therefore paired (and it follows that the two lists of numbers contain the same number of values). Typically the analytical tools required fall into the general statistical areas of regression and correlation. Knowing the form of enquiry should help them to select suitable analytical tools at the A [analysis] stage of their investigation. (pp. 223–224)

### 2.7.3. Summary – investigative questions and nature of data

I would conjecture that the complex classifications and use of language related to the nature of data, as raised in this section, leads to potential problems for both students and teachers alike. Without a clear understanding of the nature of the variables, students can pose unsuitable investigative questions. With secondary data, students should carefully unpack what each variable is and how the data were collected. They should decide on the nature of

the variable, i.e. whether it is categorical, discrete numeric or continuous numeric. These distinctions can help especially when it comes to relationship-type questions. However, it is not uncommon for students, teachers and even statisticians to pose investigative questions such as: “I wonder if reaction time is related to gender?” (Wild, 2007) or “Is there a relationship between gender and height or ethnicity and height?” While these investigative questions are valid questions and can be explored, students may think that these are relationship-type investigative questions (as they have the word relationship in them), when in fact they are comparison-type investigative questions. The idea of what a relationship is in a statistical sense is worth discussing with students (Burgess, 2007).

Posing summary, comparison and relationship questions requires students and teachers to have a clear idea of what the variable(s) are that they are interested in, what they want to do (summarise, compare, relate), and what the population of interest is. Once these factors are sorted out, decisions about the nature of data and analysis follow, but students should have some clarity about what will be involved given the investigative question that was posed.

### **2.8. Summary**

The statistical enquiry cycle is a major part of the statistics curriculum in New Zealand. Posing questions from given data is a requirement in the assessment for qualifications (NCEA) in year 11 (ages 15–16). Posing investigative questions is at the heart of the statistical enquiry cycle.

In this review I have argued that within the enquiry cycle there are two distinct purposes for posing questions, and hence two types of questions that might be posed: the investigative question and the survey question. There are also two purposes for asking questions, and hence two types of questions that might be asked: interrogative questions and analysis questions. There is limited research where the distinction between the two purposes for posing questions is made clearly, the majority of this being recent research from New Zealand. While there is more research around survey questions, survey questions are not always clearly defined as being something different from the statistical question (investigative).

At the schooling level the investigative question broadly falls into one of three types: either it can be a summary, comparison or relationship question. Survey questions or other data collection methods are used to collect the data. The nature of the data can be qualitative

(categorical or ordinal) or quantitative (discrete or continuous). Summary situations work with univariate data, whereas comparison and relationship situations work primarily with bivariate data. The investigative question and the nature of the data have a direct bearing on the type of analysis that is undertaken with the data. The link between these two is clear and the influence of one on the other varies depending on whether primary or secondary data is used.

All stages of the enquiry cycle need to be interrogated thoroughly, including a looking forwards and backwards process. At the analysis stage there are questions that should be asked to help describe and infer from the data. While initially the teacher might drive the question asking, ultimately students need to be asking the questions themselves as they undertake statistical investigations and develop their statistician's craft.

There are a number of research areas that the literature review has highlighted as being shallow in their depth of coverage. These include the investigative question, interrogating the enquiry cycle, students asking analysis questions, the link between the investigative question and the analysis, and the link between the investigative question and the conclusion. The literature review has also highlighted that there is no consensus on definitions and classifications of the different types of data.

## **Chapter 3. Research Methods**

### **3.1. Introduction**

This chapter details the theoretical basis and research methods used in this thesis. It starts with a theoretical discussion of the paradigms that underpin the research and of the researcher's paradigmatic position, values, beliefs and assumptions. This is followed by a discussion on the design research method, the data collection methods used, and the data analysis. The chapter finishes with short discussions on ethics, triangulation, validity and reliability.

### **3.2. Theoretical discussion**

The researcher brings to the study her values, beliefs and assumptions, and these influence and even drive decisions around research design and methodology. In this section connections are made to the paradigm and theoretical basis that support my values, beliefs, and assumptions.

#### **3.2.1. Paradigms**

A paradigm is a basic set of beliefs about the world, based on ontological, epistemological and methodological assumptions (Cohen & Manion, 1994; Ernest, 1998; Fontana & Frey, 1994; Guba & Lincoln, 1994; McNiff & Whitehead, 2006; Punch, 2009; Schram, 2003; Sikes, 2004; Silverman, 2000). Guba and Lincoln (1994) more formally define a paradigm as:

A paradigm may be viewed as a set of basic beliefs (or metaphysics) that deals with ultimates or first principles. It represents a worldview that defines, for its holder, the nature of the "world," the individual's place in it, and the range of possible relationships to that world and its parts, as, for example, cosmologies and theologies do. The beliefs are basic in the sense that they must be accepted simply on faith (however well argued); there is no way to establish their ultimate truthfulness. (p. 107)

Ontological assumptions look at what is the nature of reality; epistemological assumptions align with what constitutes knowledge, "How do we come to 'know' the world?" (Schram, 2003, p. 29), and the relationship between researcher and those researched; and methodological assumptions deal with appropriate methods for generating and analysing data

(Ernest, 1998; Guba & Lincoln, 1994; Punch, 2009; Schram, 2003; Sikes, 2004; Silverman, 2000).

### **3.2.2. My paradigmatic position**

Positioning oneself in relation to a paradigm is not a clear-cut task as aspects of a number of paradigms speak to underlying values, beliefs and assumptions. In part it is about gut instinct and placing a line in the sand and saying, “I stand here.” Primarily I feel comfortable with the “interpretative, naturalistic, subjective, qualitative paradigm” (Sikes, 2004, p. 18) and its link with constructivism. Interpretive research is that which is based in natural settings where researchers offer descriptions and explanations of what they observe. The data tends to be qualitative and there is a bantering back and forth often between the observed and the researcher as the story is developed (McNiff & Whitehead, 2006).

Constructivism is based in locally constructed realities, i.e. participants construct their own meanings through experiential and social interactions, a paradigm which fundamentally connects with my beliefs about teaching and learning. It makes sense that my set of beliefs about research should have a direct link to my beliefs about teaching and learning. I am first and foremost a teacher and this drives my research. It would be fair to say that constructivism has been a subtle influence in the nature and type of research undertaken.

Research is about getting and communicating knowledge and using this knowledge to inform practice, to inform policy, and to inform future developments. Research has the potential to improve things in some way and this is ultimately the goal (Sikes, 2004).

### **3.2.3. Values, beliefs, and assumptions**

Over the last 25 years working in education, my fundamental values and beliefs about teaching and learning have been influenced heavily, almost from the beginning, by the work of the EQUALS group based in the Lawrence Hall of Science, University of Berkeley, California (Afflack, 1982; Downie, Slesnick, & Stenmark, 1981; EQUALS & California Mathematics Council, 1989). This group of women and men have worked endlessly to address equity issues in science and mathematics. Much of the work they have done has been built around using cooperative learning, having students working in pairs and groups, and the use of hands-on concrete materials. These foundational ideas influence any preparation of

material for use with students. This cooperative group approach has been shown as beneficial for both Māori and Pasifika students (Gardiner & Parata, 2007).

The idea of using concrete materials and building through to abstract ideas developed further following exposure, through involvement with the numeracy development project in New Zealand, to work by Quinlan and his colleagues (Quinlan, Low, Sawyer, & White, 1993) at the Australian Catholic University and the Pirie-Kieran materials-imaging-abstract model (Pirie & Kieran, 1989; Pirie & Kieran, 1994).

More recent influences include Martignon's (2008) EIS-T model, which is based on the work of Bruner (1960). I was introduced to the EIS-T model at Martignon's presentation at the 2008 International Association for Statistical Education (IASE) round table. This proved to be an excellent model to use in activity design as it was cognisant of the big part that technology (T) has to play in statistics education. It also includes an en-active process (E), where students need to be enacting the ideas; use of iconic (I) representations (for example, data cards represent people); and symbolic (S) representations of ideas. All of these approaches and ideas have had a direct influence on the teaching and learning materials that were developed for the research and on the approach to the research itself.

### **3.3. Design research**

Design research involves an intervention, researching the intervention, and developing theories based on findings from the intervention. Design research is iterative with subsequent cycles building on previous cycles. Design research is suitable for the study of innovative practice related to posing investigative questions. This section details what design research is, and how its main phases link to the area of research for this thesis.

Design research is a common approach to statistics education research; for example, both Bakker (2004a) and Makar (2004) used design research methods in their doctoral theses. In New Zealand it has been used in the Teaching and Learning Research Initiative (TLRI) project which is based in statistics education (Pfannkuch, Arnold, & Wild, 2011).

#### **3.3.1. What is design research?**

Design research, or design experiments, has its foundations in design science (Brown, 1992) and typically involves a planned intervention that develops ideas based on theoretically grounded innovations to inform practice while simultaneously conducting research on the

intervention (Brown, 1992; Cobb, 2000a, 2000b; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Kelly & Lesh, 2000). In particular, design research looks at the types of learning that differ from common or current practice and explores new and novel practices with the intent to change systems by being innovative (Bakker, 2004a; Cobb et al., 2003; Hjalmarson & Lesh, 2008; Schwartz et al., 2008).

A design experiment is a form of interventionist research that creates and evaluates novel conditions for learning. The desired outcomes include new possibilities for educational practice and new insights on the process of learning. Design experiments differ from most educational research because they do not study what exists; they study what could be. (Schwartz et al., 2008, p. 47)

Design research involves the development of instructional materials to engineer learning, developing theories based on studying this learning, and feeding back into the instructional materials and the developing theories, thereby becoming a cyclical process (Bakker, 2004a; Brown, 1992; Cobb, 2000a, 2000b; Cobb & Gravemeijer, 2008; Gravemeijer, 1998; Hjalmarson & Lesh, 2008; Lesh & Kelly, 2000; Roth, 2005; Steffe & Thompson, 2000). The theoretical models grow out of a repeating empirical process (Bakker, 2004a; Brown, 1992; Gravemeijer, 1998; Roth, 2005), and most design research activities take place over an extended period of time (Brown, 1992; Cobb, 2000a; Confrey & Lachance, 2000; Lesh & Kelly, 2000).

### **3.3.2. The phases of design research**

Design research's iterative or cyclic process has a number of phases. These include an initial preparation and design phase, followed by a teaching experiment, then a retrospective analysis phase (Bakker, 2004a; Cobb, 2000a; Cobb & Gravemeijer, 2008; Steffe & Thompson, 2000; van Nes & Doorman, 2010). The retrospective analysis phase often leads to a new problematic situation which informs a new cycle of preparation and design, teaching and retrospective analysis (Hjalmarson & Lesh, 2008). These phases are interconnected with and by the hypothetical learning trajectory (see next section).

Figure 3-1 (next page) shows diagrammatically how the phases, the problematic situation, and the hypothetical learning trajectory intertwine. The dashed arrow of the hypothetical learning trajectory is to indicate that it is not a direct linear process. Each of these phases and the hypothetical learning trajectory are now discussed.

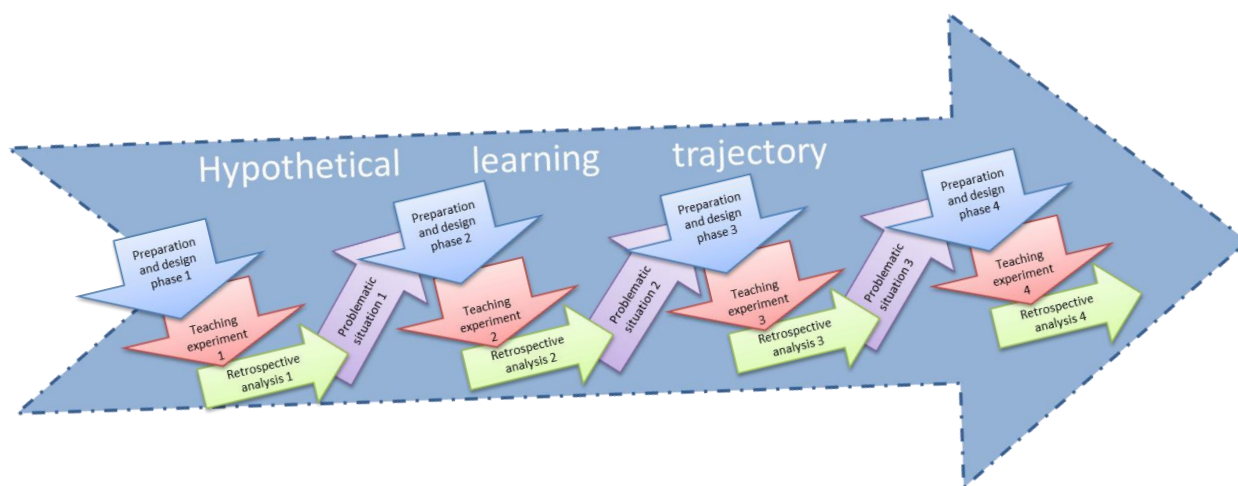


**Figure 3-1. Design research phases intertwined with hypothetical learning trajectory**

### ***The hypothetical learning trajectory***

The hypothetical learning trajectory (HLT) is a “best guess” plan for teaching and learning. It involves defining a learning goal, considering possible learning activities and the types of student thinking and understanding they might evoke, and the hypothetical learning process (Simon, 1995). The HLT systemises and extends what good teachers do, with the difference being that within the design research context it is a deliberate act: the teacher and researcher are actively and consciously planning, reflecting and recording these actions and thoughts.

The HLT is the connection between the instruction theory and the in-class teaching experiment. It is informed by general domain-specific and conjectured instruction theories. It guides and supports the design of instructional materials and is developed during the design phase. The HLT acts as a guide in teaching, interviewing and observing. It also helps to structure the analysis and guides the development of an instruction theory (Bakker, 2004a; Simon, 1995; van Nes & Doorman, 2010).

Simon (1995) places the HLT within the mathematics teaching cycle. Figure 3-2 (next page) shows the relationships between the various domains of teacher knowledge, the HLT and student interactions, and has been modified from Simon’s mathematics version (p. 137) to reflect this research which is statistics based.



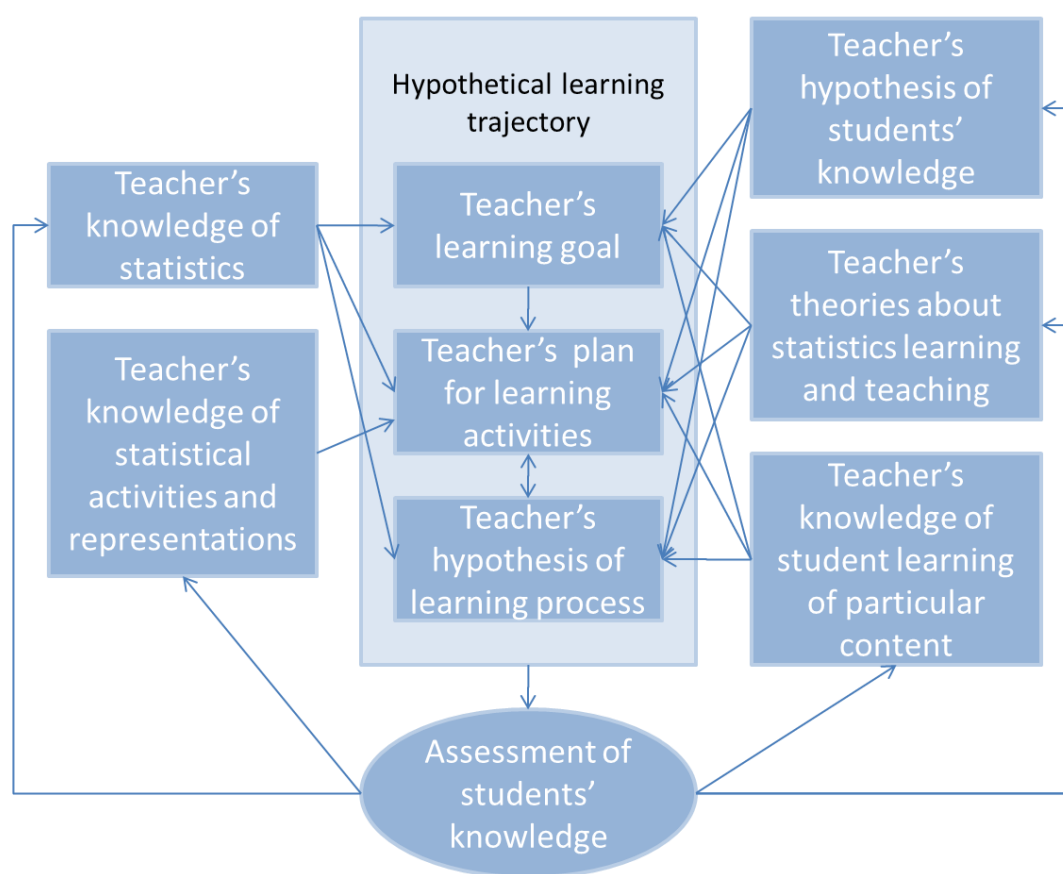


**Figure 3-2. Statistics teaching cycle**

*Note:* Reprinted from “Reconstructing mathematics pedagogy from a constructivist perspective,” by M. Simon, 1995, *Journal for Research in Mathematics Education*, 26(2), p. 137. Copyright 1995 by the National Council of Teachers of Mathematics. Reprinted with permission.

In its simplest form the HLT begins with defining the learning goal. The teacher and researcher together develop a hypothesis of students’ understanding. This initial hypothesis is based on information from a wide range of sources and experiences; for example, current students’ experiences in a related area, the experiences of a similar group of students, information that has come to light from pre-testing, and data and information from research. At this point the teacher and researcher are working with a best guess at how the learning might proceed, thereby establishing an initial goal and plan for instruction. Further information becomes available as the research progresses and the initial hypothesis is able to be improved.

The teacher and researcher consider possible learning activities and the types of thinking and learning these activities might provoke. The teacher and researcher’s hypotheses of students’ statistics knowledge, their theories about statistics teaching and learning, their knowledge of

learning in the statistics context, and their knowledge of statistics activities and representations all intersect and come into play when considering possible learning activities (Simon, 1995). Additionally, other influencing factors in New Zealand education are, for example, *Ka Hikitia* (Ministry of Education, 2009a) and the *Pasifika Education Plan* (Ministry of Education, 2009b), and the implications of these documents in classes with high Māori and Pasifika student numbers (as is the case for both of the schools where the research was based). Beliefs and interests also play a role in the development of learning activities; for example, the researcher's beliefs and interest in EQUALS (Afflack, 1982; Downie et al., 1981; EQUALS & California Mathematics Council, 1989), and the use of cooperative and group-based activities. In addition the researcher's education theories contribute to deciding on and creating learning activities. For example, when developing activities the Pirie-Kieran model (Pirie & Kieran, 1989; Pirie & Kieran, 1994) from the New Zealand numeracy project (Ministry of Education, 2005a, 2005b) with its theory of materials, imaging, and abstracting/generalising, and the EIS-T model (Bruner, 1960; Martignon, 2008) with its en-active, iconic, symbolic, technology components were considered. These hypotheses, theories and knowledge collectively underpinned the decisions that were made about which activities to use and which ones not to use, as well as what activities needed to be developed and what they would look like.

The hypothetical learning process is “a prediction of how students' thinking and understanding will evolve in the context of the learning activities” (Simon, 1995, p. 136). This is a best guess at what will happen. The hypothetical learning process is continually modified as time progresses. This is a result of the teacher and researcher developing a broader understanding of students' conceptions in the area through a process of reflection based on interactions with and observations of students. The teacher's and researcher's thinking is modified as they make sense of what is happening in the classroom.

The working backwards and forwards between all parts of the HLT is critical. Reflection, based on assessment of students' thinking, leads to constant adjustment and fine-tuning of the HLT, the goal, the activities and the hypothetical learning process (Simon, 1995).

### ***Preparation and design phase***

Each cycle in design research starts with an initial preparation and design phase in which members of the research team work together to plan and prepare for the teaching experiment. The HLT is developed and informs the design of instructional materials (Bakker, 2004a). The

initial preparation and design phase has elements of conjecture, invention, innovation, bricolage and flexibility.

During the preparation and design phase the research team is clarifying instructional goals, considering the content area of the curriculum, what students need to cover, and the central objectives and principles of statistics teaching (Cobb & Gravemeijer, 2008; Confrey & Lachance, 2000; Roth, 2005; Steffe & Thompson, 2000; Wittman, 1998). Conjectures begin to emerge while considering these practical issues and the conjectures start to determine the choice, sequence and duration of what is presented as well as the types of activities. These conjectures need to be situated within a theoretical context and should be aiming to shift perspectives (Cobb, 2000a; Confrey & Lachance, 2000).

Design research involves simultaneously designing new and innovative teaching and learning materials and researching the impact of these materials. The idea of invention and innovation is fundamental to design research. Innovative ideas in substantial teaching units and novel lessons need to be invented, but activities that are invented are done so because of their potential towards the HLT (Bakker, 2004a; Schwartz et al., 2008; Wittman, 1998). This process is iterative – design, test, and redesign. It is about reconceptualising ways to approach the content and the pedagogy and it is about thought experiments, envisioning what the teaching and learning might look like (Confrey & Lachance, 2000; Gravemeijer, 1998). Gravemeijer's (1998) idea of bricolage (collecting activities that could be useful, and using what happens to be available) adds to the broader mix of activities that have the potential to contribute towards the end goal.

The HLT needs to be flexible and easily adapted because it potentially may need to be adapted and changed at any time based on feedback and reflections of how students have understood, attempted or interacted with an activity or sequence of activities. The plan needs to tread the fine line between having sufficient detail and remaining flexible (Roth, 2005).

### ***Teaching experiment***

In the teaching experiment phase the teacher and researcher together experience the students' learning and reasoning in the classroom. The teacher and researcher develop a unit of work that is based on what they think is their best bet on possible outcomes for students. The HLT is considered at all stages of the process. Each lesson is reflected on and feeds into the next

lesson. The desired outcome for the wider sector from design research is the development of prototypical instructional materials that can be used by others in their own settings.

During the teaching experiment phase evidence is collected for retrospective analysis. This can include, but is not limited to, field observations, interviews, pre- and post-tests, and student class work. Interview questions are motivated by the HLT and interview guides can be fine-tuned following initial analysis of student responses in, for example, their pre- and post-tests. Regular classrooms are ideally used and the researcher attends all lessons in the teaching experiment.

### ***Retrospective analysis***

Retrospective analysis happens at two levels in design research. There is the ongoing reflection and analysis that happens after each teaching episode and then there is the overall reflection and analysis that happens at the end of the teaching experiment. The ongoing retrospective analysis informs the planning of the next instructional activity and is motivated by what is best for these students (Cobb, 2000a; Roth, 2005).

The retrospective analysis that happens at the end of a teaching experiment allows the researcher to reflect on a number of fronts. For example: How does what has happened relate to conjectures made? How does what the students have actually learned compare with the HLT? and Does the evidence confirm expectations? (Bakker, 2004a; Cobb, 2000a; Schwartz et al., 2008). As classroom events are located within the broader theoretical context there is also a need to consider the following: In what ways do classroom events exemplify the HLT or conjectures? Are they generalisable? Repeatable? Trustworthy? (Cobb, 2000a; Cobb & Gravemeijer, 2008). The HLT and conjectures orientate and focus the retrospective analysis and provide a basis for developing the instruction theory (Bakker, 2004a; Cobb, 2000a).

It is during the retrospective analysis that often new problematic situations arise. These highlight areas for further development.

### ***Problematic situation***

A problematic situation is the motivation for the research and the design of prototypical instructional materials, and it lies within the context being investigated (Hjalmarson & Lesh, 2008). As the planning and preparation phase starts and flows through to the teaching

experiment and retrospective analysis, new problems can emerge. The cycle begins again, adapting as the new problematic situation is acknowledged and incorporated into the design.

### **3.3.3. Supporting methods**

Grounded theory is theory built based on iterative cycles of data collection and analysis. Action research brings together the acting and the researching. Grounded theory and action research methodology provide structural support for the main research method – design research.

#### ***Grounded theory***

The area of interest for this thesis, posing statistical investigative questions, was shown through an initial review of the literature to be lacking any clear description or explanation. There was no instruction or learning theory for posing statistical investigative questions; hence there was a need to develop an instruction or learning theory, from the ground up and systematically from data collected through a number of iterations. An opportunity existed for theory generation built out of cycles of data collection and analysis – in other words, a theory generated from data and grounded in the data.

Grounded theory is a research strategy whose purpose is to generate theory from data. “Grounded” means that the theory will be generated on the basis of data; the theory will therefore be grounded in data. “Theory” means that the objective of collecting and analysing the research data is to generate theory to explain the data. The essential idea in grounded theory is that theory will be developed inductively from data. Grounded theory, then, is an overall strategy for doing research. (Punch, 2009, p. 130)

Grounded theory is an iterative process that starts with research questions and as open a mind as possible and through a series of data collection and data analysis cycles moves towards a theory. Subsequent cycles are guided by emerging theories in the analysis and continue until new data no longer shows new theoretical ideas – instead the data confirms theories already established through the research (Kieran, 1998; Punch, 2009; Schram, 2003; Smith & Davies, 2010; Strauss & Corbin, 1994). Grounded theories are always traceable to the data that gave rise to them and they are fluid and provisional (Schram, 2003; Strauss & Corbin, 1994).

Schram (2003) describes three key approaches to grounded theory research: systematic, emerging and constructivist. The systematic approach is more prescriptive with the coding of categories and subcategories, and visual diagram development to present the theory being central; the emerging approach is more flexible and looks at connecting categories and the

emerging theory; while the constructivist approach places a “more subjective emphasis on feelings, assumptions and meaning making of study participants” (p. 73).

Underpinning the grounded theory approach are three orienting concepts (Schram, 2003). *Theory* looks at the relationships, plausible and provisional, that are suggested between and within concepts and sets of concepts. Continued research can strengthen suggested theories and lead towards developing substantive theory that stays connected to the data. The developing theory contains multiple conceptual relationships, i.e. it is *conceptually dense*. In the process of analysis and theory development the process of *constant comparison* is employed. Researchers are constantly challenging the data by asking what is going on here and how is this similar or different (Schram, 2003; Strauss & Corbin, 1994).

In a grounded theory approach data are systematically collected and analysed, with subsequent cycles building on emerging theories from the analysis. This process of theoretical sampling continues until theoretical saturation is achieved. As subsequent cycles confirm theories rather than further developing them, the process grinds to a natural conclusion (Punch, 2009; Schram, 2003).

The guiding premises of grounded theory, i.e. creating a conceptually dense theory based on iterative cycles of data collection and analysis, with constant comparison, provide an overarching construct to provide a basis for this research into students’ posing investigative questions. Time did not allow for theoretical saturation in all the problematic areas researched in this thesis; however, the guiding premises provide a research strategy base.

### ***Action research***

Action research is concerned with both action and research/evaluation, a process of trying out something different (an intervention) and examining carefully what happens (Banister, Burman, Parker, Taylor, & Tindall, 1994; Cohen & Manion, 1994; Fernie & Smith, 2010; McNiff & Whitehead, 2006; Punch, 2009). Action research is predominantly based around people taking action to change something they are doing and concurrently researching or evaluating the action and results (Fernie & Smith, 2010; McNiff & Whitehead, 2006). Action research can be used to improve practice, develop theoretical knowledge, and develop new methods of learning (Cohen & Manion, 1994).

The action research cycle – plan, act, observe and reflect (Banister et al., 1994; Fernie & Smith, 2010; Kemmis & McTaggart, 2005) – shows alignment with the design research

process: plan – take stock and identify a concern or issue to address; act – think of a way forward to address and improve the situation and then try it out; observe – use different methods to collect data to show what is happening and to evaluate the success of the action; and reflect – evaluate and reflect on the evidence, the processes and the consequences. The cycle then repeats itself. In this way action research is a process aligned with design research. The difference in design research is that the whole cycle sits within the HLT and is more deliberately focused on designing new and innovative ways of teaching and learning that show potential towards the HLT.

Action research typically involves practitioners and researchers collaborating together. This is usually referred to as participatory action research (Ferne & Smith, 2010; McNiff & Whitehead, 2006; Punch, 2009; Slavin, 2007). Regardless of whether the action research involves the teacher being the action researcher or the teacher and researcher working together, one question needs addressing with care: “Does the end justify the means?” (Banister et al., 1994, p. 112). Ethically one needs to consider the consequences arising from the actions because it is difficult, having changed social practice, to reset it (Banister et al., 1994; Ferne & Smith, 2010).

### **3.4. Data collection methods**

Marshall and Rossman (2006) and Punch (2009) describe four ways of collecting qualitative data. These are through participant observation, direct observation, interviews and analysis of documents. Quantitative data can be collected through experiments, quasi-experiments and correlational surveys (Punch, 2009).

Research data in this thesis has been collected using multiple methods (Confrey & Lachance, 2000; Silverman, 2000). The main methods of data collection are questionnaire/survey (pre- and post-tests), observations (video, field notes, artefacts of student work), and interviews (including reflective discussions). These different methods have generated both qualitative and quantitative data. By definition, therefore, the data collection methodology can be described as mixed methods research, which is “empirical research that involves the collection and analysis of both qualitative and quantitative data” (Punch, 2009, p. 288).

In addition, combining data collection methods allows for limitations in one method to be offset by the strengths in another method (Marshall & Rossman, 2006). Combining

quantitative and qualitative methods also allows for triangulation, the use of multiple methods to study a problem (Patton, 1990; Punch, 2009).

### **3.4.1. Pre- and post-tests**

Pre- and post-tests allow researchers to investigate the progress of participants in innovative curriculum programmes. However, they can also influence what is accomplished and are not always able to reflect the project's desired outcomes as it might not be possible to specify these at the outset (Lesh & Kelly, 2000). Ellerton and Clements (1997) reported on externally set pencil-and-paper mathematics tests of the short-answer variety and found that there was as high as 30% “mismatches”, a mismatch being where correct student test responses on the written test could correspond to incorrect reasoning in an interview, or partial or incorrect responses could correspond to correct reasoning in an interview. They concluded that what students know cannot be accurately summarised from even an expertly constructed pencil-and-paper test and suggested using an interview as part of the data analysis process.

Questionnaires provide further insight into pre- and post-tests. Questionnaires are a widely used procedure for data collection and can be used to obtain information relatively quickly. However, the design of a questionnaire is not as straightforward as might be thought, and relative care is needed to ensure that the questions are clear and unambiguous, the layout is clear and allows for easy reading of the questions, sufficient space is given to answer the questions, and the question order has been carefully considered (Cohen, Morrison, & Manion, 2007; Cozby, 2009; Crano & Brewer, 2002; Fowler & Cosenza, 2009; Gay, Mills, & Airasian, 2012; Opie, 2004).

Pre- and post-tests can be used to compare groups of students using a control group matched to the experimental group. Typically this type of quasi-experiment design has a pre-test for both the experimental group and the control group, some intervention for the experimental group but not for the control group, followed by a post-test for both groups (Cohen et al., 2007; Langbein & Felbinger, 2006; Mark & Reichardt, 2009; Slavin, 2007). In a quasi-experiment the control group has not been selected by randomisation, hence the term *quasi-experiment* – like an experiment but the participants have not been randomly assigned to a condition. Quasi-experimental design also includes the one group pre-test-post-test design where there is not a control group. This type of quasi-experimental design suggests a number of methodological concerns (Cohen et al., 2007; Langbein & Felbinger, 2006; Mark & Reichardt, 2009; Marsden & Torgerson, 2012; Slavin, 2007); for example, history,



maturation, test effects, instrumentation, attrition and regression towards the mean. Some of these concerns, like history and maturation, may not be as big an issue in studies with short time frames (Mark & Reichardt, 2009).

### **3.4.2. Observations**

“Observation entails the systematic noting and recording of events, behaviours and artefacts (objects) in the social setting chosen for the study” (Marshall & Rossman, 2006, p. 98). Observations in a social setting can be structured or unstructured. Structured observations tend to use checklists and observation schedules and generally generate quantitative data; unstructured observations allow for a more holistic analysis, working towards a theoretical saturation (Marshall & Rossman, 2006; Punch, 2009; Smith & Bowers-Brown, 2010).

The researcher takes on one of the following roles as an observer/participant when doing observations: complete observer, observer as participant, participant as observer, or complete participant. The level of participation influences the balance of objectivity and subjectivity in the research. For example, for the observer as participant, the researcher interacts with subjects by prompting for answers to issues, so interacting with subjects but not taking a formal role in the group. Similarly, for the participant as observer, the researcher engages more in the day-to-day activity of the classroom; for example, as a part-time teacher (Opie, 2004; Punch, 2009). Regardless of the level of participation versus observation, it is difficult to “avoid having an effect on the social phenomena being studied” (Banister et al., 1994, p. 36).

Field notes, videos and photographs are common data collection methods used in observations (Banister et al., 1994; Opie, 2004; Roth, 2005). It is easy to collect too much data and care needs to be taken not to do this; for example, by following one group rather than trying to cover multiple groups (Confrey & Lachance, 2000). Field notes, in particular, can be used to record thoughts and reflections and to track thinking (Confrey & Lachance, 2000; Roth, 2005; Smith & Bowers-Brown, 2010).

Observations supplement other methods of data collection and are seen as a “fundamental and highly important method” (Marshall & Rossman, 2006, p. 99).

### ***Advantages and disadvantages of observations***

Advantages of observations include allowing for data to be collected in natural settings and for complex interactions to be documented with contextual information. Observations are also good for triangulation with other data collection methods. Disadvantages of observations include difficulty in replicating, possibility of ethical dilemmas, and the data being more affected by researcher presence (Marshall & Rossman, 2006).

### **3.4.3. Interviews**

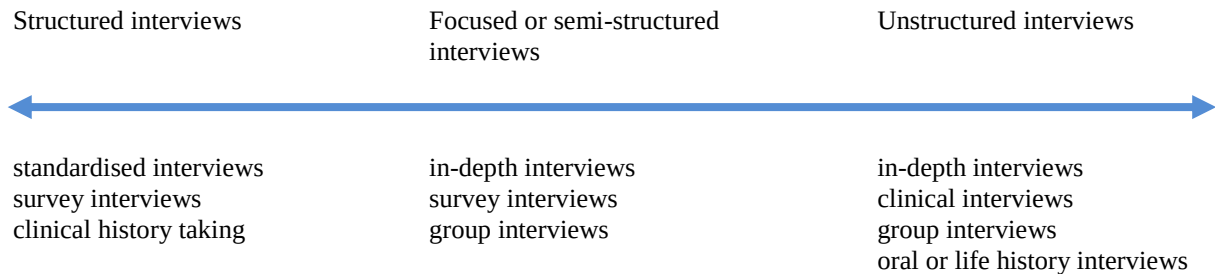
We interview people to find out from them those things which we cannot directly observe. (Patton, 1990, p. 278)

Interviews work well in mixed methods research. They allow researchers to explore issues that are too complex to do through quantitative means, and to investigate and prompt for ideas and thoughts that cannot be observed through participant observations. They can be used for triangulation or in conjunction with other methods. Interviews are a tool of great flexibility and can easily be adapted to a wide range of research situations (Banister et al., 1994; Cohen & Manion, 1994; Fontana & Frey, 1994; Fontana & Frey, 2005; Gay et al., 2012; Punch, 2009; Wellington, 2000). Interviews can be used as a principal means of gathering data, to test hypotheses or suggest new ones, and to clarify meanings. They are a powerful and common way to understand fellow human beings, and they are an explanatory device to help identify variables and relationships in research situations (Cohen & Manion, 1994; Fontana & Frey, 2005; Walford, 2001; Wellington, 2000). Those who caution about the use of interviews as a single source of data, also recommend that data about the same topic could be generated in a variety of ways supporting the mixed methods approach (Walford, 2001).

Interviews run the gamut from completely structured, where the specific words to say at every step are given, through to completely unstructured, where a starter question or issue is used and the interview meanders along following the direction of the interviewee's thinking (Banister et al., 1994; Cohen & Manion, 1994; Silverman, 2000; Wellington, 2000). Different authors classify interviews in different ways; for example, Fontana and Frey (1994, 2005), Opie (2004) and Punch (2009) classify interviews as structured, semi-structured and unstructured; Wellington (2000) as structured, unstructured and non-standardised; Patton (1990) as informal conversational, standardised open-ended, and closed fixed-response; and Cohen & Manion (1994) as formal, less formal, completely informal, and non-directive.

Punch (2009) uses two key criteria when classifying interviews: “degree of structure in the interview, and how deep the interview tries to go” (p. 145).

Punch (2009, p. 145) uses the following continuum model for interviews:



The next sections give an overview of the three broad categories of interviews – structured, semi-structured and unstructured – using Punch (2009), Fontana and Frey (1994; 2005) and Opie’s (2004) classification.

### ***Structured interviews***

Structured formal interviews are not dissimilar to completing a questionnaire, with the major difference being that the questions are asked of the participant rather than being read by them, and the responses are noted by the interviewer rather than written by the participant (Opie, 2004; Wellington, 2000). Structured interviews have pre-set questions, which typically have pre-set responses or simple short answers (Cohen & Manion, 1994; Fontana & Frey, 1994, 2005; Gay et al., 2012; Punch, 2009). Structured interviews are one way to collect a lot of data. There is an element of formality which leads to results often being used to make generalisations (Opie, 2004).

### ***Semi-structured interviews***

The semi-structured interview is less formal than the structured interview. It is a compromise between the structured and unstructured interviews (Cohen & Manion, 1994; Wellington, 2000). When preparing for a semi-structured interview, an interview guide or checklist is created from which the interviewer is given flexibility over the range and order of the questions within the framework given (Cohen & Manion, 1994; Davies, 2010; Opie, 2004; Smith & Bowers-Brown, 2010; Wellington, 2000). Advantages include longer responses and more latitude for responses. Disadvantages include the possibility of researcher bias and the analysis is more complex (Opie, 2004).

As a basis for interviewing, the semi-structured situation provides a reasonable fit with the research that is the focus of this thesis. Two specific situations related to semi-structured interviews are now discussed as they form the foundation of the types of interviews used in the data collection phase.

**Focused interviews:** Focused interviews are where prior analysis has been conducted before the interview (Cohen & Manion, 1994; Merton, Fiske, & Kendall, 1956/1990). According to Merton, Fiske and Kendall (1956/1990), the focused interview differs from other types of research interviews in the following respects: the participants have previously been involved in a particular situation that is known to the interviewer; specific aspects of the situation have been analysed by the researcher and hypotheses made, the aspects being ones which are deemed significant; the analysis provides the foundations for the interview guide, identifying the most relevant areas of inquiry; and finally the interview is focused on the subjective experiences of the participant who was involved in the particular situation, the participant's responses enabling the researcher to examine the strength or validity of her hypotheses and lead the way to possible fresh hypotheses (Cohen & Manion, 1994; Merton et al., 1956/1990).

The feature that distinguishes focused interviews is that prior analysis of the situation (for example, pre- and post-test) in which the students/participants have been involved has been done by the researcher/interviewer (Cohen & Manion, 1994; Merton et al., 1956/1990).

**Group interviews:** A group interview is where the interviewer works with several people simultaneously, either in a formal or informal setting (Fontana & Frey, 2005; Punch, 2009). The group interview can be seen as a more secure environment, one where participants can be at ease amongst their peers. In this situation they can prompt and jog one another's memories, and this often leads to richer and deeper discussions (Merton et al., 1956/1990; Punch, 2009; Wellington, 2000). The interviewer is more of a conductor or facilitator of the proceedings, encouraging participation from all members of the group. Regardless of the perceived role, the interviewer will be directing the interaction by supplying topics, questions or prompts (Fontana & Frey, 1994, 2005; Punch, 2009).

Advantages of group interviews include them being relatively inexpensive to conduct and they provide rich data. They can also be stimulating and aid in recall (Fontana & Frey, 1994, 2005). Disadvantages of group interviews include the group culture interfering with

individual expression, a single person dominating, or group dynamics not working (Fontana & Frey, 1994, 2005; Punch, 2009).

### ***Unstructured interviews***

If the semi-structured interview is less formal than the structured interview, then the unstructured interview presupposes nothing. Unstructured interviews develop and progress following the interviewee's flow of thoughts and ideas (Fontana & Frey, 2005; Opie, 2004; Roth, 2005). Most likely the interviewer will have some topic areas and/or a list of issues to be raised within the informal conversation that is the unstructured interview (Banister et al., 1994; Cohen & Manion, 1994; Gay et al., 2012). The unstructured interview allows for a greater breadth than other types of interviewing (Fontana & Frey, 2005). This type of interviewing requires a high level of expertise, with some serious training in the art of unstructured interviews, and is certainly not for novices (Punch, 2009; Wellington, 2000).

### ***Interview as a conversation***

The interview process is like a conversation. It involves two or more people in an exchange of ideas or views with the purpose of mutually creating a story. The conversation is instigated by the interviewer, with the purpose of obtaining information that is relevant to the research at hand (Cohen & Manion, 1994; Fontana & Frey, 2005; Wellington, 2000). It is a “conversation with a purpose” (Banister et al., 1994, p. 51; Wellington, 2000, pp. 71, 101), one where normal structures of conversations are suspended.

### ***Interview process***

The process of the interview has similarities across the different authors. The underlying key themes relate to the participants, the interview – schedule and actual interview – and recording (Banister et al., 1994; Davies, 2010; Punch, 2009; Wellington, 2000).

Things to consider around participants include deciding who and how many to interview, when and where the interview will take place, and how contact will be made with the potential participants (Banister et al., 1994; Punch, 2009; Wellington, 2000). Considerations for the interview schedule include how to translate research questions/objectives into interview questions, how much structure will the interview have, the types of questions to be asked, the order of the questions, and how to eliminate confusing, ambiguous or insensitive questions (Banister et al., 1994; Patton, 1990; Wellington, 2000). Within the actual interview

the following are considered: How do we establish rapport or break the ice? Why are we doing the interview and why were they (the participants) chosen? How do we communicate and listen within the interview? and How do we close the interview, thank the participant, round it up? (Banister et al., 1994; Davies, 2010; Fontana & Frey, 1994; Fontana & Frey, 2005; Merton et al., 1956/1990; Punch, 2009; Wellington, 2000). Issues or considerations around recording include deciding the type of recording (such as notes, tape record or video record), seeking permission to record the interview, assuring anonymity, agreeing to terminate the recording at any time, and involving the participants in verifying the notes (Banister et al., 1994; Patton, 1990; Punch, 2009; Wellington, 2000).

### ***Approach of the interviewer***

For the interview process to be a successful data collection tool, the approach of the interviewer is paramount. The interviewer's attitude to participants is critical, and interviewers need to convey to participants, through their actions, that the participants' views are important and informative (Banister et al., 1994; Opie, 2004; Wellington, 2000). Interviewers need to be able to listen carefully and be skilful in the interview process. They need to be able to frame questions well, probe gently to seek further clarification or elaboration of ideas, and be adept at personal interactions. They can put participants at ease, through some previously established relationship, and will have chosen a non-threatening location for the interview (Davies, 2010; Walford, 2001; Wellington, 2000).

### ***Advantages and disadvantages of interviews***

There are a number of obvious advantages to using an interview in a research situation. If the interview is tape or video recorded, then the natural language is preserved, both of the participant and the interviewer. During the interview itself, the interviewer can immediately follow up responses and seek further clarification. In addition, as the interview proceeds, the questions can be tailored based on participant responses to earlier questions or probes. A wide range of issues can be covered and/or issues can be covered to a greater depth and there are volumes of data (Banister et al., 1994; Walford, 2001; Wellington, 2000).

Disadvantages, limitations and weakness of interviews fall into a few categories: those around the participants, those around analysis and those around process. Participants may be unwilling or uncomfortable with what the interviewer hopes they will share and therefore not share this information, they may not be truthful, or they may only reveal what they want to

reveal (Walford, 2001; Wellington, 2000). In addition participants can be subject to interview fatigue if the schedule is unduly long. The analysis phase brings its own set of issues for interviews: there can be simply too much data, or the material can be “misinterpreted or over interpreted, and manipulated to produce meanings that were not ‘originally’ there” (Banister et al., 1994, p. 64). The interview process is prone to subjectivity and bias, and poor data can result if the interviewer prompts the interviewee or uses two-in-one, leading or loaded questions. Interviewing can also be time consuming (Banister et al., 1994; Cohen & Manion, 1994; Wellington, 2000).

### ***Ethical issues in interviewing***

Ethical issues in interviewing are best summed up by Fontana and Frey (2005):

Because the objects of inquiry in interviewing are humans, extreme care must be taken to avoid any harm to them. Traditionally, ethical concerns have revolved around the topics of informed consent (receiving consent by the respondent after having carefully and truthfully informed him or her about the research), right to privacy (protecting the identity of the respondent), and protection from harm (physical, emotional, or any other kind). (p. 715)

Fontana and Frey (2005) also point out other ethical considerations such as overt/covert field work and surreptitious use of tape-recording devices, and the degree of involvement with the group under study and how this involvement is used to gain access and information.

### **3.5. Data analysis**

The process of bringing order, structure, and interpretation to a mass of collected data is messy, ambiguous, time-consuming, creative, and fascinating. It does not proceed in a linear fashion; it is not neat. (Marshall & Rossman, 2006, p. 154)

The introductory quote above by Marshall and Rossman (2006) in their section on generic data analysis strategies sums up nicely the process of data analysis. It is the messy and often ambiguous process that Marshall and Rossman allude to that ultimately realises a coherent story that has been nurtured by the researcher as he or she seeks clarity and meaning from his or her data. Data analysis includes organisation, description, analysis and interpretation of data and, as intimated by the introductory quote, these do not necessarily happen in a linear or even logical fashion (Marshall & Rossman, 2006; Silverman, 2000).

The massive amount of data from interviews, observations and field notes is reduced by looking for significant patterns in the data. Coding these and generating themes and

categories helps to make meaning from the data and ultimately concepts and ideas fall into place (Marshall & Rossman, 2006; Silverman, 2000; Smith & Davies, 2010; van Nes & Doorman, 2010). Successive iterations allow for themes to be tested and refined (Huberman & Miles, 1994). Strategies to reduce the data include grouping common answers to questions or common interview responses or working chronologically and telling a story from the beginning to the end when working with observations (Patton, 1990). Findings are interpreted for meaning; this allows explanations to be given and conclusions drawn. Alongside this process of interpretation sits reflection, the testing of ideas and looking for alternative explanations (Huberman & Miles, 1994; Marshall & Rossman, 2006; Silverman, 2000; Smith & Davies, 2010).

### 3.5.1. SOLO taxonomy

The Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs & Collis, 1982; Hook & Mills, 2011; Uniservices asTTle team, 2008; Watson, 2006; Wikipedia, 2012) has been used by a number of researchers in statistics education including, and most notably, Watson (Watson, 2005, 2006; Watson & Moritz, 1999). The SOLO taxonomy has also been used as a foundational model for describing the different levels of thinking in the mathematics and statistics achievement standards (the assessment tools for national qualifications in New Zealand). Pragmatically, therefore, the SOLO taxonomy is a useful tool to use in grading student responses in pre- and post-tests due to the major impact it has on assessment in years 11–13 (ages 15–18, curriculum levels 6–8) which is where the curriculum level 5 (ages 13–15) material, the focus in this thesis, is heading.

The SOLO taxonomy is hierarchical, has five levels and is based on the quality of a student response to a task:

**Pre-structural** – The task is not attacked appropriately; the student hasn't really understood the point and uses too simple a way of going about it.

**Uni-structural** – The student response only focuses on one relevant aspect. Presents simple and obvious information.

**Multi-structural** – The student response focuses on several relevant aspects but they are treated independently and additively. Uses two or more facts, which are not necessarily related to each other.

**Relational** – The different aspects have become integrated into a coherent whole. This level is what is normally meant by an adequate understanding of some topics.



**Extended abstract** – The previous integrated whole may be conceptualised at a higher level of abstraction and generalised to a new topic or area. Makes connections not only with given subject material but also beyond it. (Wikipedia, 2012)

In the mathematics and statistics achievement standards in New Zealand, the step up in thinking within an achievement standard is described in terms of *achieved*, *achieved with merit*, and *achieved with excellence* (achieved, merit, excellence). The link between the step ups in achievement standard thinking and the SOLO taxonomy is: at achieved, students present uni-structural or multi-structural evidence (they are using simple and obvious information and may use several relevant aspects); at achieved with merit, students present relational evidence (making connections among the pieces of information); and at achieved with excellence, students present extended abstract evidence (conceptualised at a higher level and make connections beyond the given subject material) (Uniservices asTTle team, 2008; Wikipedia, 2012).

In terms of the statistics achievement standards (the assessment tools for national qualifications in New Zealand), the descriptors for achieved, merit and excellence for achievement standard AS91035 (the achievement standard for the curriculum on which this thesis is based) are given in Figure 3-3. This is a good example of how the levels of thinking are described in the statistics setting.

**Achievement**

*Using the statistical enquiry cycle* involves using each component of the statistical enquiry cycle to make comparisons.

**Achievement with Merit**

*Using the statistical enquiry cycle with justification* involves linking aspects of the statistical enquiry cycle to the context and the population and making supporting statements which refer to evidence such as summary statistics, data values, trends or features of visual displays.

**Achievement with Excellence**

*Using the statistical enquiry cycle with statistical insight* involves integrating statistical and contextual knowledge throughout the statistical enquiry cycle, and may involve reflecting on the process or considering other explanations for the findings.

**Figure 3-3. AS91035: Investigate a given multivariate data set using the statistical enquiry cycle**

*Note:* Retrieved from <http://www.nzqa.govt.nz/ncea/assessment/view-detailed.do?standardNumber=91035>

Figure 3-3 (previous page) shows that while descriptors and levels from the SOLO taxonomy are not explicitly used, the descriptions for each of the three levels (achieved, merit, excellence) connect the SOLO descriptions with aspects of the different types of thinking and the interrogative cycle in Wild and Pfannkuch's (1999) four-dimensional framework. The map between mathematics and the SOLO taxonomy is more straightforward; see, for example, AS91031: Apply geometric reasoning in solving problems (<http://www.nzqa.govt.nz/ncea/assessment/view-detailed.do?standardNumber=91031>). For the mathematics and statistics achievement standards with a statistics focus, relational thinking is labelled *with justification* and extended abstract thinking is labelled *with statistical insight*.

### 3.6. Ethics

The data collection period for this research was spread over five years. Ethics approval was sought and granted from the University of Auckland Human Participants Ethics Committee three times to cover the entire period of data collection. The first two years were covered by ethics approval 2007/085 and involved the first two rounds of data collection (2007 and 2008); the next two years were covered by ethics approval 2009/042 and covered the third round of data collection (2009); and the final year was covered by ethics approval 2011/258 and involved the fourth and final round of data collection (2011).

### 3.7. Triangulation

Triangulation is the use of more than one source of evidence to strengthen the validity of results (Banister et al., 1994; Cohen & Manion, 1994; Gay et al., 2012; Huberman & Miles, 1994; Patton, 1990; Punch, 2009; Schoenfeld, 2007; Slavin, 2007).

By self-consciously setting out to collect and double-check findings, using multiple sources and modes of evidence, the researcher will build the triangulation process into ongoing data collection. (Huberman & Miles, 1994, p. 438)

... the use of multiple lenses on the same phenomena is essential. In some cases, that means employing multiple methods to look at the same phenomena. Thus, observations, questionnaires, and interviews can all be used to challenge, confirm or expand the information gathered from each other. (Schoenfeld, 2007, p. 87)

Triangulation generally falls into four categories: data triangulation, analyst/investigator triangulation, method triangulation and theoretical triangulation. *Data triangulation* could involve using different data collection methods, getting perspectives from different people, or

collecting data over time; *analyst/investigator triangulation* involves the use of more than one analyst or investigator to collect or provide viewpoints on the data; *method triangulation* involves using mixed methods to collect data – this could be a mix of qualitative and quantitative methods; and *theoretical triangulation* is less common and uses different theoretical perspectives to look at the data (Banister et al., 1994; Cohen & Manion, 1994; Patton, 1990).

### 3.8. Validity and reliability

Unless you can show your audience the procedures you used to ensure that your methods were reliable and your conclusions valid, there is little point in aiming to conclude a research dissertation. (Silverman, 2000, p. 175)

Reliability in research centres around two key notions: (1) the degree of consistency in categorising instances or events; and (2) the extent to which repeating the research would elicit the same results. Suggested processes include documenting the procedures used and keeping an audit trail of interpretations and decisions (Banister et al., 1994; Gay et al., 2012; Punch, 2009; Silverman, 2000; Slavin, 2007).

Validity is concerned with the truth in the claims. Silverman (2000) suggests aiming for more valid findings by considering the following ways of thinking critically as qualitative data is analysed:

- *the refutability principle*: “... qualitative researchers seek to refute their initial assumptions about their data in order to achieve objectivity” (p. 178)
- *the constant comparative method*: test emerging hypothesis from a small data set on a bigger data set
- *comprehensive data treatment*: all cases of data are treated in the data analysis, so that all pieces of the data have contributed to the generalisations
- *deviant case analysis*: actively seeking out and addressing deviant cases, making strong connections with comprehensive data treatment where all pieces of data are considered
- *using appropriate tabulations*: quantitative methods using research-derived categories can afford access to data that has previously been lost in intensive qualitative research.

### **3.9. Summary**

The research methods and procedures that have been used within this thesis have been described from a theoretical point of view in this chapter. Chapter 4 – The Teaching Experiments – details the participants, teaching and learning cycle, and the data collection methods that were used in each of the four teaching experiments.

## **Chapter 4. The Teaching Experiments**

### **4.1. Introduction**

This chapter details the methodology in practice: how the design research methodology was enacted in this research; the data collection methods used and how they were used; who the participants were; and details of the teaching and learning programmes used in each of the teaching experiments.

### **4.2. Design research**

#### **4.2.1. Preparation and design phase**

Each of the four iterations of the research started with an initial planning and design phase in which members of the research team worked together to plan and prepare for the teaching experiment.

The research and development team ebbed and flowed with changes, which is inevitable when working with teachers. The first teaching experiment was part of a wider focus on specialised statistical content knowledge and had 15 members in the broader research group. Of these, only the classroom teacher and the researcher were directly involved in the teaching experiment. The broader research group were instrumental in the development of the teaching and learning materials that were used in the in-class teaching experiment. The second and fourth teaching experiments consisted of a much smaller research team, primarily the teacher and researcher, with input from the researcher's supervisors. The third teaching experiment was conducted simultaneously within the Teaching and Learning Research Initiative (TLRI) project "Building students' inferential reasoning: statistics curriculum levels 5 and 6 (ages 13–16)" (Pfannkuch et al., 2011). The broader research group for the TLRI project consisted of two principal researchers (including the author of this thesis), two practising statisticians and seven teachers (including the teacher for teaching experiments 3 and 4). Collectively the research group, including the teacher and researcher, form a pedagogical community (Cobb, 2000a).

Regardless of the research group or classroom teacher, the following two aspects were fundamental in preparation for the teaching experiment: (1) the planning and development of

the hypothetical learning trajectory (HLT) involved collaboration between the classroom teacher and researcher – this meant planning was a joint effort with the teacher and researcher sharing responsibility for the development of resources and teaching activities for use in the classroom; and (2) the teacher and researcher had shared understandings about the process and the end goals (Cobb, 2000a; Confrey & Lachance, 2000; Steffe & Thompson, 2000).

### **4.2.2. Teaching experiment**

During the course of this research study, two different teachers, at two different schools, were the collaborating teachers. Both teachers were working with year 10 students (ages 14–15), the first in a coeducational school, the second in a single-sex (girls only) school. Both teachers were involved in two iterations of the teaching experiment. The first teacher was involved in 2007 and 2008, and the second teacher in 2009 and 2011 (because the first teacher left to work overseas). The teaching experiment allowed the researcher to get first-hand experience of students' learning and reasoning. The researcher was present in the classroom for every lesson of the teaching experiment (Cobb, 2000a; Steffe & Thompson, 2000). In the fourth teaching experiment the researcher taught five lessons due to the collaborating teacher getting laryngitis and being unable to teach the class.

The teacher and researcher met to reflect and discuss the day's lesson at the end of each teaching session. The length of the discussion was dependent on time available and need from the lesson of the day. For the first teaching experiment these sessions were videotaped, whereas for the second, third and fourth teaching experiments only brief notes were made. These short daily debriefing sessions were an important part of the teaching experiment (Cobb, 2000a; Roth, 2005) because they allowed the teacher and researcher to develop a consensual or shared interpretation of the day's lesson and what was going on in the classroom.

As a result of the reflections, changes and adaptations were made to the HLT, activities updated or changed, local learning goals refined, interpretations improved, conjectures modified, and subsequent lessons revised (Cobb, 2000a; Confrey & Lachance, 2000; Gravemeijer, 1998; Hjalmarson & Lesh, 2008; Steffe & Thompson, 2000).

Capturing these changes and adaptations is a key aspect of the design research process as they are essential for careful retrospective analysis. Changes need to be well reported and these are detailed in chapters 6, 7 and 8.

## Chapter 4 – The Teaching Experiments

In the context of this research the evidence that was collected included student class work, pre- and post-tests, video recordings of lessons, video recordings of student interviews, video recordings of reflective discussions with the teacher, video recordings of the focus group (see next paragraph), and field notes.

The first teaching experiment captured, by video, the teacher when she was talking, various student groups when the teacher was not interacting with the class, and one group of students for the entire time. In the second and third teaching experiments a similar set up was used. In addition, in the second and third teaching experiments, three students were interviewed before and after the teaching experiment based on their pre- and post-test responses (a focused interview). The final teaching experiment built on experiences of the first three experiments. In an effort to get the best data, the video recording was made of the teacher when she was talking, and then of one specific group of six students. As well, this particular group was closely observed by the researcher and she asked prompting questions (mini interviews) to get the students to articulate their thinking or to clarify comments they had made. These six students were interviewed in pairs before and after the teaching experiment. The interviews were based on their pre- and post-test responses.

Table 4-1 is a summary of the year, the number of students involved, the types of data that were collected, and the number of lessons in each teaching experiment.

**Table 4-1. Summary information from teaching experiments with year 10 (ages 14–15) classes**

| Number of students<br>(permission given)<br><i>Year of teaching experiment</i> | Data collection types                                                                                                                                        | Number of<br>lessons |
|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| 26 (15)<br><b>2007</b>                                                         | Class video recording, pre- and post-tests, field notes                                                                                                      | 10                   |
| 25 (25)<br><b>2008</b>                                                         | Class and interview video recording, pre- and post-tests, student interviews following pre- and post-tests, field notes                                      | 10                   |
| 24 (24)<br><b>2009</b>                                                         | Class and interview video recording, pre- and post-tests, student interviews following pre- and post-tests, field notes                                      | 15                   |
| 29 (29)<br><b>2011</b>                                                         | Class and interview video recording, pre- and post-tests, student interviews following pre- and post-tests, mini group interviews, field notes, student work | 16                   |

### ***Observations***

Field notes, videos and photographs are common data collection methods used in observations (Banister et al., 1994; Opie, 2004; Roth, 2005), and in this research video recordings were used as the main method of collecting observational data. Initially as much activity as possible in the class was captured and it became clear that care needed to be taken not to do this. The focus of the video recordings was refined over time to focus on the teacher during whole class discussions and then one group during class activity time (Confrey & Lachance, 2000). Field notes were used to capture bigger ideas, reflections and to track thinking (Confrey & Lachance, 2000; Roth, 2005; Smith & Bowers-Brown, 2010).

The level of participation by the researcher influences the balance of objectivity and subjectivity in the research. The role that the researcher took in these sessions was, for the majority of the times, observer as participant; however, in five lessons in teaching experiment 4, the researcher was participant as observer since she was teaching the class because the teacher was unwell. In the case of observer as participant, the researcher interacted with students by prompting for answers to issues and seeking deeper explanations of student thinking, hence interacting with students but not taking a formal role in the group. In the participant as observer role, the researcher engaged more in the day-to-day activity of the classroom – in this case, as the teacher (Opie, 2004; Punch, 2009).

### ***Pre- and post-tests***

Pre- and post-tests were used in all four teaching experiments. These were adapted as the problematic situation changed to reflect the changing focus of the research. The changes to these are detailed in chapters 6, 7 and 8 and the tests are in Appendix B.

### ***Interviews***

In the data collection phase the students were interviewed following their pre- and post-test. These interviews were semi-structured (see page 56), with an interview guide that allowed for changes to the order and range of questions, and also allowed for additional prompts to be added. What was notable about these interviews was that they were based on the students' responses to the pre- and post-tests. Merton, Fische and Kendall (1956/1990), and Cohen and Manion (1994) describe the situation where prior analysis has been conducted before the interview as a focused interview.



During the fourth teaching experiment a group of six students was observed and videoed during the class activity time. When the teacher was teaching the class it was possible for the researcher to observe the group directly. During these times, when need arose for clarification or a deeper response the researcher conducted mini group interviews (Fontana & Frey, 2005; Punch, 2009). This action was not possible during the five lessons when the researcher was the teacher as the whole class needed the teacher/researcher's attention, not just the specific group of six students.

### **4.2.3. Retrospective analysis**

The retrospective analysis across the teaching experiments involved careful analysis of the students' pre- and post-tests, class interactions and interviews, looking for significant patterns in the data through coding and generating themes (Marshall & Rossman, 2006; Silverman, 2000; Smith & Davies, 2010; van Nes & Doorman, 2010). As these patterns were identified through one of the data collection methods (for example, pre- and post-tests), the other data (for example, student interviews and class interactions) was cross-referenced, thus challenging and confirming patterns in the data (Schoenfeld, 2007). Key teaching moments were explored and student in-class interactions and interview responses were analysed to ascertain student understanding of underpinning statistical concepts.

The patterns in the data frequently led to the development of frameworks or categorisation of ideas that were strongly grounded in the data (Punch, 2009; Schram, 2003; Strauss & Corbin, 1994). That is, the frameworks and categorisations were developed using the data and also tested against the data through subsequent teaching experiments or by exploring data from different data collection methods (Banister et al., 1994; Cohen & Manion, 1994; Patton, 1990). Initial retrospective analysis at the end of each teaching experiment realised new problematic situations and these motivated adaptations to the HLT and the design of new instructional materials (Hjalmarson & Lesh, 2008).

The SOLO taxonomy (Biggs & Collis, 1982; Hook & Mills, 2011; Uniservices asTTle team, 2008; Watson, 2006; Wikipedia, 2012) was used pragmatically to grade student responses in their pre- and post-tests for posing investigative questions, making the call and describing distributions. The use of the SOLO taxonomy across the different themes allowed for consistency of grading and aligned with the nature of the assessment in senior secondary classes in the national qualifications.

The participants and an outline of the teaching unit (HLT) for each teaching experiment are now described.

### 4.3. Teaching experiment 1: Recognising the initial problematic situation

#### 4.3.1. Participants – Year 10 coeducational class

The first teaching experiment occurred in a state, coeducational, decile four, multicultural, suburban school. (In New Zealand a decile one school has students from the lowest socio-economic level while a decile 10 school has students from the highest socio-economic level). The teacher (T1, 2007) of a year 10 class (ages 14–15) and her students were involved. The teacher was in her fifth year of teaching. There were 26 students (14 girls and 12 boys) in the class, of which 15 gave permission for data related to them to be used in the study. The 15 students comprised ten girls and five boys. The class had a mix of ethnicities including New Zealand European, Māori, Pasifika, Indian and Chinese. The students were in an average to below-average streamed class.

#### 4.3.2. Teaching and learning cycle

A ten-lesson teaching unit was developed collaboratively by the researcher and the teacher to follow the PPDAC cycle. The material in the teaching unit recognised that statistics was not taught to any of the year 9 (ages 13–14) students in the school in the previous year. The teacher and researcher were involved in a reflection process after each lesson. This allowed adjustments to the original plan as issues came to light and it seemed sensible to make the changes immediately. Table 4-2 gives the teaching outline for teaching experiment 1. The learning goals and instructional activities are given for each lesson.

**Table 4-2. Lesson detail, teaching experiment 1**

| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"><li>• Pose “I wonder” statements about the data; classify as summary, comparison or relationship.</li><li>• Begin to understand what is meant by cleaning the data.</li></ul> <u>Learning activities</u> <ul style="list-style-type: none"><li>• Use data cards created from CensusAtSchool data.</li><li>• Write and classify “I wonder” statements.</li><li>• Look for anything unusual or possible errors.</li></ul> |

## Chapter 4 – The Teaching Experiments

| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Re-write their “I wonder” questions to allow them to be answered from the data.</li> <li>• Construct graphs using data cards.</li> <li>• Answer their questions using the starter “I notice ...” (Shaughnessy, 1997) and statistical values.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Review questions from previous lesson.</li> <li>• Select a question to investigate, sorting the data and making possible graphs using data cards.</li> <li>• Use the starter “I notice ...” to generate statements about the data.</li> <li>• Introduce maximum, minimum, range, mode, median.</li> </ul> |
| 3             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Write a conclusion using “I notice” statements and relate this back to the original question.</li> <li>• Draw a dot plot.</li> <li>• Find statistical values such as maximum, minimum, range, median.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Create dot plots.</li> <li>• More work on “I notice” statements.</li> <li>• Writing a conclusion example.</li> </ul>                                                                                                                                                                                                                             |
| 4             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Calculate summary statistics for box and whisker plots.</li> <li>• Draw box and whisker plots.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Explore “I wonder what the difference in right foot lengths is from year 5 to year 10.”</li> <li>• Use provided dot plots to create box plots.</li> </ul>                                                                                                                                                                                                                                                                                               |
| 5             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Write comparison questions.</li> <li>• Draw dot plots.</li> <li>• Use box and whisker plots to compare.</li> <li>• Write a conclusion from box and whisker plots.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Investigating comparison questions; for example, “I wonder if year 10 boys are taller than year 10 girls?”</li> </ul>                                                                                                                                                                                                                                                                |
| 6             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Draw dot plots and box and whisker plots.</li> <li>• Write analysis and conclusion about the comparison situation being explored.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Pose a comparison question to compare our class to another year level.</li> <li>• Use dot plots, box and whisker plots and draw conclusions relating back to the question.</li> </ul>                                                                                                                                                                                                                                |
| 7             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Write an analysis when given the dot plots and box plots.</li> <li>• Write a conclusion from analysis and plots.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Reflect on previous “I notice” statements.</li> <li>• Use the PPDAC cycle with a focus on analysis and conclusions.</li> </ul>                                                                                                                                                                                                                                                                                                        |
| 8             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Understand key concepts needed for conclusion.</li> <li>• Understand the PPDAC cycle.</li> <li>• Write an analysis and conclusion.</li> <li>• Calculate the mean for a given data set.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Writing frame using PPDAC cycle.</li> </ul>                                                                                                                                                                                                                                                                                                                     |
| 9             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>• Undertake a statistical investigation using the PPDAC cycle.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 10            | <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>• Speedsters <a href="http://new.censusatschool.org.nz/resource/speedster/">http://new.censusatschool.org.nz/resource/speedster/</a>.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

#### **4.3.3. Data collected**

Each lesson was videotaped and transcribed at a later date. All students in the class were pre- and post-tested. See Appendix B.1 and B.2 for pre- and post-tests for teaching experiment 1.

#### **4.4. Teaching experiment 2: What makes a good investigative question?**

##### **4.4.1. Participants – Year 10 coeducational class**

The second teaching experiment occurred in the same state, coeducational, decile four, multicultural, suburban school as the first teaching experiment. The same teacher (T1, 2008) and her current year 10 class (ages 14–15) were involved. The teacher was then in her sixth year of teaching. There were 25 students in the class and all gave permission for data related to them to be used in the study. The 25 students comprised sixteen girls and nine boys. The class had a mix of ethnicities including New Zealand European, Māori, Pasifika, Indian and Chinese. The students were in a top-stream class doing year 11 (ages 15–16) work.

##### **4.4.2. Teaching and learning cycle**

The teaching unit from the previous year was updated and further developed collaboratively by the researcher and the teacher, following the PPDAC cycle. An additional emphasis was given to posing investigative questions. Following each lesson, the teacher and researcher met and discussed the lesson, reflecting on specific elements of the lesson; for example, posing investigative questions. In some instances changes were made to the teaching for the next lesson as a result of the discussion in these reflective sessions. The lessons where prototypical instructional material was developed or new lessons were created that supported the new problematic situation are highlighted in Table 4-3 (next page). The learning goals and instructional activities are given for each lesson, and for the highlighted lessons the hypothetical learning process is also given.

Table 4-3. Lesson detail, teaching experiment 2

| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Familiarisation/review of PPDAC cycle.</li> <li>Familiarisation with data cards.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 2             | <ul style="list-style-type: none"> <li>Cleaning data.</li> <li>Posing investigative questions</li> <li>“Growing scatterplots”.</li> <li>Noticing the relationship.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 3             | <u>Learning activities</u> <ul style="list-style-type: none"> <li>Growing scatterplots <a href="http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S6-1">http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S6-1</a>.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>Students notice the different variables on the data cards and use their knowledge of the CensusAtSchool questionnaire and different body measurements to ascertain what the variables are; they recognise that the different coloured cards indicate gender.</li> <li>Students will use tape measures and common sense to check values they are suspicious about. They discard data cards that have values they think are too large or too small for a particular variable, using the other measurement variable as a guide and also appreciating a sensible range for the particular variable.</li> <li>Students pose investigative questions that have the variable, population and intent clear.</li> <li>Students “grow” scatterplots noticing the relationship, for example, that the taller students have longer arm spans.</li> </ul> |
| 4             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Describing scatterplots.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Set of eight provided scatterplots with data and short background story.</li> <li>Descriptions include considering: middle group, spread, shape, cluster and unusual values.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 5             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>To critique and improve previously posed summary and comparison investigative questions.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Selection of questions posed in students’ pre-tests.</li> <li>Sort questions, critique and improve using criteria for a good question.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>Students are able to differentiate between summary and comparison questions.</li> <li>Students are able to critique previously posed questions using the criteria given by the teacher; they can identify whether the question is a summary or comparison question; they are able to improve the question using the target population and variable.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 6             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Writing analysis and conclusions for summary and comparison questions.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Provided graphs for the questions chosen.</li> <li>Analysis and conclusions written.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 7             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Posing and answering comparison questions.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 8             | <u>Learning activities</u> <ul style="list-style-type: none"> <li>Sleeping sheep <a href="http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S6-1">http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S6-1</a>.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 9             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Using the PPDAC cycle to explore comparison questions.</li> <li>Making a claim, answering the comparison question.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 10            | <u>Learning activities</u> <ul style="list-style-type: none"> <li>Data sets provided.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

#### **4.4.3. Data collected**

Each lesson was videotaped and transcribed at a later date. All students in the class were pre- and post-tested. Three self-selected students were interviewed following the pre-test and again following the post-test. Interview prompts based on findings from the previous year (Arnold, 2008b) were used in interviews following both the pre- and post-tests. See Appendix D for interview schedules.

Pre- and post-tests were used to collect evidence about students' investigative question posing and other aspects of the statistical enquiry cycle. See Appendix B.3 and B.4 for pre- and post-tests for teaching experiment 2.

### **4.5. Teaching experiment 3: Answering the investigative question – making the call**

#### **4.5.1. Participants - Year 10 single-sex (girls only) class**

Because the original teacher left New Zealand to teach overseas, a second teacher was approached and the third teaching experiment occurred in her school. This school was a state, decile five, multicultural, inner-city girls' school. The teacher (T2, 2009) and her year 10 class (ages 14–15) were involved. The teacher was in her ninth year of teaching. There were 24 students in the class and all gave permission for data related to them to be used in the study. The class had a mix of ethnicities including New Zealand European, Māori, Pasifika and Chinese. The students were in a middle-stream class and of average ability.

#### **4.5.2. Teaching and learning cycle**

The teaching and learning cycle was developed to meet the needs of the new school and to work from the lessons learnt from the previous two teaching experiments. In particular, there was a further focus on critiquing investigative questions as well as posing them, the idea of sampling was introduced and, from this, describing graphs and making the call in comparison situations. The lessons where prototypical instructional material was developed or new lessons were created that supported the new problematic situation are highlighted in Table 4-4 (next page). The learning goals and instructional activities are given for each lesson, and for the highlighted lessons the hypothetical learning process is also given.

**Table 4-4. Lesson detail teaching experiment 3**

| Lesson number                                                                                                                                                                                                                                                                    | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Plans for lessons 1–4 can be found at: <a href="http://new.censusatschool.org.nz/resource/statistical-investigations-part-1-introduction-to-the-ppdac-cycle/">http://new.censusatschool.org.nz/resource/statistical-investigations-part-1-introduction-to-the-ppdac-cycle/</a> . |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 1                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Reflect on year 9 work and the PPDAC cycle.</li> <li>• Introduce CensusAtSchool data set and set up the context.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Exploring the PPDAC cycle using data cards.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                  |
| 2                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Posing investigative questions and looking at what makes a good investigative question.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Sorting the investigative questions into different types and using this to inform a classification and elements of a good investigative question.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students will identify some or all of the criteria for what makes a good investigative question through their critique and classification of the investigative questions (see Figure 6-3).</li> </ul> |
| 3                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Describing summary graphs.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Using prepared writing frame to describe summary graphs.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| 4                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Describing comparative graphs.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Using prepared writing frame to describe comparative graphs.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Plans for lessons 5–10 can be found at: <a href="http://new.censusatschool.org.nz/resource/statistical-investigations-part-2/">http://new.censusatschool.org.nz/resource/statistical-investigations-part-2/</a> .                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 5                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Revisit posing investigative questions.</li> <li>• Identifying and clarifying the population.</li> <li>• Exploring sampling variability.</li> <li>• What is a sample and why sample?</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Karekare College data set.</li> <li>• What are typical popliteal lengths of students at Karekare College?</li> <li>• Make plots for different samples; look at what is different, what is the same across samples.</li> </ul>                                                                                                                       |
| 6                                                                                                                                                                                                                                                                                | <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students are able to pose an investigative question about the population of Karekare College and critique their questions and improve as necessary.</li> <li>• Students “discover” the need to use a sample to answer the question about the population.</li> <li>• Students anticipate what the graph of popliteal lengths will look like using any prior knowledge they may have.</li> <li>• Students start to acknowledge that samples from the same population for the same variable are similar but have differences, and to articulate what these similarities and differences are.</li> </ul>                |
| 7                                                                                                                                                                                                                                                                                | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Can a sample tell us something about the population?</li> <li>• Informal inferential reasoning.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Noticing from their graphs and displays and wondering about what might be happening “back</li> </ul>                                                                                                                                                                                                                                                                                                                                     |

## Chapter 4 – The Teaching Experiments

| Lesson number                                                                                                                                                                                                    | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 8                                                                                                                                                                                                                | <p>in the population”.</p> <ul style="list-style-type: none"> <li>Exploring this also for both summary and comparative situations.</li> <li>Hypothetical learning process</li> <li>Students are prepared to make statements about what they think the population(s) might look like based on the samples they have taken.</li> <li>Students are making statements about whether they think one group tends to have bigger values than another group “back in the population(s)”.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 9                                                                                                                                                                                                                | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Introduction to middle group and spread.</li> <li>Describing middle-group position relative to one another.</li> <li>Formalising box plots.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 10                                                                                                                                                                                                               | <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Exploring comparison investigative questions through identifying middle groups and moving to using box plots.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| Plans for lessons 11–15 can be found at: <a href="http://new.censusatschool.org.nz/resource/statistical-investigations-part-3/">http://new.censusatschool.org.nz/resource/statistical-investigations-part-3/</a> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| 11                                                                                                                                                                                                               | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Using Fathom (Finzer, 2007) to draw dot plots and box plots.</li> <li>Describing sample data.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Use boysgirls.ftm; juniorsenior.ftm; buswalk.ftm to explore dot plots and box plots, to get summary tables and to describe the data.</li> <li>Generate height and time-to-school graphs, recording the “box” only.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 12                                                                                                                                                                                                               | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Exploring the direction and amount of shift of the middle 50% when comparing two groups.</li> <li>Exploring the consistency across multiple samples of the same size from the same populations.</li> <li>Initial exploration of “making the call”.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Using previously prepared height and time-to-school graphs to look at patterns across samples and between situations (height and time to school).</li> </ul> <p><u>Hypothetical learning process</u></p> <ul style="list-style-type: none"> <li>Students recognise the patterns in the height graphs – shift is inconsistent, sometimes the middle 50% of the boys’ data is further to the right than the middle 50% of the girls’ data, sometimes it is the other way around; the boxes overlap in all cases and sometimes this is a complete overlap; sometimes the boys’ median height is higher, sometimes the girls’ median height is higher; and the median heights are both within the overlap of the middle 50%.</li> <li>Students recognise the patterns in the time-to-school graphs – shift is consistent, the middle 50% of the bus travel time to school is always further to the right than the middle 50% of walk travel time to school; in most cases the boxes do not overlap, for some cases there is a small overlap of the boxes; median time to school for bus is always higher than the median time to school walking; at least one of the medians is outside the overlap of the boxes.</li> </ul> |
| 13                                                                                                                                                                                                               | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Reinforcing the “making the call” message using “movies”.</li> <li>Developing clarity around the message.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Heights “movie” <a href="http://new.censusatschool.org.nz/wp-content/uploads/2012/11/heights_2samp_dots_30.pdf">http://new.censusatschool.org.nz/wp-content/uploads/2012/11/heights_2samp_dots_30.pdf</a> .</li> <li>Time-to-school “movie” <a href="http://new.censusatschool.org.nz/wp-content/uploads/2012/11/times_2samp_dots_30.pdf">http://new.censusatschool.org.nz/wp-content/uploads/2012/11/times_2samp_dots_30.pdf</a></li> </ul> <p><u>Hypothetical learning process</u></p> <ul style="list-style-type: none"> <li>Students are fluent in which situation they are looking at and what the “call” would be.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 14                                                                                                                                                                                                               | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Exploring other data sets to “make the call”.</li> <li>Constructing conclusions for comparison situations.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Senior/junior bag weights; male/female incomes.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |



| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                             |
|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 15            | <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>Students are able to “make the call” from a new data set using the ideas developed in previous lessons.</li> <li>Students are able to write a conclusion that is consistent with the samples they have in answer to their comparative investigative question about the populations.</li> </ul> |

### 4.5.3. Data collected

Each lesson was videotaped and transcribed at a later date. All students in the class were pre- and post-tested. Three students were interviewed following the pre-test and again following the post-test. These students were in the particular group that was videotaped. Interview prompts based on findings from previous two iterations were used in interviews following both the pre- and post-tests. See Appendix D for interview schedules.

Pre- and post-tests were used to collect evidence about students’ understanding of different components of the statistical enquiry cycle, including posing investigative questions, describing distributions, making a call in a comparison situation, and writing conclusions. See Appendix B.5 and B.6 for pre- and post-tests for teaching experiment 3.

## 4.6. Teaching experiment 4: Answering the investigative question – describing distributions

### 4.6.1. Participants - Year 10 girls’ class

The fourth teaching experiment occurred in the same state, decile five, multicultural, inner-city girls’ school as the third teaching experiment. The same teacher (T2, 2011) and her year 10 class (ages 14–15) were involved. The teacher was then in her eleventh year of teaching. There were 29 students in the class and all gave permission for data related to them to be used in the study. The class had a mix of ethnicities including New Zealand European, Māori, Pasifika and Chinese. The students were in a top-stream class and of above-average ability.

### 4.6.2. Teaching and learning cycle

The teaching and learning cycle was updated further in this final teaching experiment with a deliberate focus on describing distributions, particularly shape and the language of shape. The lessons where prototypical instructional material was developed or new lessons were created that supported the new problematic situation are highlighted in Table 4-5 (next page). The

## Chapter 4 – The Teaching Experiments

learning goals and instructional activities are given for each lesson, and for the highlighted lessons the hypothetical learning process is also given.

**Table 4-5. Lesson detail teaching experiment 4**

| <b>Lesson number</b> | <b>Lesson content</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|----------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                    | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Introduce/reflect on the PPDAC cycle.</li> <li>• Working with year 10 girls data cards to make “dot” plots.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Creating dot plots using data cards.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 2                    | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Sketching the shape of the distribution for large samples.</li> <li>• Identifying patterns in the distributional shapes.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Shape activity part 1 (adapted from <a href="http://new.censusatschool.org.nz/resource/engaging-with-shape/">http://new.censusatschool.org.nz/resource/engaging-with-shape/</a>).</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students will sketch “smoother” shapes due to the larger sample size and the briefness of the view of the original data.</li> <li>• Students will group the shapes into subgroups and use their own language to describe the shape.</li> <li>• Agreement will be reached across the class as to the number and description of the groups.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 3*                   | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Predicting distributions for various variables and then matching the actual graph to the variable.</li> <li>• Firming up the description of shape for distributions introducing the statistical language for shape.</li> <li>• Developing a description for a distribution.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Shape activity part 2.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students make predictions for the different variables reflecting on the shapes that had been identified in the previous lesson. They are using their language of shape to describe the graph.</li> <li>• Students are able to match the variable to the graph using the predictions already made to confirm the match. Students are starting to develop a sense of what types of variables match with which shape.</li> <li>• Students are able to describe the shape of a graph using statistical terms such as symmetric, left and right skew, uniform, unimodal, and bimodal.</li> <li>• Students are able to identify some features of a distribution with which to describe the graph of the distribution.</li> <li>• Students have “identified” variable, population, values and units as features of a distribution description.</li> </ul> |
| 4*                   | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Developing a library of distributional shapes.</li> <li>• Describing the distribution for summary situations.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>• Mix’n’match activity with 12 distributions.</li> <li>• Using the situations from lessons 2 and 3 to write descriptions of distributions.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students can sort the new graphs into one of the four main types (symmetric, uniform, left skew, right skew) of distributional shapes.</li> <li>• Students can write a description of a distribution and include at least four features of</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |

## Chapter 4 – The Teaching Experiments

| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | <p>the graph.</p> <ul style="list-style-type: none"> <li>Students' descriptions include the context in the form of variable, values and units; some are starting to include the population.</li> </ul>                                                                                                                                                                                                                                                                                                            |
| 5*            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Posing investigative questions and looking at what makes a good investigative question.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Sorting the investigative questions into different types and using this to inform a classification and elements of a good investigative question.</li> </ul>                                                                                           |
| 6*            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Identifying and clarifying the population.</li> <li>Exploring sampling variability.</li> <li>What is and why sample?</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Karekare College data set.</li> <li>What are typical popliteal lengths of students at Karekare College?</li> <li>Make plots for different samples; look at what is different, what is the same across samples.</li> </ul> |
| 7*            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Can a sample tell us something about the population?</li> <li>Informal inferential reasoning.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Noticing from their graphs and displays and wondering about what might be happening “back in the population”.</li> <li>Exploring this also for both summary and comparative situations.</li> </ul>                                               |
| 8             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Introduction to middle group and spread.</li> <li>Describing middle-group position relative to one another.</li> <li>Formalising box plots.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Exploring comparison investigative questions through identifying middle groups and moving to using box plots.</li> </ul>                                                                           |
| 9             | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Using Fathom (Finzer, 2007) to draw dot plots and box plots.</li> <li>Describing sample data.</li> </ul> <p><u>Learning activities</u></p>                                                                                                                                                                                                                                                                                         |
| 10            | <ul style="list-style-type: none"> <li>Use boysgirls.ftm; juniorsenior.ftm; buswalk.ftm to explore dot plots and box plots, to get summary tables and to describe the data.</li> <li>Explore bag weights, time to school, height and print graphs.</li> <li>In addition for height and time-to-school graphs, make a second set of graphs recording the “box” only.</li> </ul>                                                                                                                                    |
| 11            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Critiquing investigative questions.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Questions posed worksheet.</li> </ul>                                                                                                                                                                                                                                                                      |
| 12            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Describing sample data.</li> </ul> <p><u>Learning activities</u></p> <ul style="list-style-type: none"> <li>Using Fathom-generated graphs (from lesson 9-10) to describe comparative situations.</li> </ul>                                                                                                                                                                                                                        |
| 13            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Exploring the direction and amount of shift of the middle 50% when comparing two groups.</li> <li>Exploring the consistency across multiple samples of the same size from the same populations.</li> <li>Initial exploration of “making the call”.</li> </ul>                                                                                                                                                                      |

| Lesson number | Lesson content                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | <u>Learning activities</u> <ul style="list-style-type: none"> <li>Using previously prepared height and time-to-school graphs to look at patterns across samples and between situations (height and time to school).</li> </ul>                                                                                                                                                                                                                               |
| 14            | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Reinforcing the “making the call” message using “movies”.</li> <li>Developing clarity around the message.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Heights and time-to-school “movies”.</li> </ul>                                                                                                                                               |
| 15            | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Clarification of the location of the median in “making the call”.</li> <li>Exploring other data sets to make the call.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Making the call on various ratings about ability, comparing boys and girls (CensusAtSchool data set).</li> <li>Male/female incomes.</li> </ul>                                   |
| 16            | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>Constructing conclusions for comparison situations.</li> <li>Using the whole PPDAC cycle.</li> </ul> <u>Learning activities</u> <ul style="list-style-type: none"> <li>Iron levels <a href="http://new.censusatschool.org.nz/wp-content/uploads/2012/11/IRON-DATA-Worksheet.pdf">http://new.censusatschool.org.nz/wp-content/uploads/2012/11/IRON-DATA-Worksheet.pdf</a>.</li> </ul> |

Note: \* lessons that the researcher taught.

#### 4.6.3. Data collected

Each lesson was videotaped and transcribed at a later date. All students in the class were pre- and post-tested. Six students, in pairs, were interviewed following the pre-test and again following the post-test. These students were in the particular group that was videotaped during class sessions. Interview prompts based on findings from previous three teaching experiments and TLRI project findings (Pfannkuch et al., 2011) were used in interviews following both the pre- and post-tests. Interviews also included two new activities for the students to do, similar to those in the pre- and post-tests, with a view to seeing the students “in action” and capturing their thinking as they attempted the activities rather than retrospectively. See Appendix D for interview schedules.

Pre- and post-tests were used to collect evidence about students’ understanding of different components of the statistical enquiry cycle, including posing investigative questions, describing distributions, making a call in a comparison situation, and writing conclusions. See Appendix B.7 and B.8 for pre- and post-tests for teaching experiment 4.

#### 4.7. Overview of the four teaching experiments

Figure 4-1 (next page) collates the four teaching experiments along with the information about the year, the school, the teacher, students and lessons. In addition the main activities

## Chapter 4 – The Teaching Experiments

are captured underneath, highlighting where each problematic situation was identified, and where subsequent research was undertaken and continued.

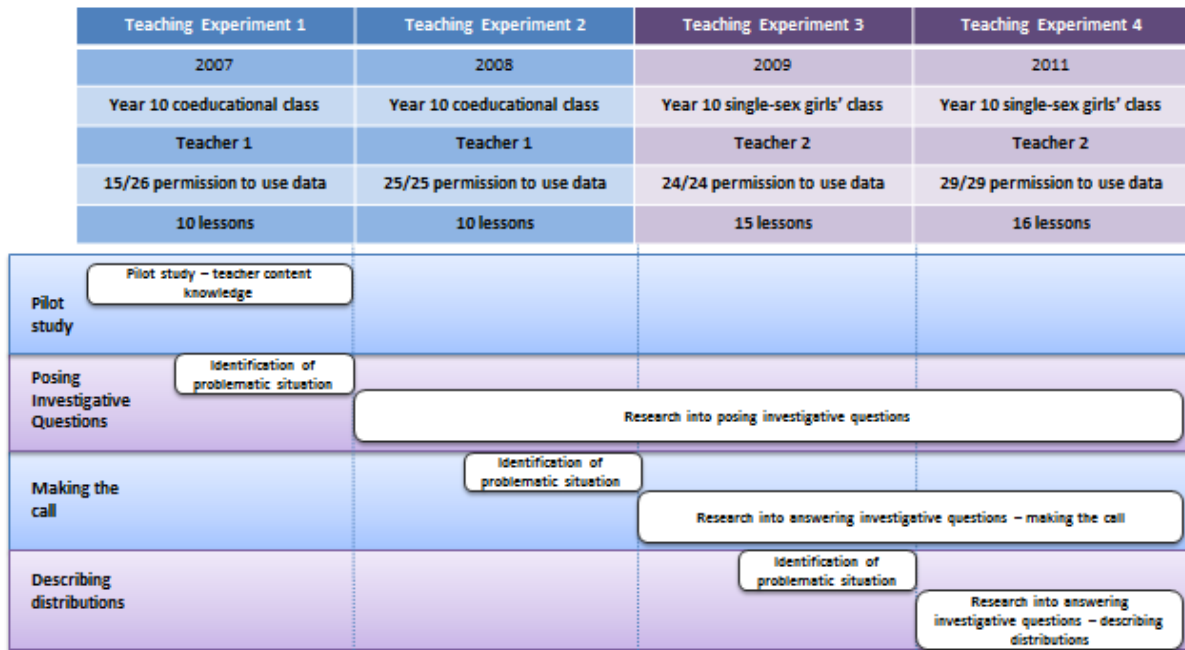


Figure 4-1. Overview of the four teaching experiments

### 4.8. Summary

Four different groups of year 10 (ages 14–15) students and two classroom teachers participated in the research. Each group brought different perspectives and knowledge to the specific teaching experiment and the resulting analysis is reflective across the four teaching experiments. Chapter 5 sets the scene by defining the initial problematic situation, and chapters 6, 7 and 8 complete the story from posing to answering investigative questions.

## **Chapter 5. Recognising the Initial Problematic Situation**

### **5.1. Introduction**

Chapter 5 introduces the initial problematic situation that forms the foundation of this doctoral thesis. Posing questions was identified as a problematic situation towards the end of the first teaching experiment, and this developed into the role of the investigative question within the statistical enquiry cycle. The scene is set with a description of the original exploratory study that was undertaken in the first year of research. Student pre- and post-test results are given and the realisation of the initial problematic situation is discussed. Finally, fledgling question categorisation ideas arising from the first teaching experiment are given and a retrospective analysis of student-posed questions is undertaken and discussed.

### **5.2. Setting the scene**

The first teaching experiment was set in the context of exploring specialised statistical content knowledge needed by teachers to meet the requirements of the statistical investigations thread in the new curriculum at level 5 (ages 13–15) (see Figure 2-1, page 11, for achievement objectives). The researcher worked with a group of secondary teachers through a series of workshops to support this exploration (Arnold, 2008a). Following the teacher workshops, one teacher was observed as she put her statistical content knowledge into effect in the classroom.

During the preparation and planning phase key statistical ideas were hypothesised based on the curriculum. A pre-test (see Appendix B.1) using this information was created and used to find out the prior knowledge of the students. Three areas, related to two aspects of the statistical enquiry cycle, were included in the pre-test: posing questions (problem), drawing data displays (analysis), and writing descriptive statements (analysis). The teacher and researcher had a detailed look at the student responses in the pre-tests to identify existing knowledge and gaps. From this analysis, a hypothetical learning trajectory was created and supporting teaching and learning activities were developed or sourced. For details on the teaching unit see Table 4-2, pages 71-72.

At the end of the unit the students completed a post-test. This was modelled on the style of the year 11 (ages 15–16) national assessment, where students are required to pose a question

for the given data, the teacher marks the questions, and then the students complete the assessment with the question they posed or, if their question was unsuitable, with a question provided by the teacher. In this teaching experiment the students posed three comparison questions which the teacher checked overnight. In the following lesson they completed the investigation for one of their questions that had been identified as suitable by the teacher. One student was given a question to work with as all three of her questions were unsuitable. Four areas, related to three aspects of the statistical enquiry cycle, were included in the post-test: posing questions (problem), drawing data displays (analysis), writing descriptive statements (analysis), and drawing conclusions (conclusion).

### 5.3. Student pre- and post-test analysis

Student pre- and post-tests were marked across the three and four areas, respectively. Each area was marked out of a possible five. Criteria were developed from the student responses (see Appendix C), the grading criteria hierarchy reflecting what the students presented as evidence across the two tests.

Table 5-1 is a summary of student pre- and post-test grades for the three areas – question, data display and data description – that were common to both tests.

- In the pre-test the question grade is for the question (summary or comparison) the student posed in order to undertake an investigation.
- For the post-test the question grade is the mean (0 d.p.) of the three comparison questions the student posed in order to undertake an investigation.

**Table 5-1. Student pre- and post-test marks**

|                |   | Post-test grade |   |   |   |   |   |       |               |   |   |   |   |   |       |                  |   |   |   |   |   |       |
|----------------|---|-----------------|---|---|---|---|---|-------|---------------|---|---|---|---|---|-------|------------------|---|---|---|---|---|-------|
|                |   | Question        |   |   |   |   |   | Total | Data displays |   |   |   |   |   | Total | Data description |   |   |   |   |   | Total |
|                |   | 0               | 1 | 2 | 3 | 4 | 5 |       | 0             | 1 | 2 | 3 | 4 | 5 |       | 0                | 1 | 2 | 3 | 4 | 5 |       |
| Pre-test grade | 0 | 0               | 0 | 1 | 2 | 1 | 0 | 4     | 0             | 0 | 1 | 3 | 2 | 0 | 6     | 1                | 1 | 2 | 1 | 0 | 0 | 5     |
|                | 1 | 0               | 0 | 0 | 0 | 2 | 0 | 2     | 0             | 0 | 0 | 0 | 0 | 0 | 0     | 0                | 0 | 0 | 1 | 1 | 0 | 2     |
|                | 2 | 0               | 0 | 2 | 0 | 1 | 0 | 3     | 0             | 1 | 2 | 1 | 1 | 0 | 5     | 0                | 0 | 0 | 1 | 0 | 0 | 1     |
|                | 3 | 0               | 0 | 0 | 0 | 2 | 0 | 2     | 0             | 0 | 0 | 0 | 1 | 1 | 2     | 1                | 0 | 1 | 2 | 1 | 0 | 5     |
|                | 4 | 0               | 0 | 0 | 0 | 1 | 0 | 1     | 0             | 0 | 0 | 0 | 0 | 0 | 0     | 0                | 0 | 0 | 0 | 0 | 0 | 0     |
|                | 5 | 0               | 0 | 0 | 1 | 0 | 0 | 1     | 0             | 0 | 0 | 0 | 0 | 0 | 0     | 0                | 0 | 0 | 0 | 0 | 0 | 0     |
| Total          |   | 0               | 0 | 3 | 3 | 7 | 0 | 13    | 0             | 1 | 3 | 4 | 4 | 1 | 13    | 2                | 1 | 3 | 5 | 2 | 0 | 13    |

Figure 5-1 shows the grade movement from pre- to post-test for the same three areas as shown in Table 5-1. There is strong evidence of movement ( $p$ -value  $\approx 0.007$ , paired  $t$ -test) from pre- to post-test in students' question grade (Figure 5-1a) with an average movement of 1.5 points; there is very strong evidence of movement ( $p$ -value  $\approx 0.001$ , paired  $t$ -test) from pre- to post-test in students' data displays grade (Figure 5-1b) with an average movement of 1.8 points; and there is weak evidence of movement ( $p$ -value  $\approx 0.094$ , paired  $t$ -test) from pre- to post-test in students' data descriptions grade (Figure 5-1c) with an average movement of 0.8 points. These results suggest that through engagement in the classroom activities for the unit, students' improved their knowledge in the three areas.

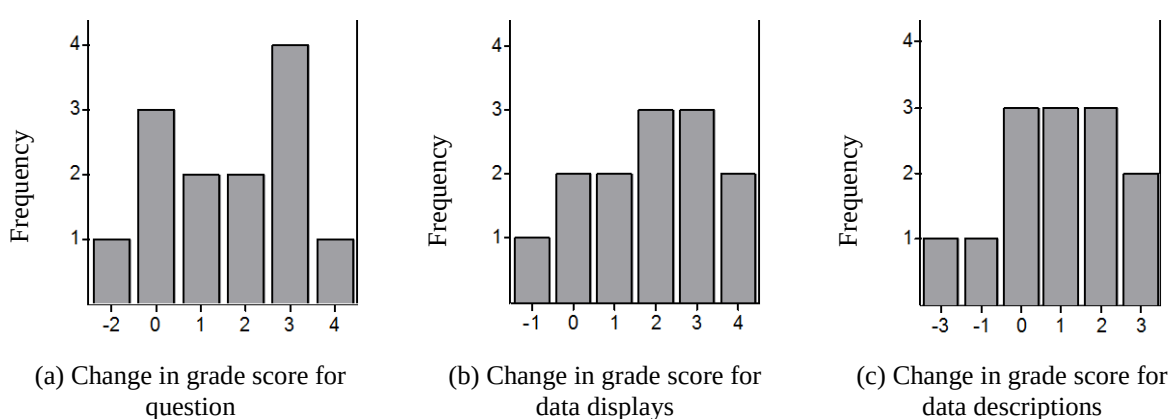


**Figure 5-1. Grade movement from pre- to post-test**

However, it is important to remember that the marking schedule was developed to look at general movement rather than focus on particular statistical concepts. The criteria were built from student responses and therefore reflected the capacity of this group of students and the current statistical knowledge imparted to students. Performance achievement was not based on pre-determined criteria that students needed to acquire to understand the statistical concepts or achieve a desired outcome. That is, the pre- and post-test were not designed in this first teaching experiment to show student improvement for a specific statistical concept; rather, the pre-test was to support the development of the hypothetical learning trajectory and to identify students' prior knowledge, while the post-test reflected the subsequent designed unit of work and also acknowledged requirements of the national assessment (NCEA level 1 achievement standards) in statistics.

#### 5.4. The problematic situation is realised

The teacher and the researcher together looked at the student post-tests with a focus on where student performance was limited and together noticed two different scenarios. The first



scenario was that a number of the students failed to complete the test despite being given additional time, mostly stumbling at the graph-drawing stage. The second scenario was that a few students had graphs with very small sample sizes and this resulted in poor descriptions and conclusions. The teacher and researcher together explored these students' scripts further to see if there were any patterns.

In the first scenario, where the students failed to complete the assessment, the common element was that the students had chosen to explore reaction time and had stumbled on constructing the scale for the graph. The problem with the scale was that most of the data was between 0.3 seconds and 0.7 seconds but there was an outlier at 2.58 seconds and they were trying to go up in 0.1 second intervals. Reflection at this point was to think about whether the data needs to be cleaned to remove outliers as the assessment is not about graph construction but more focused on the description of the data and the conclusion. On further reflection a solution to the problem was to use technology for graph construction. The road block to scale creation could be avoided, as technology does this automatically, and the student would be able to concentrate on describing the data.

In the second scenario some of the students had drawn their graphs correctly, but for one of the groups they were comparing the group had very small sample sizes, five or less. It was in exploring these particular students' test scripts that the problematic situation was realised. The questions (Figure 5-2) the students had chosen to use for their investigation had been marked as suitable by the teacher and the researcher had also agreed that the question looked suitable, but in fact the questions were not suitable for the data set supplied.

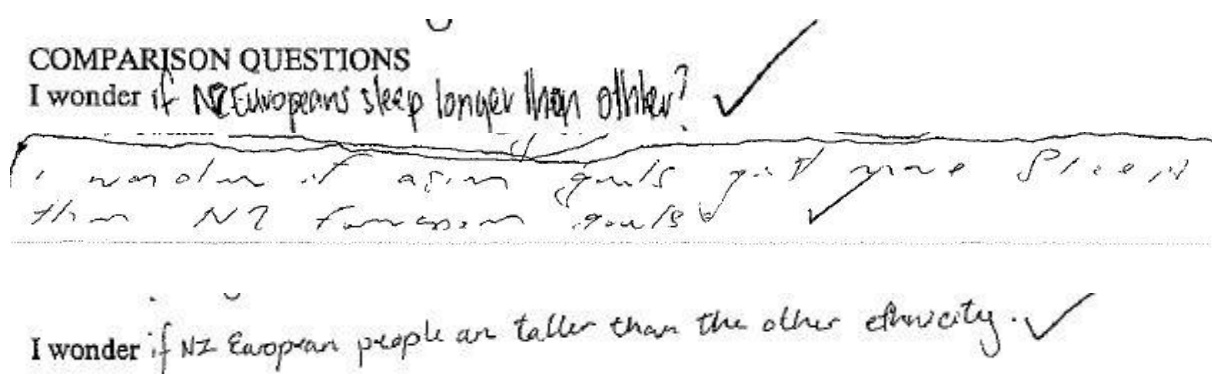


Figure 5-2. Examples of student-posed questions where the sample size was too small

The problem was that for the three comparison questions in Figure 5-2, the particular groups chosen to compare – Other or Asian girls – had five and two students in the given data set,

respectively. For the data set given, these questions were in fact not suitable due to sample-size issues and should never have been marked as suitable by the teacher. This was definitely a case of a gap in teacher statistical content knowledge. At this stage the initial problematic situation was clarified and formulated: What makes a good statistical question and what statistical content and conceptual knowledge is needed to support students to successfully pose statistical questions from given data sets?

Hence an extensive review of the literature was conducted about “What makes a good statistical question?” As was shown in the literature review (chapter 2), very little research had been done in the area of statistical questions. From the researcher’s thinking about the issue and her consequent classification of the different purposes of questions, the problem being dealt with would be categorised as posing investigative questions.

### **5.5. Retrospective analysis of investigative questions**

With this new insight, the questions in the pre- and post-test and from the first two teaching lessons were re-analysed, initially looking for patterns in student-posed questions. What became clear reasonably quickly was that a lot of the summary questions posed asked “How many [of a particular category]?” or “What was the most popular?” In addition the grade allocation descriptors for questions in the pre- and post-test didn’t capture the whole story. The researcher, along with her supervisor, set about classifying the investigative questions into categories, looking at both summary and comparison questions. This classification will now be discussed.

#### **5.5.1. Classifying summary investigative questions**

As the summary questions were sorted, five categories emerged. These included questions such as “How many ...?” and “What was the most popular?”, and also questions that asked who was the tallest/shortest, what was typical, and others that could not be answered with the data in hand. Additional categories developed as the “How many ...” questions could be about a particular category or within a specified range and the “What was typical” question could also reflect the population for which conclusions could be drawn. Finally eight categories were realised for summary questions. Of these, seven were evident in the class discourse and in the test-posed questions, while the eighth category was conjectured using Graham’s (2006) definition. The summary question categories are given in Figure 5-3 (next page) (Arnold, 2008b) and beside each is an example of a student-posed question, the

exception being category 7 where there were no examples so a hypothesised question is given in *italics*.

|                       | Summary question category |                                                                                                                                                    | Student example                                                           |
|-----------------------|---------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Non-summary questions | 1.                        | Nonsense or not a summary question.                                                                                                                | Is the Asians circled above related?                                      |
|                       | 2.                        | A question that is partially related to the data, but not answerable by the given data.                                                            | What region is the most preferred to live in?                             |
|                       | 3.                        | A question that asks about an individual case.                                                                                                     | I wonder which person has the biggest neck size?                          |
| Pre-summary questions | 4.                        | A question that asks how many of a particular category.                                                                                            | I wonder how many people look up to their family.                         |
|                       | 5.                        | A question that asks how many within a specified range.                                                                                            | I wonder how many Year 6 girls go to bed in between 8:30 p.m. and 10 p.m. |
|                       | 6.                        | A question that asks for the most popular or most common.                                                                                          | I wonder what is the most popular ethnicity.                              |
| Summary questions     | 7.                        | A question that asks about the overall distribution of the data or what is typical.                                                                | <i>What is the typical way students usually travel to school?</i>         |
|                       | 8.                        | A question that asks about the overall distribution of the data or what is typical and reflects the population for which conclusions can be drawn. | I wonder what a typical time a Year 10 student goes to bed.               |

**Figure 5-3. Summary question categories**

These eight categories are hierarchical and naming them all as summary question categories is actually misleading as not all of these types of questions would be considered summary questions. It was proposed at this stage that categories 1 to 3 were non-summary questions, categories 4 to 6 were pre-summary questions, and only categories 7 and 8 would be considered actual summary investigative questions.

At the school level the investigative questions that students are posing need to be clear in terms of their purpose and intent as the novice statistician is unlikely to be able to infer about a wider universe from poorly formed pre-summary questions, whereas an expert statistician will be able to.

### 5.5.2. Classifying comparison investigative questions

The comparison questions sorted more easily, into five broad categories. The categories are given in Figure 5-4 (next page) (Arnold, 2008b) and beside each is an example of a student-posed question, the exception being category 5b where there were no examples so a hypothesised question is given in *italics*. The five broad categories are hierarchical, but category 4a and 4b are the same level as are 5a and 5b. There was no pre-comparison question category, only questions that could not be answered due to sample size or at least one of the variables not available in the data set.

|                | Comparison question category |                                                                                                                                                                      | Student example                                                                                                                                                    |
|----------------|------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Not answerable | 1.                           | Nonsense or not a comparison question.                                                                                                                               | Is there more 15 year olds or more 14 year olds?                                                                                                                   |
|                | 2.                           | A question that is partially related to the data, but not answerable by the given data.                                                                              | I wonder if the students in Wellington take longer to get to school than students in Auckland. (Region was not a variable in the data set but time to school was.) |
|                | 3.                           | A question that is related to the data but not answerable due to sample size issues.                                                                                 | I wonder if Asian girls have a longer arm span than Indian boys. (Two Asian girls and two Indian boys in the data set.)                                            |
| Answerable     | 4a.                          | A question that is answerable by the data.                                                                                                                           | I wonder if boys are taller than girls.                                                                                                                            |
|                | 4b.                          | A question that is answerable by the data and requires recategorisation of categories to be compared.                                                                | I wonder whether people who have the longest time travel get less sleep than those who travel for less time.                                                       |
|                | 5a.                          | A question that is answerable by the data and reflects the population for which conclusions can be drawn.                                                            | I wonder if Year 10 boys have a longer arm span than Year 10 girls.                                                                                                |
|                | 5b.                          | A question that is answerable by the data and requires recategorisation of categories to be compared and reflects the population for which conclusions can be drawn. | <i>I wonder if Year 10 New Zealand European students are taller on average than Year 10 students of other ethnicities.</i>                                         |

Figure 5-4. Comparison question categories

### 5.5.3. Results from student-posed questions

Using these new classifications the students' questions were now recategorised.

#### **Summary questions**

Of the 13 students who sat the post-test, only 11 students had both a pre- and post-test classification for summary questions; the 11 students' results are summarised in Table 5-2 (next page). In the pre-test the student grade is based on one question, while in the post-test the grade is based on three questions, so the mean (0 d.p.) is given.

**Table 5-2. Pre- and post-test summary question categories**

|                             |   | Post-test summary categories |   |   |   |   |   |   |   | Total<br>Pre |
|-----------------------------|---|------------------------------|---|---|---|---|---|---|---|--------------|
|                             |   | 1                            | 2 | 3 | 4 | 5 | 6 | 7 | 8 |              |
| Pre-test summary categories | 1 | 1                            | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 4            |
|                             | 2 | 0                            | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2            |
|                             | 3 | 0                            | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0            |
|                             | 4 | 0                            | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 3            |
|                             | 5 | 0                            | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0            |
|                             | 6 | 0                            | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 2            |
|                             | 7 | 0                            | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0            |
|                             | 8 | 0                            | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0            |
| Total Post                  |   | 1                            | 0 | 3 | 5 | 2 | 0 | 0 | 0 | 11           |

Six questions posed in the pre-test were classified as non-summary questions and five as pre-summary questions. In the post-test, four questions were classified as non-summary and seven as pre-summary. The majority of the questions in the post-test were “How many ...?” questions. Five students “improved” the category of their summary question from the pre-test to the post-test, three remained the same and three regressed. It is noteworthy that none of the questions posed by the students were classified in the summary investigative question categories.

### ***Comparison questions***

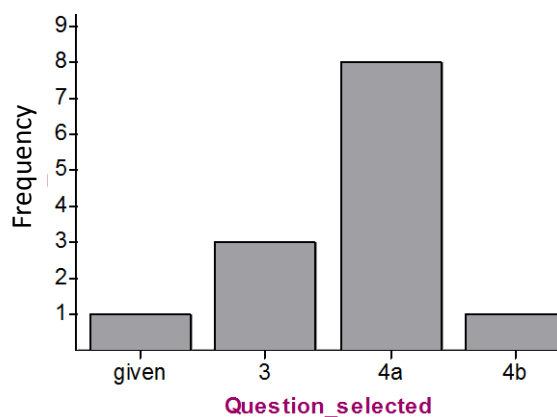
Thirteen students sat the post-test and all of them posed comparison questions. Since only two of the 13 students posed comparison questions in the pre-test, a pre-/post-test comparison is not appropriate. Hence the types of comparison questions the students posed in the post-test only are analysed. Each student posed three comparison questions in the post-test, 39 questions altogether, which are summarised in Table 5-3 (next page).

**Table 5-3. Post-test comparison question categories**

| Post-test comparison category | Number of questions in the post-test that were this category |
|-------------------------------|--------------------------------------------------------------|
| 1                             | 11                                                           |
| 2                             | 1                                                            |
| 3                             | 11                                                           |
| 4a                            | 14                                                           |
| 4b                            | 1                                                            |
| 5a                            | 1                                                            |
| 5b                            | 0                                                            |
| Total                         | 39                                                           |

Sixteen of the 39 post-test comparison questions were suitable (categories 4a, 4b and 5a) based on this current question classification. Eleven additional questions posed were marked as suitable (category 3) by the teacher, but these 11 questions are in fact not suitable due to at least one of the groups selected having too few in the sample. When considering all the questions posed by the students in the post-test, 41% were suitable comparison investigative questions (categories 4a, 4b and 5a) and 59% were unsuitable (categories 1 to 3).

The questions were marked by the teacher before the students continued with the investigation. The category of question that the students used to complete their post-test is shown in Figure 5-5. Category 3, 4a and 4b questions were marked as suitable by the teacher. Twelve of the 13 students had questions that had been marked as suitable to use. However, as Figure 5-5 shows, while nine of these were comparison questions (categories 4a and 4b), three were category 3 and therefore not answerable due to sample-size issues. The “given” category is for the one student who was given a question so that she could complete her investigation because her questions were all category 1 and so not suitable.

**Figure 5-5. Category of question selected to complete the post-test**

When the 11 category-3 posed questions from the post-tests were explored further, it was found that these questions were posed by six different students. However, only three of these six students actually used their category-3 questions in their investigation, while the other three used a category-4a question to do their investigation (as shown in Figure 5-5). The three students who selected a category-4a question to complete the test had posed two category-3 questions and one category-4a question and selected their single category-4a question. It is not known whether these students started with category-3 questions and moved to category-4a questions when they struggled to complete the investigation. Of the remaining three students, two students had posed two category-3 questions and one category-1 question; both of these students used a category-3 question to complete the test. The final student had posed two category-4a questions and one category-3 question; they had then chosen the category-3 question and, as a result, struggled to complete the investigation. This misconception by the teacher had led to potential issues for the students.

### ***Teacher examples of question posing in class lessons***

Posing questions was explicitly addressed in the first two lessons in the following ways. The teacher gave examples of one category-4 summary question in the first lesson and two category-8 summary questions in the second lesson; she also gave examples of two comparison questions (4a and 5a) in the first lesson. In the second lesson the teacher reviewed the questions recorded from the first lesson and looked at rewriting the questions so that they would come up with a summary question to investigate. In the following excerpt she takes a category-4 summary question from the previous day and changes it to a category-8 summary question, but it should be noted that she did not address why the original question was unsuitable.

*T1, 2007: Today we are going to look at how we can answer these questions, but before we do that we want to make sure we have a good question to start with so we are going to re-write the questions just a little bit. Let's look at the question here: "I wonder if all year five students go to sleep at 7 p.m.?" Can anyone think how we could reword that to make it a little easier to investigate or think of another way we could say that? Maybe we want to know what a typical time year 5 students go to bed – let's see how we re-write this. How about we wonder what a typical time a year 5 student goes to bed is?*

## 5.6. Discussion

The retrospective analysis associated with this first teaching experiment addresses the different categories of investigative questions (summary and comparison) students are posing at the beginning of a statistical investigation based on a given multivariate data set. The results suggest that most of the students can pose comparison questions but none were successful at posing summary questions. Of the 13 students who sat the post-test, 10 of them posed at least one suitable comparison question of the three comparison questions they posed, and although none of the students had posed a suitable summary question, five of them had posed predominantly pre-summary questions. These questions varied from the “How many [of a particular category]?” through to “What was the most popular [category]?”.

The teacher modelled both a pre-summary question and two suitable summary questions within the initial two lessons; however, she didn’t address what made a good summary question with the students. It would appear that modelling the questions without addressing what made a good question was insufficient to facilitate a major change in the category of summary questions that students were posing. Students have been exposed to “summary-type” questions for a number of years. Predominately these have been textbook examples which usually involve a given data display and then a series of questions of the pre-summary type to answer. There seems to be no examples in current New Zealand textbooks requiring students to pose the question to answer, or even the situation where the given question to answer is a summary question. The types of questions asked in textbooks are “read the data” and “read between the data” (Friel et al., 2001, p. 130) types of questions. These questions match with the pre-summary posing questions. I was unable to find any evidence of “reading beyond the data” (Friel et al., 2001, p. 130) types of questions that match with the summary investigative questions.

All these questions (read the data, read between the data, and read beyond the data (Friel et al., 2001)) are valid within the statistical investigation cycle. However, there is a distinction between question *posing* and question *asking*: the summary investigative questions involve an aggregate view and need to be *posed* both by teachers and students, whereas the individual case questions and the pre-summary questions need to be *asked* by students when they are writing their analysis.

The teacher modelled suitable comparison questions in class, but again at this point there was no realisation that the rationale for these questions should be addressed. Generally the



modelling was of the category of questions that the students used in the post-test. Comparison questions are new for nearly all of the students at this level. It would appear that having less previous exposure to comparison questions means the students predominately use the model they had seen in class. The one issue that came up with the comparison questions was students selecting a group for comparison that had too small a sample size. This was the situation for three students, although the teacher didn't identify these categories of questions as being not suitable.

When comparing the findings of this first teaching experiment with Franklin and Garfield's (2006, p. 353) framework, most of the student-posed summary-type questions would be at Level A: "beginning awareness of the statistical question distinction", whereas most of the comparison-type questions would be at Level C: "students can make the statistics question distinction". However, while Franklin and Garfield's (2006) framework provides a holistic view of the statistical investigative cycle, it is too broad to use to determine the categories of the questions posed; for example, it does not distinguish between questions that can be answered with the given data, and questions that cannot. Furthermore, it does not appear to be based on how students think.

The key finding from this first teaching experiment was a descriptive framework for the classification of the *investigative* questions (summary and comparison) that students pose. Based on this teaching experiment and the subsequent literature review in chapter 2, it is conjectured that a number of moderating questions would be useful to consider when posing an *investigative* question from a given multivariate data set (secondary data):

- What was the original question that was used to collect the data (survey question)?
- What type of data is being used?
- What graph or display of the data will be made?
- What hypothesis can be made about the data?
- Is the question interesting?
- Who would be interested in the answers to this question?
- Is there enough data available to answer the question (issues around sample size)?
- What background information is available about the data (how it was collected, who it was collected from, when it was collected, etc.)?
- Is the variable of interest in the data set?

All of these issues might help to answer the question “Is the *investigative* question ‘right’?”

Hence further work is needed to explore the relevance of these issues raised in relation to posing questions for investigation. More work is needed to consolidate the framework of the types of investigative questions students can pose (summary and comparison) and to build a framework for relationship questions. Other questions to consider include:

- Are the data continuous or discrete and does this matter?
- Is the question going to be about the sample or about the population?
- What sort of modelling by the teacher is required?
- What help do the teachers need before being able to do this?
- What are the links between the questions posed by the students and their analyses and conclusions?

As Konold and Higgins (2003, p. 212) said: “Helping students raise questions that interest them and that they can productively pursue is a challenge for the teacher.”

### 5.7. Summary

Posing questions in the statistical enquiry cycle has been part of the New Zealand curriculum landscape since the implementation of NCEA assessments in 2001. Questions posed were based on naïve understandings of what a good statistical question was. This naïvety became abundantly clear when the teacher and researcher reviewed the post-test responses of the students in this first teaching experiment. Statistical content and conceptual knowledge around posing investigative questions was identified as a major concern at the schooling level and provided the catalyst for the direction of this doctoral thesis.

At this early stage on the journey, summary and comparison investigative questions were placed into different categories using a hierarchical linear classification system. Moderating questions were suggested and a number of issues were raised. In fact, there are more questions than answers, such as the question “What makes a good *investigative* question?” Answers to this question are explored further in the next chapter.

## **Chapter 6. What Makes a Good Investigative Question**

### **6.1. Introduction**

The role that posing investigative questions plays in assessment for qualifications in New Zealand and the identification of a lack of teacher knowledge in this area has highlighted “What makes a good investigative question?” as a problematic situation. Teachers and assessors need to know what makes a good investigative question at the school level, the components and concepts underpinning a good investigative question, and the learning that students need to be immersed in to support their posing good investigative questions. A “good” investigative question is one that allows for rich exploration of the data in hand, discovery, and thinking statistically. A good investigative question has an element of open-endedness and students are engaged in interesting work. A good investigative question is not too restrictive and is crucial in the teaching and learning process at the school level.

This chapter starts by outlining the overall journey that was taken to answer the question “What makes a good investigative question?” based on the initial problematic situation that was identified in chapter 5. The specific planning and preparation for the various teaching experiments is detailed, providing context and rationale for the different decisions made about teaching and learning activities. The retrospective analysis that was undertaken is described and findings from this are given, including criteria for what makes a good investigative question. A classification framework for investigative questions is proposed and this forms the basis for the analysis of investigative questions. This framework grew as subsequent teaching experiments exposed unforeseen and innovative investigative questions posed by the students. Finally, using the classification framework, a quantitative analysis of student responses is conducted in order to determine: (1) the progress of the students from pre- to post-test, and (2) in what areas the overall quality of the investigative questions posed improved from 2008 to 2011 in cognisance of the research development incurred as a result of the teaching experiments.

### **6.2. Problematic situation**

Posing investigative questions was identified in the first teaching experiment as a problematic situation that was in need of further exploration (see chapter 5). In the second teaching

experiment this problematic situation was addressed and there was a focus on ensuring that the variable and the target population were clear in the question and that the question was asking about “some type of relationship or comparison” (T1, 2008, lesson 2). At the end of the second teaching experiment two further issues arose, one was around language and the other was around contextual and conceptual knowledge underpinning posing investigative questions. In the third teaching experiment criteria for what makes a good investigative question were used and underlying conceptual knowledge needed to understand the investigative question was focused on in the teaching and learning. By this stage the hierarchical linear classification system for investigative questions (see chapter 5) proved to be too simple and not able to capture the subtleties of the population description in questions and the quality of the question posed. As a result a two-way classification matrix (framework) was developed.

The problematic situation involves three research questions: (1) What makes a good investigative question? (2) What are the underpinning concepts that are needed to support the teaching and learning around posing investigative questions? and (3) What level of comparative investigative questions are year 10 (ages 14–15) students posing?

### **6.3. Planning and preparation**

This section discusses the planning and preparation for the final three teaching experiments with a focus on posing investigative questions. The specific activities and actions that were undertaken are described to show the development of the teaching and learning progression for developing good investigative questions and how this links to the problematic situation.

#### **6.3.1. Planning and preparation for teaching experiment 2 (2008)**

Planning and preparation for the second teaching experiment expanded the work from the previous year (2007) and reflected the requirements for the students in the 2008 class – a class that was working at a level that was one year ahead of the rest of their age group. Of note in this experiment was the deliberate teaching focus on posing investigative questions. Adaptations specific to posing investigative questions were made to the teaching and learning activities and to the pre- and post-tests.

***Adaptations to teaching and learning activities – teaching experiment 2***

In summarising questions within the statistical investigation cycle (section 2.7.3, page 39), three points were noted: posing investigative questions require students and teachers to have a clear idea of what the variable(s) are that they are interested in, what they want to do (summarise, compare or relate), and what the population of interest is. The teacher also had three criteria that the school she was working in had been using for some time to support students in posing questions for the national assessment in year 11 (ages 15–16) and these criteria aligned with the points noted above. The teacher's criteria were:

1. reference to the target population
2. reference to the variable being measured
3. some kind of relationship/comparison to be investigated.

The planning involved deliberately teaching these criteria to the students and providing sufficient examples to allow them to practise with a number of different variables and populations.

***Adaptations to the pre- and post-tests – teaching experiment 2***

The pre- and post-tests were changed for the second teaching experiment. The tests focused on three parts of the PPDAC cycle – posing investigative questions, analysis and conclusions – and were designed so that each section could be given out separately. In terms of posing investigative questions, the changes were (for both the pre- and post-test):

**Section on posing investigative questions**

- The data table given was part of a larger sample, rather than the whole sample which had been the case previously, so that sample size was not an issue for the groups that the students chose to compare.
- The numbers of categorical variables were restricted in an attempt to limit the number of categories the students might use when defining their groups to compare; for example, right-handed girls and left-handed girls, 15-year-old Asian boys and 16-year-old Asian boys.

- The population that the sample had come from was repeated in the test question that asked the students to pose investigative questions. Previously it had only been mentioned in referring to the table of data, but not in the actual prompt for posing investigative questions. This is the broader New Zealand CensusAtSchool population, but the students still had to sort out the specific subgroup that was sampled from; for example, year 11. It was felt that only having it with the data table made it harder for students to make the connection with the population they were posing investigative questions about.
- Students were asked to pose three summary-type questions, three comparison-type questions and three relationship-type questions. This allowed for the three different types of questions to be posed. Previously they had posed only summary-type and comparison-type questions; the inclusion of relationship-type questions was to cover the range required for the curriculum, and for this teaching experiment the decision had not yet been made to focus on summary and comparison only.

### Section on analysis

- Students were not expected to graph the data as graph construction (including sample size and scale issues) had previously been a barrier to determining students' ability to reason from the data.
- “Good” models of investigative questions were used in the analysis section where students were asked to describe distributions and draw conclusions from given questions and graphs. These investigative questions had been posed through dialogue with the researcher's supervisor.

### 6.3.2. Planning and preparation for teaching experiment 3 (2009)

Between the second and third teaching experiments there was extensive dialogue between the researcher and colleagues at the university based largely on the retrospective analysis of student responses in the post-test. This dialogue addressed language, and in particular, the use of the small words like *a* and *the* in investigative questions and the implications of these as to which group the question was about. Through this dialogue and through analysis of student responses, particularly poorly posed investigative questions, ideas of suitable criteria for “What makes a good investigative question?” were generated. The researcher trialled some of the ideas with a year nine (ages 13–14) class at another school. This was not recorded as it was not part of the research for which permission to video was granted, but it did provide an

opportunity to trial some of the material before using it with the class in the third teaching experiment.

***Adaptations to teaching and learning activities – teaching experiment 3***

The teaching and learning activities around posing investigative questions in this third teaching experiment built on the work from the second teaching experiment.

Students posed investigative questions in class, similar to what they did in the pre-test, and then a selection of these was used in the following lesson. During that lesson the students grouped the questions and explained how they had grouped them; the teacher then elicited from the students which questions they thought were investigative questions and which ones not, and through the discussion brought out a number of criteria that could be considered when posing investigative questions. An additional activity was built in later in the topic where students critiqued questions that had been posed and improved them based on the developed criteria.

During the work on using samples to answer investigative questions about populations, care was taken to reinforce the actual population about which the students were posing and answering investigative questions. Instead of the CensusAtSchool database being used, a fictitious school was invented (discussed fully in section 7.4, Identifying and clarifying the population, page 151) and data cards for each “student” were created to help to develop the concept of population and sample. The “population”, Karekare College students, was constantly referred to, and this population was also physically shown as the data cards in a bag. This material representation of the population, coupled with the actual drawing of samples from the bag, was designed to reinforce the connection between sample and population and the investigative question.

***Adaptations to the pre- and post-tests – teaching experiment 3***

The pre- and post-tests were changed for the third teaching experiment. In terms of posing investigative questions, the changes were (for both the pre- and post-test):

- The students were asked to pose six investigative questions that they could answer with the data and then asked to go back and label them as summary, comparison or relationship questions. The idea was to not restrict their question posing, but to still see if they could identify the questions according to the different types.

For the post-test only:

- The population that the students were posing questions about changed from year 11 in the pre-test to secondary students (years 9–13) in the post-test. This was to better reflect the nature of the population that they had been working with in class and also to counter the posing of questions about a particular age group.
- A number of investigative questions were given. For each question the students had to comment on whether they thought the question was a good investigative question or not. They were asked to give reasons as to why or why not. If the question was not a good investigative question, the students were asked to change it to make it a better investigative question. The reason for the addition of this section was to see which, if any, of the criteria for what makes a good question had become part of the students' knowledge base.

### **6.3.3. Planning and preparation for teaching experiment 4 (2011)**

In the fourth and final teaching experiment the experiences from the previous three teaching experiments were refined.

#### ***Adaptations to teaching and learning activities – teaching experiment 4***

The teaching and learning sequence did not start with posing investigative questions, as it had previously, and the various types of questions (summary and comparison) were dealt with separately as they became necessary through the statistics unit.

When doing entire investigations using the PPDAC cycle, the number of populations used in the examples was broadened. Previously this had been restricted to a couple of situations and it was felt that this narrow selection prevented students from fully grasping the idea of what the population was for a particular sample.

#### ***Adaptations to the pre- and post-tests – teaching experiment 4***

The pre- and post-tests were changed for the fourth teaching experiment. In terms of posing investigative questions the changes were (for both the pre- and post-test):



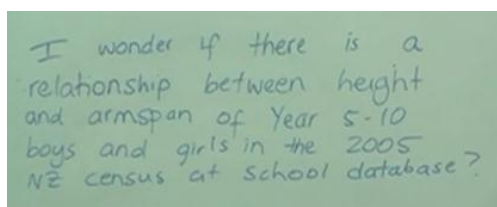
- Students were asked to pose three summary-type questions and three comparison-type questions. Relationship-type questions were left out as these were no longer a focus in the research. The format of the question went back to asking for the specific types of questions because in the free situation from teaching experiment 3 there were issues. For example, the students incorrectly labelled the questions or did not label them at all, and many times there were not examples of a particular type of question which made it difficult to classify or grade the students' question-posing level.
- In addition, the following detail was added into the question: "The population that the sample comes from and the variables available are detailed in the data sheet." The purpose here was to invoke the concept of posing investigative questions about a population.

## 6.4. Teaching experiments

This section describes teaching moments that relate to posing investigative questions and are pertinent to the outcome of the research; for example, deliberate teacher acts, "aha" moments for teacher and researcher, or particular student insights or actions.

### 6.4.1. Teaching experiment 2

In the second teaching experiment, the teacher (T1, 2008) deliberately discussed and highlighted the three criteria (section 6.3.1, page 98) for posing good investigative questions. She first discussed them with the students in the second lesson when posing relationship investigative questions. The transcript below of part of the lesson captures this discussion. The discussion is based on the investigative question given in Figure 6-1. When the teacher wrote notes after the discussion, comparison situations were included in the examples given.



I wonder if there is a relationship between height and armspan of Year 5-10 boys and girls in the 2005 NZ census at school database?

Figure 6-1: Relationship question posed in class

*Teacher: This is a good question, so we'll just go over why that's a good question. So, first of all, it helps to start with the starter "I wonder". It's a relationship question, so you want a **relationship sort of word**, "Is there a relationship?" Usually, we've got two variables – two things we're looking at. So **you need to mention both of***

*those two variables, or those two things that you're actually investigating. So some of you would have height and arm span, some of you would have what?*

*Student: Neck circumference, wrist circumference.*

*Teacher: Neck circumference, wrist circumference. Okay, now, you can't just say "of boys and girls". Why can't I just say "of boys and girls"?*

*Student: 'Coz that's everyone.*

*Student: There are 11-, 12-, 13-, 14-, 15-year-old boys and girls in the class, and we're looking at year 5–10 boys and girls.*

*Teacher: Good. So we can't say that, based on our sample, we know what the height of year 13s are, can we? So you need to actually say what your sample was, or **what the population was where it came from** was year five to year 10 students, and that was boys and girls. The other thing is, where the data is ... so the data came from 2005 New Zealand CensusAtSchool database.*

The three criteria that the teacher refers to are: (1) using a relationship (or comparison) word to show whether it is a relationship or comparison question; (2) mentioning the variables; and (3) noting what the population was that the sample came from.

The teacher decided to spend one lesson sorting, critiquing and improving investigative questions that had been posed by others. This involved the students first sorting the questions into the different types (summary, comparison and relationship) and then improving the investigative questions by making sure the investigative questions met the three criteria given by the teacher. In this lesson a number of points were mentioned that have subsequently been linked to posing a good investigative question or understanding the question they have posed.

- The teacher made mention a couple of times during the lesson about the question being worth investigating. This links to Graham's (2006) second consideration (see page 28).
- The actual variable that could be investigated was clarified; for example, they weren't investigating foot size; they were investigating right foot length.
- Using comparing words when posing comparison questions to make it clear what type of question it is; for example, using longer, taller or faster. Linked to this was the use of the appropriate comparing word; for example, use longer for right foot length, but not for right foot width (in this case you would use wider).

- The teacher started to explore how students might go about answering their question, making links to the types of graphs and/or statistics that they might use. This relates to understanding the types of questions and therefore the analysis that would be appropriate for the given situation.

Posing investigative questions became integral throughout the topic – any investigation started with an investigative question. The PPDAC cycle was overtly used and the teacher pointed to the specific aspect they were working on, on the PPDAC poster (see Figure 1-2, page 2), as the lessons and topic progressed. The criteria for what makes a good investigative question were specifically covered in three lessons: the second lesson to initially give the criteria, and then in the fourth and fifth lessons the criteria were reviewed as part of the delivered lesson. When writing conclusions, the teacher also made the connection clearly back to the investigative question a number of times; for example, “You have to relate your conclusion back to the question and you have to have a good question in the first place” (2008 teacher, lesson 4).

### 6.4.2. Discussions with colleagues post teaching experiment 2

Discussions with colleagues at the University of Auckland also contributed to what should be considered when developing investigative questions for statistical enquiry. These discussions ranged across the topics of language and underlying concepts, such as sample and population, and notions of individual and aggregate (Bakker & Gravemeijer, 2004; Konold & Higgins, 2002, 2003).

The language used by the teacher and students in posing investigative questions needs to be tight and precise. From the second teaching experiment it became apparent that this still needed further work. Review of student-posed investigative questions with colleagues highlighted two language issues. Firstly, it seems that it is very common to drop in *a* or *the* in the question, which can very quickly change the whole meaning of the question. Using *a* or *the* before boys, for example, can change it from being about the population (boys – note no article used) to an individual (*a* boy) or to the sample (*the* boys). Similarly, when used in front of the word typical, the question can change from asking about the aggregate (typical heights) to asking about a singular measure (*the* typical height). Secondly, what the students were defining as typical varied. In a statistical investigation *typical heights* (or arm spans or methods of transport), i.e. typical variables, are explored – not *typical boys* or *typical girls*.

Table 6-1 highlights examples of these issues in student-posed questions in the 2008 post-test.

**Table 6-1. Student-posed investigative questions highlighting aspects of language**

| Student-posed question                                                                                                                                           | Commentary                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I wonder what <i>the</i> typical hair length is for <i>a</i> year 11 girl from the 2007 NZ CensusAtSchool database? (2008 student, post-test)                    | <ul style="list-style-type: none"> <li>• Uses <i>the</i> for typical – this is interpreted as a singular view not as an aggregate view.</li> <li>• Uses <i>a</i> – are they meaning an individual?</li> </ul>                                                                                                                                                                                                                                                                                                                                                                            |
| I wonder what <i>the</i> typical height is for <i>a</i> year 11 girl? (2008 student, post-test)                                                                  | <ul style="list-style-type: none"> <li>• Uses <i>the</i> for typical – this is interpreted as a singular view not as an aggregate view.</li> <li>• Uses <i>a</i> – are they meaning an individual?</li> </ul>                                                                                                                                                                                                                                                                                                                                                                            |
| I wonder if most of <i>the</i> boys are taller than 160 cm? (2008 student, post-test)                                                                            | <ul style="list-style-type: none"> <li>• Uses <i>the</i> – this is interpreted as the boys in the sample, i.e. the boys given.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| I wonder if <i>the</i> boys are taller than <i>the</i> girls, <i>who were selected from</i> the 2007 CensusAtSchool? (2008 student, post-test)                   | <ul style="list-style-type: none"> <li>• Uses <i>the</i> and qualifies with <i>who were selected from</i> – referring to the sample.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| I wonder if <i>the</i> boys <i>sampled</i> in the 2007 NZ CensusAtSchool have a longer arm span than <i>the</i> girls <i>sampled</i> ? (2008 student, post-test) | <ul style="list-style-type: none"> <li>• Uses <i>the</i> and qualifies with <i>sampled</i> – referring to the sample.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| I wonder what <i>the range</i> of the height is for year 11 boys in the 2007 CensusAtSchool database? (2008 student, post-test)                                  | <ul style="list-style-type: none"> <li>• Uses <i>the</i> for range – this is interpreted as a singular measure and not as an aggregate view.</li> <li>• Uses <i>range</i> here – this is interpreted as the “range”, i.e. highest value minus lowest value because of the use of <i>is</i>.</li> <li>• There is an unanswered question here: Was the student meaning something broader than just the value of the range, almost like a distribution idea?</li> <li>• [Note: It is the “<i>is</i>” more than the “<i>the</i>” that indicates singular and height not heights.]</li> </ul> |
| I wonder what is <i>the</i> height of <i>the typical</i> NZ year 11 student? (2008 student, post-test)                                                           | <ul style="list-style-type: none"> <li>• Uses <i>the</i> for both height and typical – this is interpreted as meaning an individual student.</li> <li>• Uses <i>the typical</i> – interest is in typical heights, not typical students, so misconception about what is being investigated.</li> </ul>                                                                                                                                                                                                                                                                                    |
| I wonder what <i>the</i> height for <i>a typical</i> NZer 15 year old boy is? (2008 student, post-test)                                                          | <ul style="list-style-type: none"> <li>• Uses <i>the</i> for height – this is interpreted as a singular view not as an aggregate view.</li> <li>• Uses <i>typical</i> – what is typical? In this instance the interest is in typical heights, not typical New Zealand 15-year-old boys.</li> <li>• Uses <i>a</i> – are they talking about an individual here? A typical New Zealander 15-year-old boy?</li> </ul>                                                                                                                                                                        |

The use of the typical height stems back to initial work by the researcher to come up with an improved format for summary questions. Previously the students had asked investigative questions such as “What is the average height?”, but this was felt to be too restrictive as it is fundamentally asking about a single statistic or singular view. The migration to the word *typical* was an attempt to broaden the perspective to a more aggregate view. Typical in the summary situation was used to mean an overall distribution idea with key features included. Upon reflection, the initial wording that was used as examples was classified as also being of a singular view. For example, in the pre-test in 2008 the students were given a summary

question to explore. The question was: “I wonder what is the typical height of NZ year 11 students?” Clearly this example and the subsequent post-test example, “I wonder what is the typical right foot length of NZ year 11 students?” are, in hindsight, less than ideal model examples for students.

During analysis of student questions from the second teaching experiment, and through discussion with colleagues, the issue was identified and defined and the wording was improved. The new and improved wording that was used in subsequent teaching experiments for summary questions was “I wonder what are typical heights of NZ year 11 students?” The plural format was aiming at the broader distribution or aggregate idea. However, it needs to be made clear that one template for summary questions is not being proposed – that would lead down a very predictable and very wrong path. The question structure proposed now allows a focus on the key aspects of a good investigative question and hence allows people to pose other good summary investigative questions.

### **6.4.3. Teaching experiment 3**

In the third teaching experiment, the teacher (T2, 2009) had been exposed to much deeper thinking about posing investigative questions prior to teaching the statistics unit. This had included workshops for all the teachers in the school on the material and the teacher was a member of the linked TLRI research project team.

In the second lesson, where the teacher had given the students questions that they had previously posed to sort, the students came to the conclusion that they did not like most of their questions by the end. They felt that the questions were not suitable as investigative questions due to various reasons including that the question was not able to be answered due to the variable not being available in one question, there was not enough data to answer another question, and that some of the questions were about an individual and not the whole group and they felt this was unacceptable.

As the teacher moved into new concepts, such as sampling, she started always with an investigative question which was posed collectively as a class and checked against the criteria that had been established.

In a wrap-up session the students again came back to the criteria about what makes a good investigative question and, as well as posing investigative questions themselves, they had to critique questions posed by others. During this activity, an interesting observation was made

by one of the students to another student in the group that was being observed: “Have you noticed that all the good ones are really long?” (2009 student, final lesson).

### 6.4.4. Teaching experiment 4

In the final teaching experiment the teacher’s (T2, 2011) approach to posing investigative questions was different to the previous experiments. She gave the students questions that had been posed by others before she required them to pose their own. An activity where the students had to sort a number of investigative questions into groups provided a catalyst to talk about what questions were good questions and what questions were not. From this discussion some of the criteria that had previously been established by the research were re-established by the students. That is, the students and teacher developed the criteria based on the class discussion about the questions they were sorting. Criteria that the students came up with included that the question needs to be about the overall distribution of the data, it must be interesting, and the variable and group need to be stated. Student reflection at the end of the lesson elicited a further criterion that had not been mentioned in class: that the type of question needed to be clear. At this point the teacher resisted the urge to “finish” the criteria and settled to leaving further criteria because they naturally arose in the teaching and learning sequence.

Defining the context, i.e. the variable and the population, became a focus and throughout the unit the teacher constantly asked the students to define the variable and the population for each situation. This was also linked to moving from questions about “these” students (the sample) to questions about the population. An example of defining the variable comes from one of the later lessons where students were exploring some situations where CensusAtSchool participants had ranked themselves as to how good they thought they were at a particular subject; for example, maths, reading, sport and the arts. The discussion is around exactly what the variable is, i.e. is it boys rating themselves higher than girls rate themselves, or is it boys rating themselves as better when they compare themselves to girls?

*Teacher: The question they were asked was how good do you think you are at maths. That was the question that they were asked. That was the survey question. ... How good do you think you are at maths? So remember we’re comparing the boys and the girls. So when we’re posing an investigative question we’re looking at the first one, so those were the survey questions. The investigative question can someone give it to me, the first one?*

## Chapter 6 – What makes a good Investigative Question

*Student: I wonder whether boys tend to think that they are better at maths than girls. Year 4–13 boys and girls.*

*Teacher: Year 4 to 13 New Zealand boys tend to think ...*

*Student: They are better at maths than girls.*

*Teacher: They are better at maths in this case than year 4 to 13 New Zealand girls.*

*Student: No not think they are. Because the boys wonder if they're better than the girls.*

*Teacher: Remember the question wasn't "Are you better than girls?", it's just how good you think you are so it's not rating against the other. But in the overall rating.*

*...*

*Teacher: What did we say up there? Boys rate themselves better at maths than girls. The boys aren't rating themselves compared to girls, it's just when they rate themselves, boys' ratings tend to be higher than girls' ratings. So the question could have been: "I wonder whether ratings for maths ability by year 4 to 13 New Zealand boys tend to be higher than ratings for maths ability by year 4 to 13 New Zealand girls."*

In addition to the discussion regarding how to frame or describe the variable, it is worthwhile noting that by this fourth experiment the teacher was clearly differentiating between the two types of questions that are posed, i.e. survey questions and investigative questions. It is also worth noting the use of the phrase “tend to” for comparison questions. This phrasing has become part of the teacher’s natural language she is using in relation to comparison questions.

The teacher repeatedly came back to the difference between the three different types of questions, resulting in a lesson about two-thirds of the way through the teaching unit where the teacher explicitly dealt with what the three different types were. She did this by eliciting ideas from the class and using these to clarify the different types of questions: summary, comparison and relationship. What came up in this discussion was confusion around what relationship-type questions were. This was clarified through the teacher questioning the students’ understanding. By the end of the discussion they had agreed that summary questions involved one variable and one group or population, comparison questions involved one variable compared across two groups, and relationship questions involved two variables taken from an individual.

The teacher persisted throughout the unit of work focusing on reinforcing the criteria for what makes a good question, getting the context sorted out by getting the students to correctly

define the variable(s) and the population, and making the questions about the population not the sample. In addition to this she required them to make predictions of what they expected, particularly in the comparison situation, asking students all the time which they thought would be bigger, taller or faster. The implication of this was around how the comparison question was framed – for example, did they have boys taller than girls or girls taller than boys? – with the expectation being that the question was framed so that it aligned with what the students expected to be true. So if they thought boys were taller than girls, then the question was framed that way.

A key realisation from this research was that students were conjecturing based on their general knowledge which group would tend to have bigger values. The students were not explicitly aware they were making such a conjecture, but their posed questions strongly suggest that they were. This was a new insight from the second and third teaching experiments that the teacher was now drawing on.

The structure of comparison questions was framed as the **variable** (of one of the groups) tending to be bigger/smaller than the **variable** (of the other group), thus putting the focus on what was being compared, i.e. the variable (see Pfannkuch, Regan, Wild, & Horton (2010)). Figure 6-2 gives three examples of teacher-modelled comparison investigative questions in class.

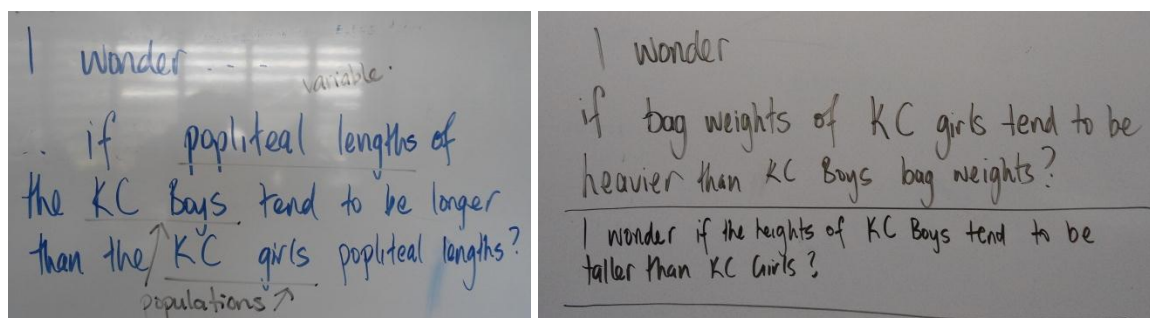


Figure 6-2. Examples of teacher-modelled comparison investigative questions

Student-posed investigative questions were marked and corrected individually, and overall feedback was given to the class on three occasions throughout the unit of work.

### 6.5. Retrospective analysis

Two of the three research questions are explored in this retrospective analysis section: (1) What makes a good investigative question? and (2) What are the underpinning concepts that



are needed to support the teaching and learning around posing investigative questions? The first question and the third question (What level of comparative investigative questions are year 10 (ages 14–15) students posing?) are reflected upon in section 6.6, quantitative analysis of test questions.

Three major findings came out of the retrospective analysis of student-posed investigative questions. The three findings are: (1) criteria for what makes a good investigative question; (2) a detailed two-way classification matrix for investigative questions that are posed; and (3) a teaching and learning framework that shows the connection between the different types of knowledge needed at the various stages of posing an investigative question (Arnold, 2009), from an inkling to a precise question (Wild & Pfannkuch, 1999). All of these are relevant to curriculum level 5 (ages 13–15) in the New Zealand curriculum (Ministry of Education, 2007). The first two findings are discussed in this chapter and are relevant to the first two research questions. The third finding, different types of knowledge needed, is not discussed in this thesis but is identified as an area for further research.

The main sources of data were student pre- and post-test responses. Classroom transcripts, student pre- and post-test interviews, and field notes have been used to provide additional data. Supervisor and in-depth research discussion seminars with other mathematics and statistics educators were used to float, discuss, firm up and discard ideas.

### 6.5.1. Criteria for what makes a good investigative question

The criteria for what makes a good investigative question combine the three features the teacher used in the second teaching experiment (see page 98), moderating questions from the first teaching experiment (see page 94) and detailed analysis of the questions that students posed in their pre- and post-tests from the four teaching experiments. The criteria and considerations for teaching and curriculum development are:

| Criteria                                                  | Considerations for teaching and curriculum development                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. The variable(s) of interest is/are clear and available | <ul style="list-style-type: none"><li>• The variable being described for a summary situation is clear.</li><li>• The variable being compared in a comparison situation is clear.</li><li>• The variables being looked at for a relationship are clear.</li><li>• The variable(s) has been correctly identified from the actual survey question that was asked; see the example given on pages 107–108 about students rating their ability in maths.</li><li>• The variable(s) is/are available. For example, students posed a question around favourite place to live when the variable given was the region that students live in. The variable <i>favourite place to live</i> was not available, only the variable <i>where students live</i> was.</li></ul> |

| Criteria                                                                                            | Considerations for teaching and curriculum development                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
|-----------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2. The population of interest is clear                                                              | <ul style="list-style-type: none"> <li>• Watch student use of the words <i>a</i> or <i>the</i> in front of the population or group they are using. For example, students write about “a boy”, “the boys” or simply “boys”. The difference between these three is that “a boy” indicates an individual, “the boys” suggests the boys in the sample, whereas “boys” most likely indicates that it is about the population. In addition students sometimes refer to “these boys”, also meaning the sample rather than the population (see pages 104–105).</li> <li>• The use of language is often the downfall of students who inadvertently refer to the sample rather than the population, but students can also pose their question directly about the sample rather than the population, missing the point about posing questions about the population. An example of this is: “I wonder if the boys sampled in the 2007 NZ CensusAtSchool have a longer arm span than the girls sampled?” (2008 student, post-test)</li> <li>• Students must identify both population groups. Students often identify one of the population groups and then revert to an abbreviated form for the second group, and as a result do not clearly identify the second population group; for example, referring to the first group as year 10 New Zealand boys and then girls for the second group.</li> <li>• Watch the use of age groups as a possible population group that is represented by the data when the data is from a single year level. The issue here is that it might be a biased sample as only the oldest, or the youngest, of a particular age group would be represented in a single year group. For example, in year 10 most students are 14 years of age but there are also 13 year olds and 15 year olds. The 13 year olds tend to be students in their second half of their 14th year and the 15 year olds tend to be students in the first half of their 16th year.</li> </ul> |
| 3. The intent is clear                                                                              | <ul style="list-style-type: none"> <li>• From the question it needs to be clear if it is a summary, comparison or relationship question. The intent of the question is critical as it indicates the relevant analysis that will be undertaken for the variable(s).</li> <li>• “Is the question clear and unambiguous?” (Wild &amp; Pfannkuch, 1999).</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| 4. The question can be answered with the data                                                       | <ul style="list-style-type: none"> <li>• The problematic situation first arose around this criterion (see chapter 5). The question is not able to be answered as the population selected for summary/relationship questions or the population groups for comparison questions result in the sample available being too small for any sensible analysis to be made. This can also be an issue when a sample is given and then students sample from this sample, resulting in too small a sample.</li> <li>• The question needs to be specific, so that it is answerable from data – questions that are too vague and general are harder to answer (Graham, 2006).</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| 5. The question is one that is worth investigating, that it is interesting, that there is a purpose | <ul style="list-style-type: none"> <li>• The information obtained by answering the question will be useful to someone, i.e. there is a purpose for the investigation.</li> <li>• The question should be personally interesting to you – not only will this bring greater motivation, but also common-sense knowledge about the context should help to ensure that the investigation proceeds along sensible lines (Ben-Zvi &amp; Amir, 2005; Burgess, 2007; Graham, 2006; Hancock et al., 1992; Ridgway et al., 2005).</li> <li>• If neither of these points is relevant, then one would need to ask: Why we are investigating this situation?</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| 6. The question allows for analysis to be made of the whole group                                   | <ul style="list-style-type: none"> <li>• The question is referring to the aggregate picture rather than an individual case (Bakker, 2004a). For example, questions that ask about who is the tallest or who has the longest arm span are not suitable.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |

Figure 6-3. Criteria for posing investigative questions

In the third and fourth teaching experiments the teacher worked with the students to get them to generate the criteria as they critiqued a mixture of good and bad questions. Through this

activity most of the criteria were realised with any additional criteria added to the list as they became an issue in a particular question or activity. The teacher used the criteria to reflect with the students as to whether a suggested investigative question would be suitable or not. In the post-test for the third teaching experiment and both the pre- and post-test for the fourth teaching experiment students were asked to critique investigative questions that had been posed by others. What came through in the analysis of student test responses, particularly at the end of the fourth teaching experiment, was that they were clearly focused on checking for the population with the vast majority signalling this as an issue in three of the four questions that were poorly posed. In the other question just over half the students used population as one of the reasons why the only well posed question was a good question. More detail and analysis of student responses follow in the results section (see pages 125–128).

### 6.5.2. Classification of summary investigative questions

The key findings from the first teaching experiment were two descriptive frameworks for the classification of the *investigative* questions that students pose. Figure 6-4 shows the descriptive framework for summary questions (this is the same as Figure 5-3).

| Summary question categories |    |                                                                                                                                                    |
|-----------------------------|----|----------------------------------------------------------------------------------------------------------------------------------------------------|
| Non-summary questions       | 1. | Nonsense or not a summary question.                                                                                                                |
|                             | 2. | A question that is partially related to the data, but not answerable by the given data.                                                            |
|                             | 3. | A question that asks about an individual case.                                                                                                     |
| Pre-summary questions       | 4. | A question that asks how many of a particular category.                                                                                            |
|                             | 5. | A question that asks how many within a specified range.                                                                                            |
|                             | 6. | A question that asks for the most popular or most common.                                                                                          |
| Summary questions           | 7. | A question that asks about the overall distribution of the data or what is typical.                                                                |
|                             | 8. | A question that asks about the overall distribution of the data or what is typical and reflects the population for which conclusions can be drawn. |

**Figure 6-4. Summary question categories**

This framework and the one for comparison questions were initially used to grade student-posed questions in the second and then third teaching experiments. The frameworks, however, did not capture the whole picture of the investigative questions that students were posing. Firstly, the different ways that the students referred to the population showed varying degrees of “correct” or acceptable populations for which conclusions could be drawn, and secondly the question categories were not complete.

Student responses from both the pre- and post-tests were then analysed in-depth from the second (2008) and third (2009) teaching experiments and, as a result of this analysis, a two-way classification matrix was developed for posing summary and comparison questions. The

two-way classification comprises: (1) rows that describe the summary or comparison question categories, and (2) columns for the population descriptors. After the fourth (2011) teaching experiment, student responses were graded using the two-way classification matrix. The two-way classification matrix was sufficient for summary situations, i.e. no additional population descriptors or question categories needed to be added. However, following the analysis of student-posed comparison questions (see page 118), additional question categories were added for comparison situations.

The two different aspects of the classification matrix are discussed separately. The population descriptors are common across all types of questions, whereas the question categories are different for summary and comparison. Summary and comparison questions categories are discussed first, and then the population descriptors.

### ***Updated summary question categories***

The final classification for summary questions (Figure 6-5) has the following changes from the original classification (Figure 6-4): not-related questions were added to the first category; sample size was added to not being answerable by given data category; how many of a particular category or within a specified range were combined as they were both “how many” questions; and an additional summary question category was added. The additional category concerns questions framed around asking about a summary statistic; for example, What is the median height of year 10 girls? or What is the typical height of year 10 girls? Finally, the population category has been removed as this is now considered across all categories of questions, with the exception of category A. The changes are signalled in italics.

| Summary question categories |    |                                                                                                                                                                     |
|-----------------------------|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Non-summary questions       | A. | Nonsense, <i>not related</i> or not a summary question.                                                                                                             |
|                             | B. | A question that is partially related to the data, but not answerable by the given data ( <i>either due to sample size issues or variable not in the data set</i> ). |
|                             | C. | A question that asks about an individual case.                                                                                                                      |
| Pre-summary questions       | D. | A question that asks how many of a particular category <i>or within a specified range</i> .                                                                         |
|                             | E. | A question that asks for the most popular or most common.                                                                                                           |
| Summary questions           | F. | A question that asks about <i>a summary statistic; for example, the mean</i> .                                                                                      |
|                             | G. | A question that asks about the overall distribution of the data or what <i>are</i> typical ( <i>more than a single value, more than the typical or the mean</i> ).  |

**Figure 6-5. Updated summary question categories**

## Chapter 6 – What makes a good Investigative Question

Examples for each of the different categories of summary questions are given in Figure 6-6. At this point the population descriptors are not considered (see pages 121–124, just the summary intent. Examples are drawn from across the four teaching experiments.

| Question category                                                                                                                                                    | Student question example                                                                                                                | Commentary                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A.</b> Nonsense, not related or not a summary question.                                                                                                           | I wonder are the second (girl) and third (boy) related? Both NZ Euro, 144 cm, ran, Yr 6. (2007 student, pre-test)                       | An example of a nonsense question.                                                                                                                                                                                                                  |
|                                                                                                                                                                      | I wonder what gender has the longest average right foot. (2007 student, pre-test)                                                       | This is a comparison question.                                                                                                                                                                                                                      |
| <b>B.</b> A question that is partially related to the data, but not answerable by the given data (either due to sample size issues or variable not in the data set). | I wonder what is the preferred region to live in. (2007 student, pre-test)                                                              | Preferred region to live in was not one of the variables in the data set.                                                                                                                                                                           |
| <b>C.</b> A question that asks about an individual case.                                                                                                             | I wonder what height the tallest 15 year old girl is. (2008 student, pre-test)                                                          | This question is an example of an <i>analysis</i> question, a question that you might ask when describing the data. It is asking about the tallest height – probably a single case; definitely asking about a single statistic (the maximum value). |
| <b>D.</b> A question that asks how many of a particular category or within a specified range.                                                                        | I wonder how many year 11 girls are taller than 160 cm. (2008 student, post-test)                                                       | This question is an example of an <i>analysis</i> question, a question that you might ask when describing the data. This one asks how many within a range.                                                                                          |
|                                                                                                                                                                      | I wonder how many students come from the Auckland region in the sample of students. (2009 student, post-test)                           | This question is an example of an <i>analysis</i> question, a question that you might ask when describing the data. This one asks how many of a particular category.                                                                                |
| <b>E.</b> A question that asks for the most popular or most common.                                                                                                  | I wonder what is the most popular ethnicity group. (2007 student, post-test)                                                            | This question is an example of an <i>analysis</i> question, a question that you might ask when describing the data. This one is an example of: What is the most common group by ethnicity?                                                          |
|                                                                                                                                                                      | I wonder what is the most common amount of languages spoken is from the 2007 C@S database (year 11 students). (2008 student, post-test) | This question is an example of an <i>analysis</i> question, a question that you might ask when describing the data. This one is an example of: What is the most common number of languages spoken?                                                  |

| Question category                                                                                                                                            | Student question example                                                                              | Commentary                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>F.</b> A question that asks about a summary statistics; for example, the mean.                                                                            | I wonder what is the typical pulse rate for NZ students. (2009 student, post-test)                    | This framing of the question is reasonably common in the second and third teaching experiments as initially the model of summary questions had this structure until it was realised that this is actually asking about “the” typical rather than “what are typical”.                                                                                                                                                                             |
|                                                                                                                                                              | I wonder what the average neck circumference is for secondary students. (2009 student, post-test)     | It was decided to move beyond the “average” questions because even though they have a sense of aggregate, they also could be answered by just calculating the mean or finding the median. Summary questions are about describing the whole distribution so posing this style of question could unintentionally restrict the type of analysis that students would do.                                                                             |
| <b>G.</b> A question that asks about the overall distribution of the data or what are typical (more than a single value, more than the typical or the mean). | I wonder how big is their armspan. (2008 student, pre-test)                                           | This question suggests an overall view rather than a single or individual view. “How big” has been interpreted to mean what is the distribution of arm spans.                                                                                                                                                                                                                                                                                    |
|                                                                                                                                                              | I wonder what the typical arm spans are of yr10 Auckland Region students. (2011 student, post-test)   | The language used moved towards thinking of typical heights or typical arm spans, for example. The rationale here was that to answer this question students needed to be thinking of more than just the middle or average. They should also be considering the range, the middle group, the mode (or peak), and the shape of the distribution, thus giving an overall picture of the data rather than a single point such as the median or mode. |
|                                                                                                                                                              | I wonder what are the typical wrist circumferences for yr 9–13 NZ students. (2011 student, post-test) | Very little difference between this example and the previous one, except the order of the words, the previous example talks about “the typical arm spans are” and this one “are the typical wrist circumferences”. In both instances the typical variable is referred to in the plural, not as a singular.                                                                                                                                       |

Figure 6-6. Summary question examples

***Reflection on final framing of summary investigative questions***

The framing of summary-type investigative questions became clearer over the four teaching experiments. The summary-type investigative question posed needed to be about the overall distribution of the data rather than focused on a single statistic, such as the average or the typical, and so this was the way that the examples that were used with students were developed. It was hard not to fall into the trap of developing a template for posing summary investigative questions (see last paragraph of section 6.4.2, page 106), i.e. a question template where students basically just change the variable and the population. The most common style of question that was used in the teaching and learning material in the fourth teaching experiment was category G (Figure 6-6): “What are typical [**variables**; for example, heights, popliteal lengths] for the [**given population**]?” But as can be seen from the examples for category G, which are typical examples of student responses, this wording was not used by

the students. The use of the word *the* is prevalent in students' phrasing; for example, the last example is "I wonder what are *the* typical wrist circumferences for year 9–13 NZ students" (2011 student, post-test), whereas a tidier version of this question would be "*I wonder what are typical wrist circumferences for year 9–13 NZ students.*"

The more common type of phrasing was category F questions: "What is the typical [**variable**; for example, heights, popliteal lengths] for the [**given population**]?" In the 2008 and 2009 student post-tests, generally the questions that had "typical" in them use this wording. In the 2011 post-tests this category F wording was still predominant, but of the 58 actual summary questions posed across the class, 18 of those were category G, compared with one out of 16 in 2009 and three out of 35 in 2008. This change seems to reflect the change in wording that the teacher was using in the class examples.

### ***Other examples of category G questions***

Questions such as 'I wonder how tall year 11s are?' (2008 student, post-test) are examples of summary investigative questions that would meet the criteria of asking about the overall distribution. To answer this question you would most likely talk about the shape of the distribution, give an indication of the range of heights, indicate where the centre is and the middle 50%, and if there is anything unusual or interesting. Other hypothesised examples include:

- What types of ice cream do year six students in our school like?
  - This is a proposed summary investigative question for category data. The description could include the all the different varieties of ice cream flavours, the most popular and least popular, and anything interesting or unusual.
- What is the distribution of heights for secondary school girls in New Zealand?
  - This is a proposed summary investigative question for continuous data. The desired description of the distribution includes the shape and at least two other features (see section 8.7.2).

### 6.5.3. Classification of comparison investigative questions

As indicated in the previous section, two frameworks – summary and comparison – were developed based on findings in the first teaching experiment. Figure 6-7 shows the descriptive framework for comparison questions (this is the same as Figure 5-4).

| Comparison question categories |     |                                                                                                                                                                      |
|--------------------------------|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Not answerable                 | 1.  | Nonsense or not a comparison question.                                                                                                                               |
|                                | 2.  | A question that is partially related to the data, but not answerable by the given data.                                                                              |
|                                | 3.  | A question that is related to the data but not answerable due to sample size issues.                                                                                 |
| Answerable                     | 4a. | A question that is answerable by the data.                                                                                                                           |
|                                | 4b. | A question that is answerable by the data and requires recategorisation of categories to be compared.                                                                |
|                                | 5a. | A question that is answerable by the data and reflects the population for which conclusions can be drawn.                                                            |
|                                | 5b. | A question that is answerable by the data and requires recategorisation of categories to be compared and reflects the population for which conclusions can be drawn. |

**Figure 6-7. Comparison question categories**

Comparison question categories were updated following the second and third teaching experiments where student responses generated new categories. The categories were updated further following the fourth teaching experiment, sparked by student responses that did not align with the current categories. In addition to these changes, two additional categories were added following consideration of higher-level analysis and the impact this has on the investigative question posed.

#### ***Updated comparison question categories***

The updated classification for comparison questions (Figure 6-8, next page) has the following changes from the original classification (Figure 6-7): not-related questions were added to the first category; the criteria regarding not being able to answer the question with the given data were combined to recognise the two situations for which this occurs (i.e. either the question cannot be answered because of sample size issues or the variable is not in the data set); the answerable categories from the original classification were updated to pre-comparison and comparison questions (levels 4–6, ages 11–16) and the criteria completely reworked because student responses in 2008 and 2009 generated a number of criteria that were not evident in the 2007 student responses; questions about how much bigger or if there is a difference were added to category G following the fourth teaching experiment. These types of questions had not been previously posed by students.



In addition, categories for curriculum level 7 (ages 16–17) and curriculum level 8 (ages 17–18) have been added to the classification. Neither of these are present in any of the student work presented but have arisen as part of the wider research on statistics in New Zealand which the researcher is involved in and are pertinent to this research in terms of horizon thinking or horizon knowledge for teachers (Ball, Thames, & Phelps, 2008). The updated classification for comparison investigative questions is given in Figure 6-8. The changes are signalled in italics.

| Comparison question category             |    |                                                                                                                                                                                                                                                                          |
|------------------------------------------|----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Non-comparison questions</i>          | A. | Nonsense, <i>not related</i> or not a comparison question.                                                                                                                                                                                                               |
|                                          | B. | A question that is partially related to the data, but not answerable by the given data ( <i>either due to sample size issues or variable not in the data set</i> ).                                                                                                      |
| <i>Pre-comparison questions</i>          | C. | <i>A question that hints at comparison.</i>                                                                                                                                                                                                                              |
|                                          | D. | <i>A question that has all of one group bigger/smaller than all of another group or compares individuals.</i>                                                                                                                                                            |
| <i>Comparison questions (levels 4–6)</i> | E. | <i>A question that compares categorical data.</i>                                                                                                                                                                                                                        |
|                                          | F. | <i>A question that compares a summary statistic.</i>                                                                                                                                                                                                                     |
|                                          | G. | <i>A question that assumes the idea of tendency. This includes questions that ask how much bigger or if there is a difference. Uses a comparing word.</i>                                                                                                                |
|                                          | H. | <i>A question that includes the idea of tendency; for example, on average, generally, tends.</i>                                                                                                                                                                         |
| <i>Comparison question (level 7)</i>     | I. | <i>A question that reflects the level of analysis that students working at curriculum level 7 have, i.e. the question reflects that analysis includes finding an informal confidence interval for the median.</i>                                                        |
| <i>Comparison question (level 8)</i>     | J. | <i>A question that reflects the level of analysis that students working at curriculum level 8 have, i.e. the question reflects that analysis can now determine if there is a difference between medians or mean and quantify this in terms of a confidence interval.</i> |

**Figure 6-8. Updated comparison question categories**

Figure 6-9 (next page) gives examples for each of the different categories of comparison questions, and as with the summary questions in the previous section, at this stage the population descriptors are not considered.

## Chapter 6 – What makes a good Investigative Question

| Question category                                                                                                                                                    | Student question example                                                                                                                                                                       | Commentary                                                                                                                                                                                                                                                       |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A.</b> Nonsense, not related or not a comparison question.                                                                                                        | I wonder Auckland region and Wellington region have the same student in year 10? (2009 student, post-test)                                                                                     | This question is irrelevant.                                                                                                                                                                                                                                     |
|                                                                                                                                                                      | I wonder if the popliteal length relates to armspan. (2009 student, post-test)                                                                                                                 | This is a relationship question.                                                                                                                                                                                                                                 |
| <b>B.</b> A question that is partially related to the data, but not answerable by the given data (either due to sample size issues or variable not in the data set). | I wonder if all the ambidextrous students are capable of kicking a ball with both left and right foot. (2009 student, pre-test)                                                                | Handedness was in the data set provided; however, there wasn't a question about ambidextrousness for "footedness".                                                                                                                                               |
|                                                                                                                                                                      | If Asian girls have a longer armspan than Indian boys. (2007 student, post-test)                                                                                                               | In the 2007 post-test there were only two Indian boys and two Asian girls.                                                                                                                                                                                       |
| <b>C.</b> A question that hints at comparison.                                                                                                                       | I wonder if more year 10 boys are physically fit than year 10 girls. (2011 student, post-test)                                                                                                 | This question suggests comparison, though as it reads it is probably only comparing a couple of categories.                                                                                                                                                      |
|                                                                                                                                                                      | I wonder if ambidextrous hand writers can speak different languages. (2009 student, pre-test)                                                                                                  | This question hints at comparing the number of languages spoken across handedness.                                                                                                                                                                               |
| <b>D.</b> A question that has all of one group bigger/smaller than all of another group or compares an individual.                                                   | I wonder if all girls have longer hair than all boys. (2008 student, pre-test)                                                                                                                 | A good example of the type of thinking, and therefore the type of question, where students are thinking something is bigger and think all of one is bigger than all of the other. They have not yet grasped the idea of tendency or tending to be bigger/longer. |
|                                                                                                                                                                      | I wonder if the average resting rate for a boy is lower than a girls? (2011 student, pre-test)                                                                                                 | Comparing a boy with a girl, comparing individuals.                                                                                                                                                                                                              |
| <b>E.</b> A question that compares categorical data.                                                                                                                 | I wonder if secondary students that live in southland region are fitter than secondary students from Auckland region. (2009 student, post-test)                                                | In the data set given, the variables that might be used to answer this question were both categorical, region they live in, and fitness levels (unfit, a little fit, ...).                                                                                       |
| <b>F.</b> A question that compares a summary statistic.                                                                                                              | I wonder if the typical right foot length for year 11 boys is greater than the typical right foot length for year 11 girls from the 2007 NZ CensusAtSchool database. (2008 student, post-test) | This question is comparing "the typical" which is interpreted as a summary statistic; for example, the median or the mode.                                                                                                                                       |
|                                                                                                                                                                      | I wonder if the average hair length of 16 year old girls is greater than the average hair length of 16 year old boys. (2008 student, pre-test)                                                 | This question is comparing the average, which could be median, mean or mode.                                                                                                                                                                                     |
| <b>G.</b> A question that assumes the idea of tendency. This includes questions that ask how much bigger or if there is a difference.                                | I wonder if secondary girl students have bigger wrist circumference than secondary boy students. (2009 student, post-test)                                                                     | This question uses the phrase "have bigger" but, unlike the example in category D, they haven't indicated that they are thinking all girls bigger than all boys, so this style of question has been categorised as assuming tendency.                            |
|                                                                                                                                                                      | I wonder if boys have longer popliteal lengths than girls. (2009 student, post-test)                                                                                                           | A second example showing a different variable; commentary above relevant for this question.                                                                                                                                                                      |

| Question category                                                                                                                                                                                                                                                           | Student question example                                                                                                                             | Commentary                                                                                                                                                                                                                                                |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>H.</b> A question that includes the idea of tendency; for example, question includes words or phrases such as on average, generally or tends.                                                                                                                            | I wonder if boys in year 10 tend to be <b>taller</b> than girls in year 10. (2009 student, post-test)                                                | This question structure has one population tending to be taller/heavier OR have a longer/shorter [ <b>variable</b> ] than the other population.                                                                                                           |
|                                                                                                                                                                                                                                                                             | I wonder if on average right handers have longer <b>hair</b> than left handers. (2008 student, pre-test)                                             | This is a similar structure to the first, but instead of using “tend”, they have used “on average”.                                                                                                                                                       |
|                                                                                                                                                                                                                                                                             | I wonder if Yr 9–13 NZ boys have typically higher <b>pulse rates</b> compared to Yr 9–13 NZ girls. (2011 student, post-test)                         | This is a similar structure to the first also, but this time they have used “typically” to express the idea of tendency.                                                                                                                                  |
|                                                                                                                                                                                                                                                                             | I wonder if the <b>popliteal length</b> of Yr 9–13 NZ girls tend to be longer than Yr 9–13 NZ boys <b>popliteal length</b> (2011 student, post-test) | This question structure has the <b>variable</b> (of one of the groups) tending to be bigger/smaller than the <b>variable</b> (of the other group). A different structure to the previous three.                                                           |
| <b>I.</b> A question that reflects the level of analysis that students working at curriculum level 7 have, that is, the question reflects that analysis includes finding an informal confidence interval for the median.                                                    | Proposed question:<br><br>I wonder if the median height of New Zealand year 11 boys is greater than the median height of New Zealand year 11 girls?  | At level 7 students can now find an estimate for the population median for each group, then find an informal confidence interval for the population median and use this to make a call. The question needs to reflect these new tools that students have. |
| <b>J.</b> A question that reflects the level of analysis that students working at curriculum level 8 have, i.e. the question reflects that analysis can now determine if there is a difference between medians or mean and quantify this in terms of a confidence interval. | Proposed question:<br><br>I wonder what the difference in <u>median</u> heights is between New Zealand year 11 boys and New Zealand year 11 girls?   | At level 8 students have more tools in their toolkit and can find a confidence interval for the difference in medians (or means). The question needs to reflect this further sophistication in analysis.                                                  |

Figure 6-9. Comparison question examples

***Reflection on final framing of comparison investigative questions***

Two reflections on the final framing of the comparison investigative questions need to be mentioned. Firstly, the use of “tend to” to describe the idea of comparison, where one group “tends to be higher” than the other for a given variable, was signalled right from the start of the work on posing investigative questions (Pfannkuch, Wild, Horton, & Regan, 2009). Researching students’ thinking about comparison situations (Pfannkuch, 2006; Pfannkuch & Horring, 2005) had already identified “tend to” as being an important consideration in teaching thinking about the question framing for comparison situations.

Secondly, from the second teaching experiment to the third teaching experiment the framing of the question used in the pre- and post-tests moved from “I wonder if Year 11 NZ boys tend to have shorter **hair** than Year 11 NZ girls?” to “Do the **hours of sleep per night** for NZ

year 5 to year 10 students who own cell phones tend to be less than the **hours of sleep per night** for NZ year 5 to year 10 students who do not own cell phones?” Both are acceptable as suitable comparison questions at this curriculum level, but the second question puts the variable (of the populations) clearly as the item that is being compared. Specifically, this second type focuses on comparing the distribution of the variable as opposed to the comparing the population groups and this is the preferable focus for comparison questions. In the 2008 post-test, there were 65 comparison questions that were one of these two types: 59 of the questions were the first type and six (9%) were the second type. In the 2009 post-test, of 29 questions, 24 were the first type and five (17%) were the second type, and in the 2011 post-test, of 67 questions, 36 were the first type and 31 (46%) were the second type.

### 6.5.4. Population descriptors

In the initial classification for summary and comparison questions, both had the top category as being a “good” question plus the population. As student pre- and post-test responses were analysed from the second (2008) and third (2009) teaching experiments, it became clear almost immediately that the “super” category of population was not going to work. Students who had similar types of questions had a wide range of populations. For example, in the 2008 post-test 22 of the 24 students posed an investigative question about one group being taller than another group. Aside from the variation in the question format, 14 different population or group descriptors were used. The descriptors fell into three main categories: (1) boys and girls; (2) various combinations of age groups; and (3) year 11 boys and girls. Table 6-2 (next page) lists the descriptions used and the number of students who used the description. The sample came from the 2007 New Zealand CensusAtSchool database and only year 11 students had been selected.

**Table 6-2. Population descriptors in 2008 post-test responses for questions about “being taller”**

| Sorted by                            | Group or population descriptors                                            | Number |
|--------------------------------------|----------------------------------------------------------------------------|--------|
| Boys and girls                       | Sample from boys and girls, 2007 New Zealand CensusAtSchool                | 1      |
|                                      | boys and girls                                                             | 1      |
|                                      | boys and girls, 2007 CensusAtSchool                                        | 1      |
|                                      | boys and girls, 2007 New Zealand CensusAtSchool                            | 2      |
| Ages with and without boys and girls | Sample from 15 year olds and 16 year olds, 2007 New Zealand CensusAtSchool | 1      |
|                                      | 15 year olds and 16 year olds                                              | 1      |
|                                      | 16 year old girls and 15 year old boys                                     | 2      |
|                                      | 15 year olds and 16 year olds, 2007 CensusAtSchool                         | 1      |
|                                      | 15 year old girls and 15 year old boys, 2007 New Zealand CensusAtSchool    | 2      |
| Year 11 boys and girls               | year 11 girls and year 11 boys                                             | 1      |
|                                      | year 11 girls and year 11 boys, 2007 CensusAtSchool                        | 2      |
|                                      | NZ year 11 boys NZ year 11 girls                                           | 1      |
|                                      | year 11 girls and year 11 boys, New Zealand CensusAtSchool                 | 2      |
|                                      | year 11 girls and year 11 boys, 2007 New Zealand CensusAtSchool            | 4      |

Within the three broader categories there are multiple ways that students can phrase a descriptor and this is based around whether they acknowledge that the broader population is New Zealand students, and that the sample has been taken from a particular CensusAtSchool database. It could be possible to make a fine graded scale for population descriptors, but pragmatism and what would be useful to teachers and students meant that fewer categories were better than more.

Initially there seemed to be three clear categories: (1) boys and girls – this is very general and could mean all boys and all girls in the world; (2) New Zealand boys and girls – this is better; but generally the data came from a subgroup of New Zealand boys and girls giving (3) New Zealand year level(s) boys and girls as the target. Summarising the three categories are:

1. Broad student population; for example, boys, girls, students.
2. Broad New Zealand student population; for example, New Zealand boys, New Zealand students.
3. Actual New Zealand student population; for example, New Zealand year 10 students, New Zealand year 11 students, New Zealand secondary school girls.

However, as can be reasonably expected, student responses did not fall nicely into the three categories. Where, for example, did year 11 boys and year 11 girls fit? Clearly it is more specific than New Zealand students, but it doesn't specify New Zealand. An additional category was needed between broad New Zealand student population and actual New Zealand student population. This category covers any age or year group that could reasonably be assumed from the year group that was selected for the sample. For example, as well as accepting year 11 students (2008 pre- and post-tests and 2009 pre-test), 15 year olds and 16 year olds were also accepted; and likewise, as well as accepting secondary students (2009 post-test and 2011 pre- and post-tests), any combination of years 9–13, ages 13–18 and teenagers were accepted as well.

Two other types of questions occurred that did not fit within these four categories. In the first type, which occurred only rarely, students went broader than boys and girls but didn't use a specific population descriptor; for example, they asked about typical heights of males and females or of people. The second type of question that didn't fit into the four categories was when students specifically or inadvertently posed their investigative question about the sample. Examples of the second type of question are: "What are typical heights of these year 11 students?" and "What are typical heights for year 11 students sampled from the 2007 NZ C@S database?"

The year and the CensusAtSchool database add a further dimension, as examples in Table 6-2 show, but in the end these were not considered a requirement. It may have been better in hindsight to have just said the sample was of New Zealand secondary students, or New Zealand year 11 students. If a student wrote 2007 CensusAtSchool, this was considered to be equivalent to having written New Zealand.

This gave six population categories that are considered as part of the overall question classification. These categories held through analysing the questions posed in the fourth teaching experiment, are relevant for both summary and comparison questions, and would also be relevant for relationship questions. The final six population categories are in Figure 6-10 (next page).

- |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"><li>1. Referring to the sample.</li><li>2. Broad population, not specifying students.</li><li>3. Broad student population; for example, boys, girls, students.</li><li>4. Broad New Zealand student population; for example, New Zealand boys, New Zealand students.</li><li>5. Any relevant student population that can be generalised about from the actual New Zealand student population used; for example, year 11 students, teenagers, secondary school girls.</li><li>6. Actual New Zealand student population; for example, New Zealand year 10 students, New Zealand year 11 students, New Zealand secondary school girls.</li></ol> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Figure 6-10. Population descriptor categories**

In the teaching and learning, the focus was always on using the category 6 population descriptor, but in assessment situations category 5 would be an acceptable population descriptor. Category 4, however, would not be acceptable because there needs to be some recognition of the limiting factor of the specific age or year level group selected in the sample. The fact that they are New Zealand students is not enough by itself.

### **6.5.5. Two-way classification framework (matrix)**

In order to classify a posed investigative question, both for summary and for comparison situations, the two categories (1) question category and (2) population descriptor category need to be considered as the student is working with both aspects (categories) at the same time. The combination of the two aspects gives rise to a two-way classification framework for both summary and comparison investigative questions. For summary investigative questions the framework would be a 7 by 6 matrix made up of the seven question categories (rows, Figure 6-5) and six population descriptor categories (columns, Figure 6-10); for comparison investigative questions (levels 4–6, ages 11–15) the framework would be an 8 by 6 matrix made up of the eight question categories (rows, Figure 6-8) and six population descriptor categories (columns, Figure 6-10). The matrices for summary and comparison investigative questions are shown in Figure 6-11 (next page).

|    | 1. | 2. | 3. | 4. | 5. | 6. |
|----|----|----|----|----|----|----|
| A. |    |    |    |    |    |    |
| B. |    | B2 |    |    |    |    |
| C. |    |    |    |    |    | C6 |
| D. |    |    |    |    |    |    |
| E. |    |    |    |    | E5 |    |
| F. |    |    | F3 |    |    |    |
| G. |    |    |    |    |    | G6 |

|    | 1. | 2. | 3. | 4. | 5. | 6. |
|----|----|----|----|----|----|----|
| A. |    |    |    |    |    |    |
| B. |    |    |    |    |    |    |
| C. |    | C2 |    |    |    |    |
| D. |    |    |    |    | D5 |    |
| E. |    |    |    |    |    |    |
| F. |    |    |    |    |    |    |
| G. |    |    |    | G4 |    |    |
| H. |    |    |    |    |    | H6 |

Figure 6-11. (a) Summary investigative question matrix      (b) Comparison investigative question matrix

The shaded portion of the two matrices in Figure 6-11 show where the two aspects combine to give different combinations to describe the investigative questions posed; for example, B2, and F3 are different combinations of the two aspects. To give more specific examples, **F3** (in Figure 6-11a) is a summary investigative question that asks about a summary statistic and has a broad student population, and **H6** (in Figure 6-11b) is a comparative investigative question that includes the idea of tendency and has the actual New Zealand student population.

## 6.6. Quantitative analysis of test questions

Two specific aspects of the pre-and post-tests are discussed in this results section: (1) post-test responses about whether a given investigative question is a good investigative question or not for the 2011 students in relation to the first research question “What makes a good investigative question?”; and (2) pre- and post-test responses on posing investigative comparison questions for 2008 and 2011 students in relation to the third research question “What level of comparative investigative questions are year 10 (ages 14–15) students posing?”

### 6.6.1. Students’ ability to discern good (or not) investigative questions

This section looks at student responses to a request to critique previously posed investigative questions as a further look at the research question “What makes a good investigative question?” In the 2009 post-test and the 2011 pre- and post-tests students were asked to comment on five previously posed investigative questions and to give reasons whether they thought the investigative question was a good investigative question or not. If the investigative question was not a good investigative question, students were asked to change it to make it a better investigative question.



Three questions from the 2011 post-test responses (questions 1, 3, and 4) have been selected to exemplify students critiquing of questions posed by others. The three questions chosen include two “bad” questions and one “good” question. The bad questions chosen reflect the responses for the four bad questions and give sufficient flavour as to where students’ focus lay.

### ***Question 1: Do girls have longer arm spans than boys?***

There were 27 responses to this question in the post-test in 2011. Of these 27 responses, 26 students said that it was a bad question and one student said that it was a good question. The suggested improvements from 25 of these 26 students mentioned adding in the population and the populations suggested are summarised in Figure 6-12. Eighty per cent of students suggested the question could be improved by including New Zealand and the appropriate age/year level group, i.e. improve to the category-6 population descriptor.

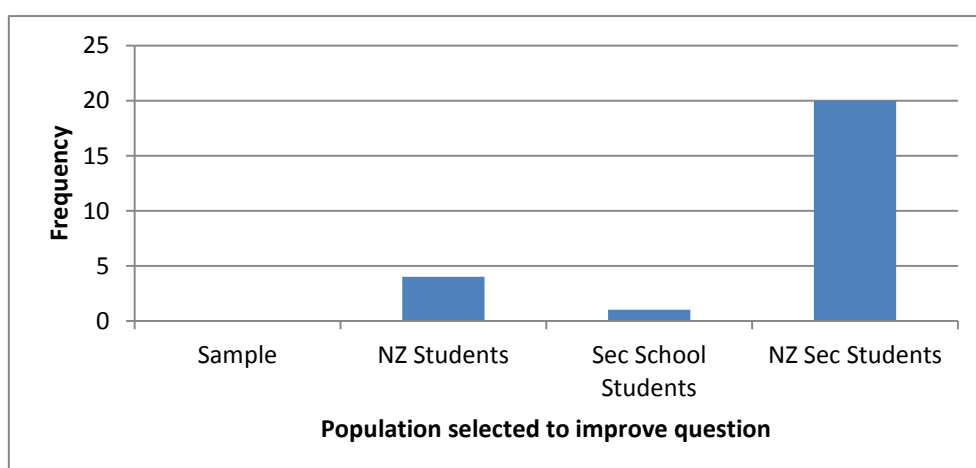


**Figure 6-12. Populations suggested to improve question 1**

The only other suggested improvement, made by one student in the 2011 post-test, was that the words “on average” should be added into the question: “Do girls on average have longer arm spans than boys?”

### ***Question 3: What are typical neck circumferences for these students?***

There were 28 responses to this question in the post-test in 2011. Of these 28 responses, 25 said that the question was not a good question and could be improved, while three of the students felt it was a good question. For those that said it could be improved 23 specified

some population improvement. Again a high percentage (70%) of these students suggested including the category-6 population descriptor as the improvement (Figure 6-13).

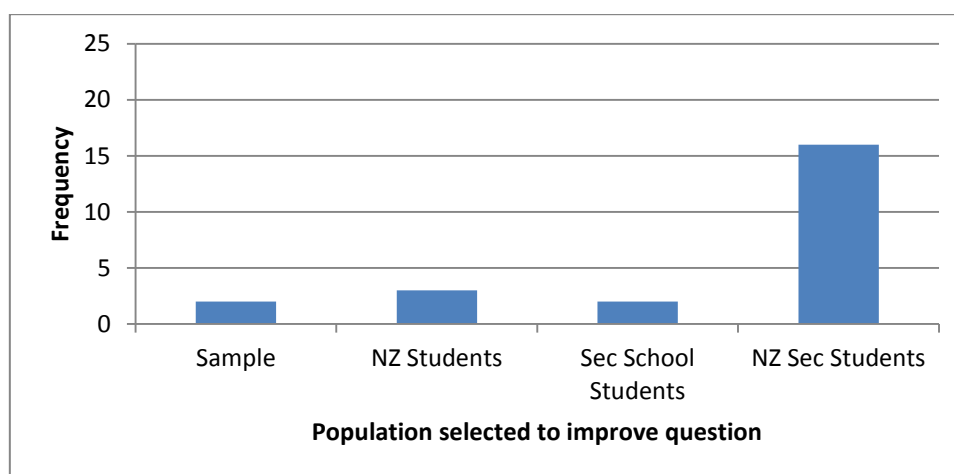


**Figure 6-13. Populations suggested to improve question 3**

From these two examples of bad questions it seems that students have generally narrowed in on the population as being a key feature of a good question. (In the two questions not analysed in detail in this section, question 2 and question 5, the students also mentioned population as a key feature of a good question.)

***Question 4: Do the popliteal lengths of NZ secondary school boys tend to be longer than the popliteal lengths of NZ secondary school girls?***

There were 27 responses to this in the post-test 2011. Of these 27 students, 25 said that it was a good question and two students said it was a bad question. This is summarised in Figure 6-14 (next page). Of the 25 students who said it was a good question, 17 had the population as one of the reasons, and of these students, 13 also included the variable being clear as part of why it was a good question.

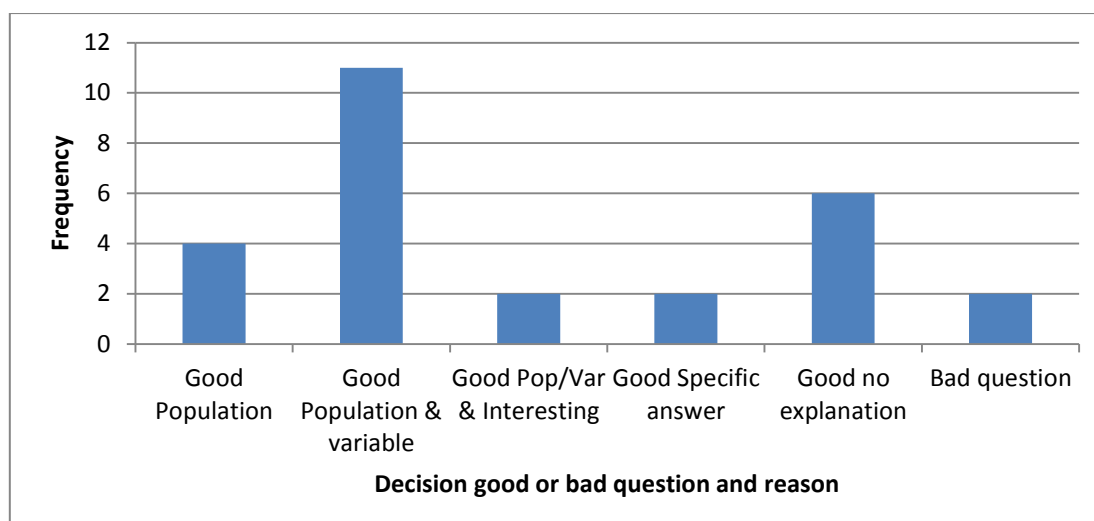


**Figure 6-14. Decision regarding question 4 and supporting reasons**

From looking at both the reasons given by the students for why a question is good and bad, it appears that the students have captured the idea that the investigative question needs to contain both the variable and the population and that generally they have an idea of what the population should be. This aligns with what they did when they posed their own questions (see final population categories on page 124), where approximately 60% of the comparison questions posed had category-6 population descriptors.

### **6.6.2. Comparison investigative questions**

This section gives the results for the third research question: What level of comparative investigative questions are year 10 (ages 14–15) students posing? In all four teaching experiments the students were asked to pose investigative questions in their pre- and post-tests. Both summary and comparison questions were analysed and informed the retrospective analysis. In this results section, only the comparison investigative questions are discussed for the 2008 and 2011 students.

The rationale for presenting only the comparison questions is that they reflect the direction that students are heading within the senior secondary school statistics curriculum so it is appropriate to focus on these types of questions. The rationale for comparing the 2008 and 2011 experiments is that the students were very similar in nature. Both classes were from mid-decile multicultural schools and both classes comprised students considered to be of above-average ability. The 2008 group were slightly higher overall in ability as these students were working one year ahead of their peers and were the top of two classes in this situation.

The 2011 students were in a higher streamed class, of which there were three, but the three classes were similar in nature as the streaming was by option line. This difference in ability between the 2008 and 2011 students is shown in the pre-test results for comparison questions (see Figure 6-15). The 2008 pre-test mean score (see next section for scoring) for comparison questions is higher up the scale than the pre-test mean score for comparison questions for the 2011 students. This suggests that the 2008 students tended to have a higher level of understanding of posing investigative questions coming into their teaching experiment than the 2011 students had coming into theirs. Details regarding these scores are covered in the next section.

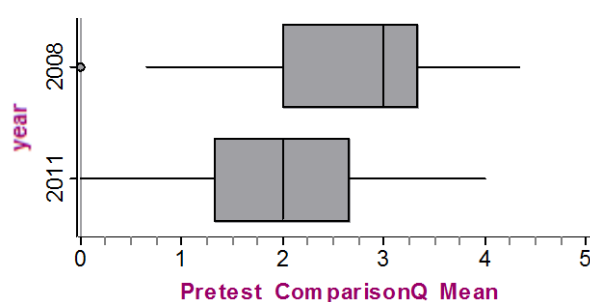


Figure 6-15. Pre-test mean score for comparison questions

This results section on comparison questions will look at the overall movement from pre-test to post-test for the two groups and also look at the overall movement from 2008 to 2011. In addition, the quality of the questions and the quality of the population descriptors are discussed.

### Overall question ranking

In the pre- and post-tests the students were asked to pose three comparison investigative questions. These questions were each individually graded according to the comparison question category (see Figure 6-8, page 118) and the population descriptor category (see Figure 6-10, page 124). For example, “*I wonder if the popliteal length of Yr 9–13 NZ girls tend to be longer than Yr 9–13 NZ boys popliteal length*” (2011 student, post-test) was graded as **H6** because as a comparison question it includes the idea of tendency and it also has the actual New Zealand student population correct. On the other hand, “*I wonder if boys have longer popliteal lengths than girls*” (2009 student, post-test) was graded as **G3** because it assumes the idea of tendency and has only specified a broad student population (boys and girls). This grading system gave more than 40 different possibilities when the question categories and the population descriptor categories were combined (see Figure 6-11b, page

125). In order to make comparisons between the two years and look at the difference within a year, the 40+ possibilities were simplified into six overall grades using the SOLO taxonomy (Figure 6-16).

| SOLO taxonomy level          | Grade | Description of evidence                                                                                                                                                                                          |
|------------------------------|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| No response or idiosyncratic | 0     | Questions that are not comparison questions, nonsense or not-related questions.<br><b>Category A questions.</b>                                                                                                  |
| Pre-structural               | 1     | Questions that are partially related to the data, but not answerable by the given data.<br><b>Category B questions, any population.</b>                                                                          |
| Uni-structural               | 2     | Questions that hint at comparison or have all of one group bigger/smaller than the other.<br><b>Category C and D questions, any population.</b>                                                                  |
| Multi-structural             | 3     | Questions that compare categorical data.<br><b>Category E questions, any population.</b><br><b>Relational or extended abstract categories (F, G and H) with population categories 1–4.</b>                       |
| Relational                   | 4     | Questions that compare summary statistics or assume the idea of tendency, including the idea of difference. Population is “acceptable”.<br><b>Category F and G questions with population categories 5 and 6.</b> |
| Extended abstract            | 5     | Questions that include the idea of tendency. Population is “acceptable”.<br><b>Category H questions with population categories 5 and 6.</b>                                                                      |

Figure 6-16. SOLO criteria for grading comparison investigative questions

A final pre-test and a final post-test score were found for each student by finding the mean of their three SOLO grades. These final scores have been analysed to look at the difference between pre- and post-test within the each year and to compare the results of the 2008 and 2011 classes.

**2008 results:** Twenty-four students completed both the pre- and post-test. Their final scores for comparison investigative questions in both of these tests are summarised in the scatterplot and summary table in Figure 6-17 (next page).

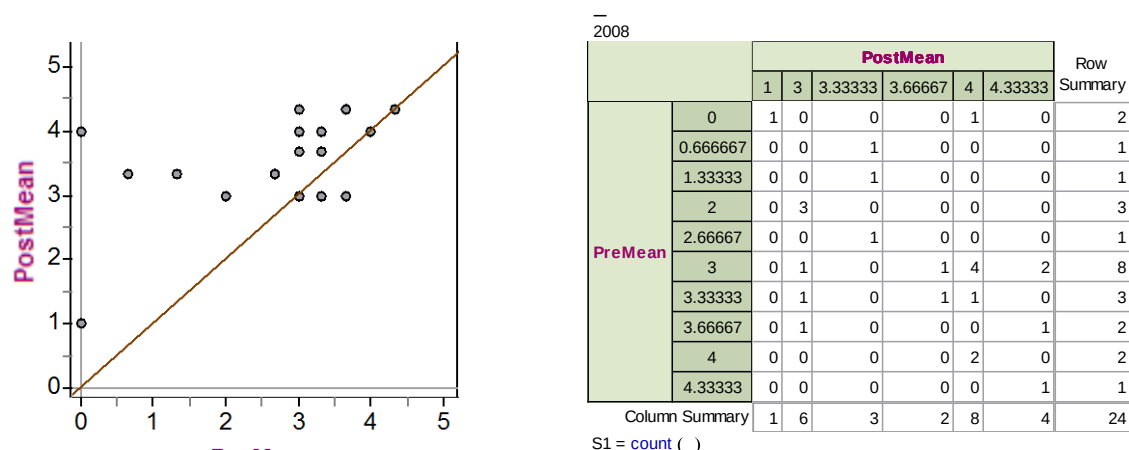


Figure 6-17. (a) Scatterplot and (b) table of 2008 student pre- and post-test mean scores for comparison investigative questions

Figure 6-17 shows a scatterplot of students' pre-test mean score and their post-test mean score. The line marked is post-test mean score = pre-test mean score.

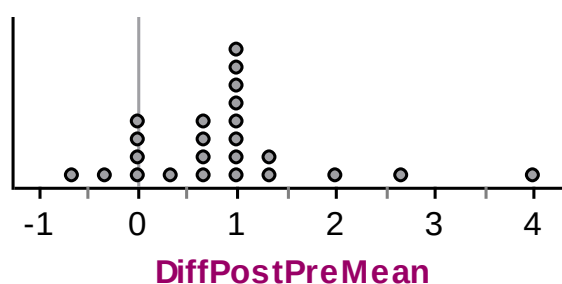


Figure 6-18. Graph of difference between post-test mean score and pre-test mean score

Figure 6-18 shows the difference between students pre-test mean score and their post-test mean score. A difference of one indicates that the student had a mean improvement of one point over their three comparison investigative questions. Figure 6-17 and Figure 6-18 show the same information, but in different ways. Of the 24 students who sat both the pre- and post-tests, 18 improved their mean score, four remained the same, and two lowered their mean score. From the post-test results, 12 students were working overall at a relational level, 11 students at a multi-structural level, and one at a pre-structural level. The student still working at a pre-structural level posed investigative questions that wondered if there was more of one type than another type; for example, more 16-year-old boys than 16-year-old girls. This student had English as a second language and language understanding may have been a confounding factor here. The students made significant improvement ( $p\text{-value} = 0.00019$ , paired  $t\text{-test}$ ) in their mean scores from pre- to post-test question posing, and on average increased their mean grade by 0.89 points (95% CI = [0.47, 1.30]).

**2011 results:** Twenty-six students completed both the pre- and post-test. Their final scores for comparison investigative questions in both of these tests are summarised in Figure 6-19.

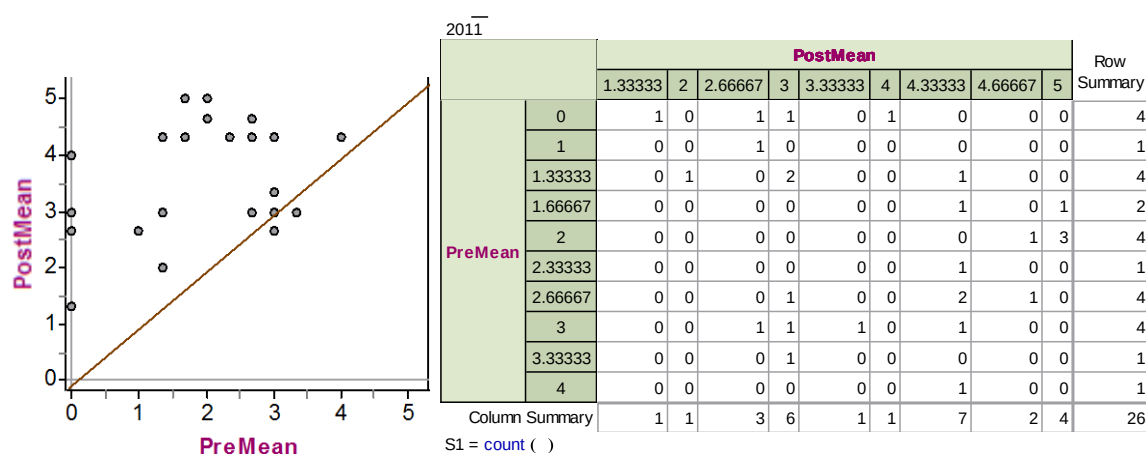


**Figure 6-19. (a) Scatterplot and (b) table of 2011 student pre- and post-test mean scores for comparison investigative questions**

Figure 6-19 shows a scatterplot of students' pre-test mean score and their post-test mean score. The line marked is post-test mean score = pre-test mean score.

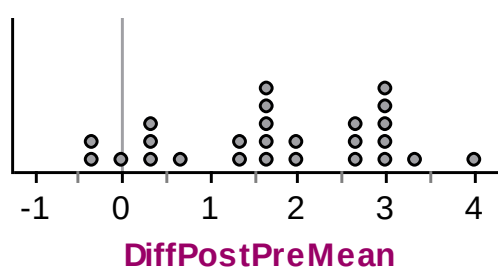


**Figure 6-20. Graph of difference between post-test mean score and pre-test mean score**

Figure 6-20 shows the difference between students pre-test mean score and their post-test mean score. A difference of two indicates that the student had a mean improvement of two points over their three comparison questions. Figure 6-19 and Figure 6-20 show the same information, but in different ways. Of the 26 students that sat both the pre- and post-tests, 23 improved their mean score, one remained the same, and two lowered their mean score. In the post-test, four students were working overall at extended abstract level, 10 at a relational level, seven at a multi-structural level; one at a pre-structural level, and four at a uni-structural level. The four uni-structural students all had a least one good question amongst their three, but were let down by a combination of the population category being low or one of the questions not being a comparison question. The pre-structural student asked questions that were about individuals (a boy, a girl) and also one non-comparison question. The

students made significant improvement ( $p\text{-value} < 0.0001$ , paired  $t\text{-test}$ ) in their mean scores from pre- to post-test question posing and on average increased their mean grade by 1.78 points (95% CI = [1.29, 2.28]).

**Comparison of 2008 and 2011 results:** There was no significant difference between post-test scores for the 2008 and 2011 students. The 2008 students overall had an average mean-grade increase of 0.89 whereas the 2011 students had an average mean-grade increase of 1.78. The 2011 students, however, started a lot lower down (see Figure 6-15, page 129). While it had been hoped to show there would be a difference between teaching experiment 2 (2008) and teaching experiment 4 (2011), this difference was not significant in their “average” grades in the post-test. A look at the question and population descriptor categories, however, in the next two sections shows a deeper understanding of the question aspects in 2011 than in 2008.

### ***Final question type***

Analysis of the different types of questions the 2011 students posed in their post-tests compared with the 2008 students shows the movement in the understanding of the researcher and teachers as to what makes a good investigative question. In 2008, the type of comparison investigative question that was modelled during the teaching experiment was category G (a question that assumes the idea of tendency), and this is reflected clearly in the student-posed questions. As shown in Figure 6-21, of the 72 questions that were posed by the students (across the three comparison investigative questions that they were asked to pose), 58 of the questions were category G-type questions.

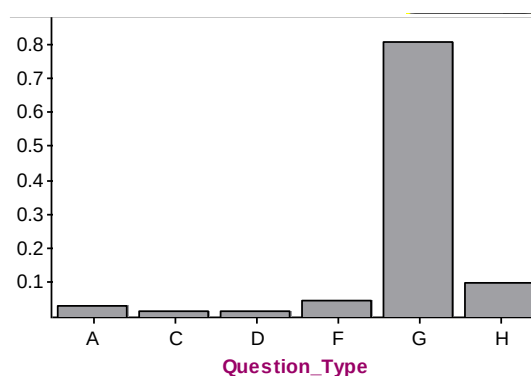


**Figure 6-21. 2008 question type for all post-test comparison investigative questions**

By 2011 the comparison questions that were modelled in class had moved to category H-type questions (see Figure 6-8, page 118). Figure 6-22 (next page) shows a higher proportion of



questions in category H (45 questions out of 87) than any of the other categories for 2011, and it also shows a higher proportion in this category compared with 2008 results (9.7% for 2008, 51.7% for 2011). An interesting observation is the number of category E-type questions (12) in 2011, which reflects students posing comparison questions about category data rather than numeric data. This contrasts with the 2008 results when there were no category E-type questions. However, there were more categories that could be compared in the 2011 post-test data set than in the 2008 post-test data set.

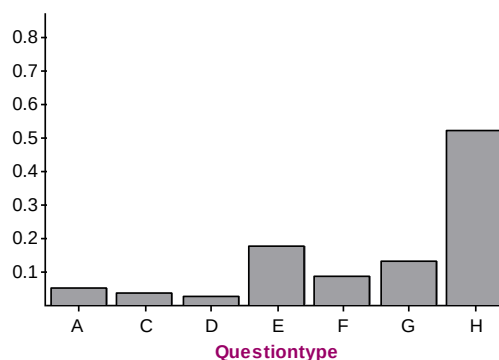


**Figure 6-22. 2011 question type for all post-test comparison investigative questions**

There appears to have been movement in the types of questions posed by students from 2008 to 2011 and this aligns with the changes in the thinking by the researcher.

### ***Final population categories***

The movement regarding population categories shows that in 2011 more students (57%) had category 6 population descriptors than in 2008 (43%) (Figure 6-23).

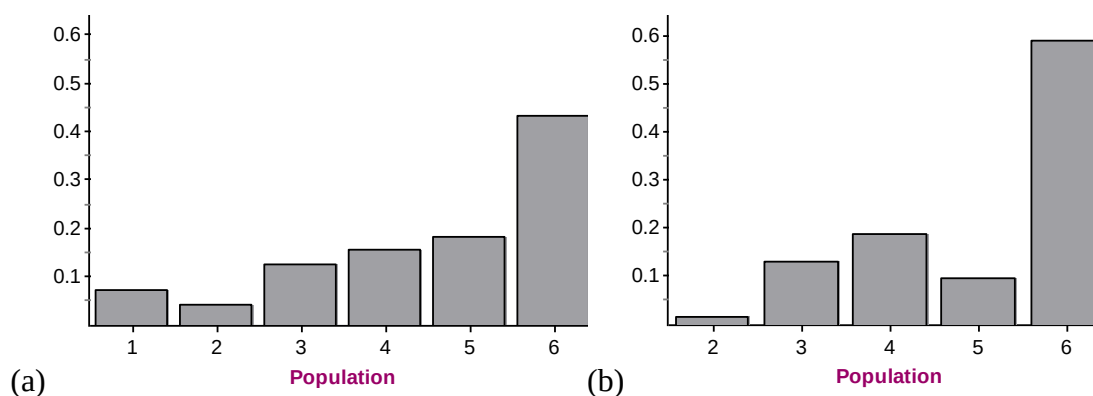


**Figure 6-23. Population categories for all post-test comparison questions: (a) 2008, and (b) 2011**

No students in 2011 used the sample (category 1) as the population and only one student in 2011 for one question used people generally as the population (category 2), and that was in

just one question. There were similar proportions of questions using boys and girls as populations (12.5% in 2008, and 12.6% in 2011), slightly more questions about populations of New Zealand boys and girls in 2011 (18.4% compared with 15.3% in 2008), and a bigger proportion of acceptable populations (category 5 and 6) overall in 2011 with 67.7% of the comparison questions posed in 2011 having acceptable populations compared with 61.2% in 2008.

### **6.7. Discussion**

In the literature review (see chapter 2) it was argued that there were two types of questions: those that are more formally posed and those that are spontaneously asked. Posed questions can be investigative questions or survey questions, with an investigative question being one that is asked of the data while a survey question being one that is asked to get the data.

This chapter has described the journey towards clarification of what makes a good investigative question in the school sector in the New Zealand setting. The journey started with the literature and initial teacher ideas, and these ideas were confirmed, added to and clarified. The researcher focused on what was important when posing investigative questions because this is an area that had been shown to be bereft of research in the literature review and an area that was identified as problematic in the first teaching experiment (see chapter 5). The chapter showed the progression to establishing criteria for what makes a good investigative question, and proposed two-way classification frameworks for summary and comparison investigative questions. The classification framework for comparative investigative questions then quantified the level of comparative investigative questions posed by students in year 10 (ages 14–15).

The first research question was “What makes a good investigative question?” A good investigative question is one that: has both the variable(s) and population(s) clear; has the intention clear; is able to be answered with the data; is about the whole group; and is interesting. These criteria were clarified through analysis of the in-class activities, student-posed questions (primarily in pre- and post-tests), and discussion with colleagues.

In addition to the criteria for what makes a good investigative question, classification frameworks were developed for summary and comparison investigative questions from a close analysis and critical reflection of student pre- and post-test data. The frameworks were grounded in the data with the summary framework reaching saturation after the third teaching

experiment, in the sense that the framework coped with all the fourth teaching experiment responses. While the comparison framework also suggested saturation after the third teaching experiment, this was not confirmed as minor adjustments were made after the fourth teaching experiment based on further student examples from this final teaching experiment. The classifications for summary and comparison investigative questions show the complexity of posing investigative questions, including the need to classify the population described in the investigative question separately from the question classification. This research explicitly articulates the thinking behind questions that had previously not been addressed, and more specifically for investigative questions, through both the criteria and the classification framework.

The second research question, “What are the underpinning concepts that are needed to support the teaching and learning around posing investigative questions?”, is now discussed. This research has shown that teaching posing investigative questions is necessary and it is something that needs careful consideration by the teacher as to how they will incorporate it into their teaching and learning programme. Pedagogical aspects to consider include: the types of investigative questions to address and when; at what point in the teaching and learning sequence will the teacher introduce posing investigative questions; and how or if the teacher will let the students “discover” for themselves the criteria for what makes a good investigative question. When students are taught about posing investigative questions it appears that they are able to make progress in this area. In addition this research has shown that there was a need for research in this area on posing investigative questions, an area that has not been researched before. The underlying big concepts and the components of what make a good investigative question include students needing: a sense of population; a sense of “tend” and “typical”; a sense of the variable(s); and an image of a hypothesised distribution(s) as they start posing their investigative question (see Figure 8-3).

The third research question in this chapter focuses on the level of question that students can pose: “What level of comparative investigative questions are year 10 (ages 14–15) students posing?” The results suggest that year 10 (ages 14–15) students are capable of posing comparative investigative questions that assume the idea of tendency and have an acceptable population descriptor; in other words they can pose “good” comparative investigative questions. In 2008 in a class of 24 students, 50% were at least at this level, and in 2011 in a class of 26 students, 54% were at least at this level. Most of the remaining students in both teaching experiments (46% in 2008, 27% in 2011) were posing comparative investigative

questions but their questions needed a bit of fine-tuning, mostly in terms of tidying up the population descriptor in the question.

Three key themes emerge from investigating this initial problematic situation about what makes a good investigative question. Firstly, language used in investigative questions needs to be precise. Precise wording is critical (Biehler, 1997; Pfannkuch, Wild, et al., 2009) as “loose” or non-precise wording can cause confusion and lead to poorly formed questions. Secondly, there are a number of statistical ideas and concepts that need to be developed concurrently. These include sample and population and the connection between the two (see chapter 7), and ideas around tendency and typical. Finally, the third key theme is that a statistical investigation is about more than just comparing or calculating simple measures; it is about students thinking distributionally (see chapter 8), describing what they see in the sample(s) they have selected (see chapter 8), and then make inferential statements (see chapter 7) about what may be happening back in the population(s) (Pfannkuch, Wild, et al., 2009).

These three themes provide the foundations for further exploration related to posing investigative questions. Having posed investigative questions, students need to be able to investigate the situation and come to a suitable conclusion. Each of the two types of investigative questions explored (summary and comparison) have different analysis needs. Exploring student post-test responses from the second teaching experiment, particularly focusing on comparison questions (as needed for the following schooling year), it appeared that students were not able to draw conclusions about what was happening back in the populations from the sample data they were given; i.e. they were unable to make the call whether A tended to be bigger/faster/longer than B. This is the focus for chapter 7.

Retrospective analysis from the third teaching experiment (where the focus was on making the call) identified describing distributional shape as an issue, and this links directly to students thinking distributionally and hypothesising about distributions. Describing distributions was an additional focus in the fourth teaching experiment as it was felt that this was a missing link. Describing distributions is discussed in chapter 8.

Figure 6-24 (next page) shows the connections between posing investigative questions and the other aspects explored in this research. Answering comparison investigative questions through making a call is the focus of chapter 7 and, in particular, connects with the concepts of sample and population, sampling variability, and making an inference. Answering

summary investigative questions through describing distributions is the focus of chapter 8, and in particular the focus is on hypothesising distribution and distributional thinking.

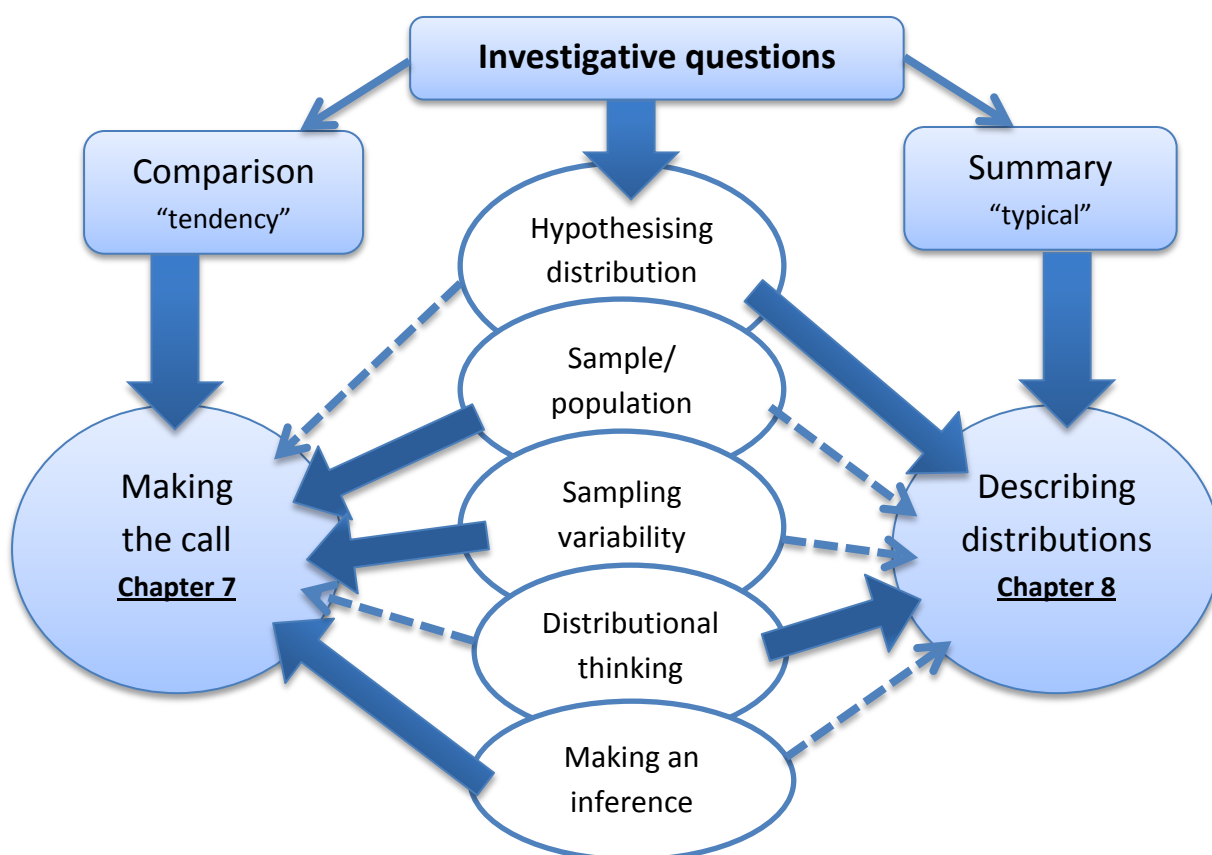


Figure 6-24. Link between investigative questions, making the call and describing distributions

## 6.8. Practical implications of the research

This research started in 2007 and aligned with the implementation of the New Zealand Curriculum (Ministry of Education, 2007). From 2007–2010, one of the achievement standards in the national qualification (NCEA level 1) in mathematics and statistics required students to pose a question and use statistical methods to respond to it. This achievement standard was similar to AS91035 which was mentioned previously.

In 2008 the process of aligning the achievement standards to the new curriculum was started. In short, all three levels of NCEA achievement standards have been realigned to the new curriculum. Posing an *investigative* question is explicitly stated in a number of the statistics achievement standards, and this reference to investigative questions is a direct result of this research. See Appendix E for specifics related to NCEA achievement standards.

## Chapter 6 – What makes a good Investigative Question

To support the implementation of the realigned achievement standards the Ministry of Education developed the *Senior Secondary Guides* and in particular the *Mathematics and Statistics Senior Secondary Guide* (<http://seniorsecondary.tki.org.nz/Mathematics-and-statistics>). Three achievement objectives at curriculum levels 6–8 (ages 15–18) have a link to investigative questions. Further details of curriculum resources can be found in Appendix E.

During the period of the research a number of presentations directly related to posing investigative questions were given to teacher and researcher groups (see Appendix F), and conference papers were presented at ICME-11 in Monterey, Mexico (Arnold, 2008b) and at the SRTL-6 Forum in Brisbane, Australia (Arnold, 2009).

## **Chapter 7. Answering the Investigative Question: making the call**

### **7.1. Introduction**

Chapter 6 explored what makes a good investigative question. Students in year 10 (ages 14–15), having posed investigative questions, need to be able to use informal inferential reasoning in comparison situations to answer the investigative question they have posed. In other words, students need to be able to “make the call” whether condition A tends to have bigger/faster/longer values than condition B back in the populations and provide statistical evidence from their statistics and plots to support their call – and this needs to be built on a solid understanding of the underpinning concepts such as sample, population and sampling variability.

The problematic situation, defined following a review of student post-test responses in 2008, is described. A review of the related literature is given in this chapter because the theme for this chapter was not identified as a focus initially and is not attended to in the literature review (chapter 2). The specific planning and preparation around sample, population, sampling variability and making the call for the 2009 and 2011 teaching experiments is detailed, providing context and rationale for the different decisions made about teaching and learning activities. The retrospective analysis that was undertaken is described and findings from this are given. Finally, the quantitative analysis of student test responses that was conducted in order to determine the progress of students from pre- to post-test in cognisance of the research development incurred as a result of the teaching experiments is presented.

The research on which this chapter is based was initiated through discussion with the researcher’s supervisors and included making strong links back to the work that Pfannkuch (Pfannkuch, 2006; Pfannkuch & Horring, 2005) had previously undertaken in year 11 (ages 15–16) classes. The end result was the undertaking of a Teaching Learning and Research Initiative (TLRI) research project in 2009 and 2010 that explored students’ informal inferential reasoning (Pfannkuch et al., 2011). The researcher was one of the two statistics education researchers in the TLRI team and led the teaching and learning programme development in the project. The research also builds on and aligns with the work undertaken on statistical inference by Wild, Pfannkuch, Regan, and Horton (Pfannkuch et al., 2010; Wild, Pfannkuch, Regan, & Horton, 2011).

## 7.2. Problematic situation

The curriculum and associated teaching and learning in year 10 (ages 14–15) mathematics and statistics are in part preparation for NCEA level 1 (New Zealand national assessment for qualifications) in year 11 (ages 15–16). NCEA level 1 is perceived as being the culmination of the first three years' work in secondary school, i.e. years 9–11 (ages 13–16). With this in mind, the student responses to the comparison situation in the post-test in 2008 were reviewed to see what level of analysis the students were doing and what types of conclusions they were drawing to better support future research and planning. Task C in the 2008 post-test required students to complete the analysis and conclusion parts of the PPDAC cycle for a comparison situation. They were provided with the problem (the investigative question), the plan, and the data in the form of graphs and tables of statistics (start of the analysis – Figure 7-1).

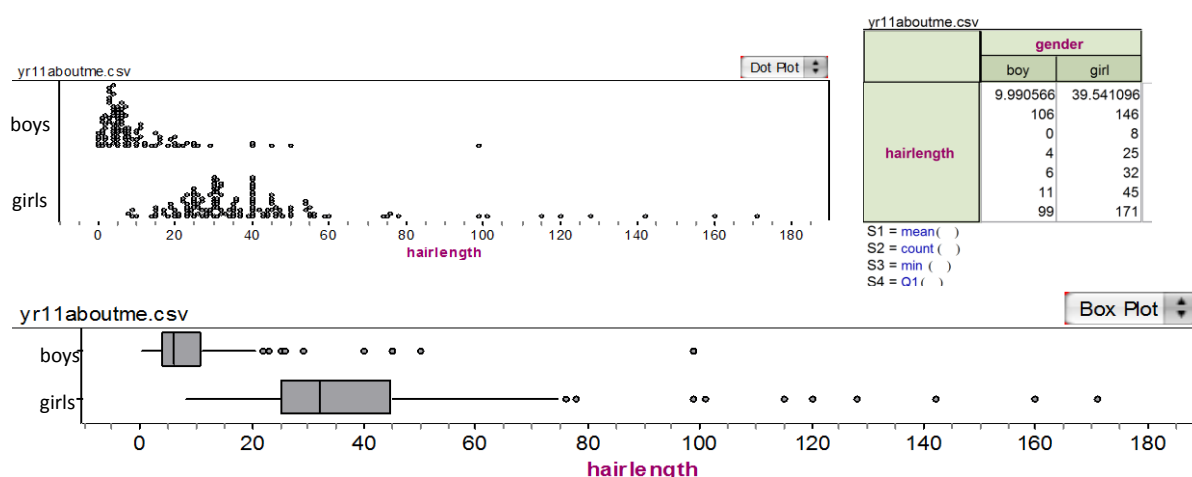


Figure 7-1. (a) dot plot of hair length, (b) associated sample statistics (c) box plot of hair length from Task C in post-test, 2008

The analysis that the students made in the post-test lacked any real depth and minimal connection to the context, which in this case was hair length of New Zealand year 11 boys and hair length of New Zealand year 11 girls. However, the analysis is not the focus for this chapter; rather, it was the students' responses in the conclusion part of the PPDAC cycle that is of interest. There were three headings for the students to respond to in the conclusion: answer to the problem, support for this answer, and generalisation of the sample findings to the population. The section where students wrote their support for their answer gave some indications of what they were using to make the call regarding New Zealand year 11 boys tending to have shorter hair than New Zealand year 11 girls. These are now discussed briefly.



### 7.2.1. Student post-test responses for conclusions 2008

The evidence that the students used to support their answer to the problem tended to include either one or both of two types of evidence. The first type of evidence (type 1) was where they compared individual features that appeared to be bigger (or smaller); for example, medians, means, maximums, minimums, and upper and lower quartiles. The second type of evidence (type 2) was where they compared blocks of data (the idea of shift (Pfannkuch, 2006)) rather than individual features; for example, “75% of the girls data is higher than 75% of the boys” (2008 student, post-test). A summary of their responses is given in Table 7-1. The examples of the type 1 and type 2 evidence are indicative; the actual individual student responses varied but were of a similar theme.

**Table 7-1. Student responses to support for the answer in 2008 post-test**

| Type 2 evidence \ Type 1 evidence         | 75% of girls data is higher than 75% of boys data | 50% of the boys are lower than all of the girls | 75% of the girls are longer than all of the boys (incorrect statement for the data) | No type 2 evidence (only type 1 evidence) | Total Type 1 evidence |
|-------------------------------------------|---------------------------------------------------|-------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------|-----------------------|
| States both mean and median higher        | 1                                                 |                                                 | 1                                                                                   | 2                                         | 4                     |
| Girls' middle 50% is longer               |                                                   | 1                                               |                                                                                     |                                           | 1                     |
| Girls' median is higher                   |                                                   | 3                                               | 1                                                                                   |                                           | 4                     |
| All statistics for girls are higher       |                                                   |                                                 |                                                                                     | 4                                         | 4                     |
| Girls' mean is higher                     |                                                   |                                                 | 1                                                                                   | 5                                         | 6                     |
| No type 1 evidence (only type 2 evidence) | 1                                                 |                                                 | 1                                                                                   |                                           | 2                     |
| Total Type 2 evidence                     | 2                                                 | 4                                               | 4                                                                                   | 11                                        | 21                    |

Students were most likely (19 out of 21) to support their answer using type 1 evidence, with just over half (11 out of 21) using this type only. There were five different responses for type 1 evidence, with four of the five including either the mean or the median, i.e. comparing a measure of centre. Eighteen of the 21 students including at least one statement that compared a measure of centre.

Ten students used type 2 evidence. Of the three ideas (75% of girls higher than 75% of boys; 50% of boys lower than 100% of girls; and 75% of girls higher than all the boys) only the

first two are correct statements. The last statement may have arisen with students being confused by the marked outliers and not including them in the data they were considering. Regardless of this confusion, both type 1 and type 2 evidence suggest that there are a range of ideas and options students are using to make the call, and while there are similarities amongst them, there is not a strong consistency in the evidence being used by the students to answer the investigative question.

### **7.2.2. New problematic situation**

Students at the end of the second teaching experiment were making the call based on a variety of statistics or blocks of data. There were nine different broad sources of evidence the students were using and these sources of evidence tended to be well spread amongst the students with no one source of evidence featuring strongly. Having said that, there were commonalities across the evidence sources, with measures of centre featuring strongly in nearly all of the evidence provided.

What was clear was that there wasn't an agreed understanding between the teacher and her students as to what constituted support for an inference. Furthermore, the investigative question was about the populations and the students' reasoning was based on describing the sample statistics not on making inferences about the populations. There were no criteria that were transparent for students and teachers to use. The problematic situation that now arose was building students' capability to answer the investigative question coherently and with consistency, using informal inferential reasoning.

The new problematic situation involves three research questions:

- (1) What underpinning concepts do students need to support them to make a call at curriculum level 5 (ages 13–15)?
- (2) Can year 10 (ages 14–15) students consistently and coherently make a statistical inference?
- (3) What evidence do students use to make the call at curriculum level 5 (ages 13–15) given suitable learning experiences for developing criteria to make a call?

### **7.3. Literature review**

#### **7.3.1. Informal inferential reasoning**

Informal inferential reasoning in statistics has recently become a focus of research (Pratt & Ainley, 2008). Zieffler, Garfield, delMas and Reading (2008) characterise informal inferential reasoning as “the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples” (p. 44). Such inferences, however, are based on arguments that are particular to a class or to a student’s development and depend on what is acceptable to a teacher. In New Zealand a problem arose when the new curriculum (Ministry of Education, 2007) and subsequent national assessment required students to make informal inferences about populations from samples in year 11 (ages 15–16). In such a situation the arguments to support inferences can no longer be an agreement between a teacher and her class but must be based on criteria that are transparent to all teachers. Therefore Wild, Pfannkuch, Regan, and Horton (2011) proposed a developmental pathway for comparative situations from year 10 to year 13 (ages 14–18) for justifying how to make a call or make a decision about whether condition A tends to have bigger values than condition B back in the populations. Enabling students to make a call depends on building their understanding of a network of underlying interrelated concepts, such as reasoning about sampling variability.

Earlier research reported that students studied could calculate means but did not know means could be used to compare two groups (e.g. Konold & Pollatsek, 2002). Other research focused on identifying and enhancing students’ conceptions when comparing distributions (e.g. Bakker & Gravemeijer, 2004). But problems arose when students were making inferences or claims in comparison situations. Pratt, Johnson-Wilder, Ainley, and Mason (2008) noted that students and teachers being studied did not know whether they were reasoning about the data as if it were the whole population or about an underlying population from which the data were a sample. At the International Forum for Statistical Reasoning Thinking and Literacy in 2007, statistics education researchers collectively agreed that students from middle school onwards should be generalising beyond the data in hand to a population or process. Konold and Kazak (2008) argue that the recognition by the forum that students need deeper understandings of inference to make comparisons meant that there was an implicit acceptance that the concepts of chance or sampling behaviour also need to be addressed. This research is specifically focused on facilitating students to make a decision in

comparison situations, to draw inferences about populations from samples, and to take sampling variability into account.

### **7.3.2. Sampling variability reasoning**

Sampling variability reasoning is at the core of statistical practice but it has only recently received attention in school curricula and instruction. Typically, students reach the final years of high school, where they are explicitly introduced to notions such as different sampling methods and basic statistical inference from confidence intervals, without fundamental knowledge or experiences of sampling behaviour. Saldanha and Thompson (2002) state, “in statistics instruction it is uncommon to help students conceive of samples and sampling in ways that support their developing coherent understandings of why statisticians have such confidence in this practice” (p. 268). Their comment clearly points to a lack of attention to conceptual development in school curricula.

A carefully structured set of learning experiences is required if students are to understand and appropriate the sampling variability reasoning underpinning statistical inference. As Garfield and Ben-Zvi (2007) stated in relation to distribution, centre and variability, students “need help in developing an understanding of what these concepts actually mean and how to reason about them in an integrated way” (p. 386).

When teaching statistical inference at the senior levels, many studies have documented impoverished conceptual development of sampling variability reasoning. For example, Rubin, Bruce, and Tenney (1991) explain that senior high school students in their study tended to believe that a representative sample was one that was sampled correctly and that randomness was “not sufficient to explain sampling variability – some mechanism or bias must be postulated to explain it” (p. 318). Similarly, Saldanha and Thompson (2002), who focused on building senior students’ conceptions about sampling distributions in their study, comment that students “image of sampling did not entail a sense of variability that extended to ideas of distribution” (p. 264). Sotos, Vanhoof, Noortgate, and Onghena (2007), in a review of statistics education research on statistical inference, state students’ misconceptions concerning sample size and sampling variability might arise from “a profound lack of insight into the idea of variability in random events that has a direct impact on the understanding of sampling processes” (p. 102). They note school and university statistics courses fail to support students in developing an understanding of basic core concepts.

According to Franklin et al. (2005) “statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in data” (p. 6). Despite the importance of considering variation in statistics, researchers have only in the last decade begun to document students’ conceptions of variability. Since there are many types and sources of variability, there is only some research on sampling variability and it is mainly limited to chance settings. In chance situations Shaughnessy (2007) notes that for students to consider variability among samples, they need to develop distributional reasoning; i.e. students need to grow beyond their propensity to focus on the expected value and develop intuitions “for a reasonable amount of variation around an expected value” (p. 982). He also reports that there is some evidence students’ distributional reasoning can be improved if they conduct hands-on simulations. Applying these findings to the statistics setting it would seem that drawing out random samples of a fixed size by hand from a population and noting the sample median might start to develop students’ intuitions about sampling variability. But at some stage hands-on simulations become laborious for students and carefully designed technology tools that attend to students’ current knowledge of statistics can help to bridge them from “naive conceptions to richer, more powerful understandings of statistical concepts” (Shaughnessy, 2007, p. 995).

The problem, however, is not that simple. When researchers have attempted to support students’ reasoning about sampling variability and the sampling distribution, misconceptions have tended to persist. For example, Meletiou-Mavrotheris, Lee, and Fouladi (2007) found no significant differences between a traditional and a technological instruction environment on students’ grasp of statistical inference concepts, and stated:

Results from our study agree with the findings of the considerable research that has been done in the last thirty years ... that people tend to think deterministically and lack awareness or understanding of variation and its relation to sample size. (p. 76)

The sampling distribution is considered a key component of statistical inference and researchers suggest school students should be introduced informally to ideas of sampling variability and sampling distribution (Garfield & Ben-Zvi, 2007). The TLRI project team (Pfannkuch et al., 2011) believed beginners should not attend to sampling distributions as a separate idea but rather experience sampling variability behaviour, which is integrated and associated with the plots, through hands-on and technological experiences.

### 7.3.3. Using technological tools

delMas (1997) provides guidelines for the development of new cognitive technology-based tools for software and teaching. Some of his suggestions for fostering understanding in statistics are: recognise the roles encoding, prediction and feedback play in the development of understanding; integrate physical activities with computer simulations; emphasise the interplay among verbal, pictorial and conceptual representations when using instructional activities designed around simulations; use representations that are familiar to students; and make it clear to students which features in the software environment are important. Wild et al. (2011) also emphasise the importance of hands-on activities, visual imagery, verbalisations and the discourse of the teacher to bring meaning to pictorial representations. Their tools present the concept of sampling variability as an integral part of the familiar box-plot representation. Colour was used to focus students' attention on the properties and structure of sampling variability in order for students to encapsulate it as part of the representation. The "movie snapshots" simulations are dynamic representations (<http://www.censusatschool.org.nz/2009/informal-inference/WPRH/>). Greer (2009) commented about an earlier version of their tools:

The sample values are not represented numerically, which may well be significant since numerical values could cue computation, whereas the visual counterpart invites comprehension. Consider also how the process unfolds in time, and leaves a history ... that could stimulate episodic memory of the process that gave rise to it. (p. 701)

Greer notes that provided teachers promote sense making and flexibility of thought, students can draw on prior episodic memories that assist them to solve problems in new settings. The technological tools and the methods for making a call devised by Wild et al. (2011) are novel. They incorporate facets suggested by other researchers, build on the chance research of Shaughnessy (2007), draw on multimedia learning theories (Clark & Paivio, 1991; Mayer, 2009; Mayer & Moreno, 2002) and, in particular, connect verbalisations and visual imagery (Arnold, Pfannkuch, Wild, Regan, & Budgett, 2011; Clark & Paivio, 1991).

### 7.3.4. Theoretical framework

To gain an insight into students' sampling variability reasoning, a theoretical framework was identified and adapted to assist in the analyses of students' thinking. The framework is based on Makar and Rubin's (2009) framework for thinking about statistical inference. Makar and Rubin identified three components that they regarded as essential to informal statistical

inference: probabilistic concepts of uncertainty and variability; generalisation realised in concepts of aggregate and going beyond the data in hand to draw conclusions about a wider universe; and using relevant evidence from the data to form a decision. Using the Makar and Rubin framework, a network of sampling reasoning concepts are proposed (Figure 7-2). These concepts underpin each of the components that were specifically for the case of comparing two samples of quantitative data randomly selected from finite populations (Pfannkuch, Arnold, & Wild, 2012).

| Statistical inference for comparison of two samples of quantitative data<br>at New Zealand Curriculum level 5 |                                                               |                                                              |                                                                                            |
|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------------------------------------|
|                                                                                                               | Components                                                    |                                                              |                                                                                            |
|                                                                                                               | <i>Probabilistic</i>                                          | <i>Generalisation</i>                                        | <i>Evidence from data</i>                                                                  |
| <i>Verbalisations</i>                                                                                         | Articulating the uncertainty embedded in an inference         | Making a claim about the aggregate that goes beyond the data | Being explicit about the evidence used                                                     |
| <i>Underpinning reasoning concepts</i>                                                                        | Sampling variability<br>Uncertainty<br>Sample size (not @ L5) | Sample Population<br>Distribution (Chapter 8)                | Shift, Overlap,<br>Position of medians<br>Decision guide (Figure 7-7)<br>Shape (Chapter 8) |

**Figure 7-2. Framework for thinking about statistical inference and sampling reasoning**

Note: Adapted from Makar & Rubin, 2009, p. 85.

When students make a claim from the comparison of two box plots they need to: (1) know what relevant aspects to notice from what they can actually see in the plots; (2) draw on relevant sampling variability knowledge from what they have experienced; and (3) invoke contextual knowledge. In terms of the framework, the component *evidence from data* requires students to be explicit about the evidence used and is therefore based on what they can see. For the comparison of two box plots, students in year 10 (ages 14–15) can see and describe the shift or the location of one box relative to the other, the degree of overlap of the two boxes, and the position of the medians relative to the overlap. At year 11 (ages 15–16), students are also expected to comment on the overall visible spread. Based on this information and the sample size for each group, they can then apply a decision guide (see Figure 7-7, page 157) appropriate to the curriculum level they are working at. For the *generalisation* component, students: (1) see sample distributions; (2) need to know that they are inferring from samples about what is happening back in the two populations; and (3) may hypothesise the shape of the population distributions (see chapter 8) based on their contextual knowledge (Pfannkuch et al., 2010). For the *probabilistic* component, students need to: (1) draw on experienced visual imagery of what the plots might look like when taking sampling variability into account (and at level 6 (ages 15–16), also consider sample size); and (2) draw

on their experience of sampling variability to know that they are “pretty confident” about the call but not absolutely certain.

#### **7.4. Planning and preparation**

This section discusses the planning and preparation for the third and fourth teaching experiments with the lens placed on sampling variability, informal inferential reasoning, sample and population, with a view to making the call and answering the investigative question. The especially designed learning trajectory is focused on facilitating students to make decisions in comparison situations, to draw inferences about populations from samples, and to take sampling variability into account. Although some researchers, such as Watson (2004), have investigated students’ sampling reasoning in the sense of simply taking a random sample, research that explored students’ sampling variability reasoning in relation to sample-to-population inferences does not appear to currently exist.

##### **7.4.1. Planning and preparation for teaching experiment 3 (2009)**

One of the foci of the third teaching experiment was making the call in comparison situations. The planning and preparation included working with a research team of 11, comprised of statisticians and classroom teachers, to look at how to create situations where students could come to the decision rule for making the call for themselves. How to teach key statistical concepts were considered in great depth as part of this preparation. The underpinning concepts of sample, population and sampling variability were regarded as very important to the ultimate outcome of students successfully making the call.

##### ***Adaptations to teaching and learning activities – teaching experiment 3***

There were significant changes to the teaching and learning sequence from the second to the third teaching experiments and these are indicated by the shading in Table 7-2 (next page). Lesson 2 was discussed in the previous chapter on posing investigative questions. Lessons 5–8 and 12–15 reflect the significant changes that were made to support the work on sample, population, sampling variability and informal inferential reasoning (i.e. making the call).



**Table 7-2. Teaching and learning sequence for 2009**

| Lesson number | Abbreviated lesson content (full details in Table 4-4, page 76-78)                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Reflect on year 9 work and the PPDAC cycle.</li> <li>• Introduce CensusAtSchool data set and set up the context.</li> </ul>                                                                                                                                                                                                                                                                                                           |
| 2             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Posing investigative questions and looking at what makes a good investigative question.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students will identify some or all of the criteria for what makes a good investigative question through their critique and classification of the investigative questions (see Figure 6-3).</li> </ul>                                                |
| 3             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Describing summary graphs.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                 |
| 4             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Describing comparative graphs.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                             |
| 5             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Revisit posing investigative questions.</li> <li>• Identifying and clarifying the population.</li> <li>• Exploring sampling variability.</li> <li>• What is a sample and why sample?</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students are able to pose an investigative question about Karekare College population and critique their questions and improve as necessary.</li> </ul> |
| 6             | <ul style="list-style-type: none"> <li>• Students “discover” the need to use a sample to answer the question about the population.</li> <li>• Students anticipate what the graph of popliteal lengths will look like using any prior knowledge they may have.</li> <li>• Students start to acknowledge that samples from the same population for the same variable are similar but have differences, and to articulate what these similarities and differences are.</li> </ul>                                  |
| 7             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Can a sample tell us something about the population?</li> <li>• Informal inferential reasoning.</li> </ul> <u>Hypothetical learning process</u> <ul style="list-style-type: none"> <li>• Students are prepared to make statements about what they think the population(s) might look like based on the samples they have taken.</li> </ul>                                                                                            |
| 8             | <ul style="list-style-type: none"> <li>• Students are making statements about whether they think one group tends to have bigger values than another group “back in the population(s)”.</li> </ul>                                                                                                                                                                                                                                                                                                               |
| 9             | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Introduction to middle group and spread.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                   |
| 10            | <ul style="list-style-type: none"> <li>• Describing middle group position relative to one another.</li> <li>• Formalising box plots.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                 |
| 11            | <u>Learning objectives/goals</u> <ul style="list-style-type: none"> <li>• Using Fathom (Finzer, 2007) to draw dot plots and box plots.</li> <li>• Describing sample data.</li> </ul>                                                                                                                                                                                                                                                                                                                            |

| Lesson number | Abbreviated lesson content (full details in Table 4-4, page 76-78)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Exploring the direction and amount of shift of the middle 50% when comparing two groups.</li> <li>Exploring the consistency across multiple samples of the same size from the same populations.</li> <li>Initial exploration of making the call.</li> </ul> <p><u>Hypothetical learning process</u></p> <ul style="list-style-type: none"> <li>Students recognise the patterns in the height graphs – shift is inconsistent, sometimes the middle 50% of the boys’ data is further to the right than the middle 50% of the girls’ data, sometimes it is the other way around; the boxes overlap in all cases and sometimes this is a complete overlap; sometimes the boys’ median height is higher, sometimes the girls’ median height is higher; and the median heights are both within the overlap of the middle 50%.</li> <li>Students recognise the patterns in the time-to-school graphs – shift is consistent, the middle 50% of the bus travel time to school is always further to the right than the middle 50% of walk travel time to school; in most cases the boxes do not overlap, for some cases there is a small overlap of the boxes; median time to school for bus is always higher than the median time to school walking; at least one of the medians is outside the overlap of the boxes.</li> </ul> |
| 13            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Reinforcing the “making the call” message using “movies”.</li> <li>Developing clarity around the message.</li> </ul> <p><u>Hypothetical learning process</u></p> <ul style="list-style-type: none"> <li>Students are fluent in which situation they are looking at and what the “call” would be.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 14            | <p><u>Learning objectives/goals</u></p> <ul style="list-style-type: none"> <li>Exploring other data sets to “make the call”.</li> <li>Constructing conclusions for comparison situations.</li> </ul> <p><u>Hypothetical learning process</u></p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| 15            | <ul style="list-style-type: none"> <li>Students are able to “make the call” from a new data set using the ideas developed in previous lessons.</li> <li>Students are able to write a conclusion that is consistent with the samples they have in answer to their comparative investigative question about the populations.</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |

Two lesson blocks (5–8 and 12–15) support two major concept development stages: (1) sampling variability ideas (including the ideas of sample and population); and (2) informal inferential reasoning, i.e. making the call about what is happening back in the population(s) from sample(s) and considering sampling variability.

**(1) Sampling variability ideas (lessons 5–8):** A number of specific adaptations were made to the teaching and learning sequence to support the development of the notion of sampling variability. These adaptations were made through discussion with the TLRI research group, where suggested activities were discussed and trialled in project meetings and workshops. In particular, the changes linked to sampling variability are described below.

***Identifying and clarifying the population*** – After looking at the investigative questions the students had posed previously (in teaching experiments 1 and 2, and pre-test responses for teaching experiment 3), it appeared that there still needed to be some work on the idea of population and describing the population. In an attempt to develop this concept of population more clearly, a tailor-made population was created. This population, as mentioned in chapter

6, was a fictitious secondary school called Karekare College. The criteria for the development of this population data set were that: it needed to be big enough that students would not want to use all of the data; it needed to be small enough that the data set could be made up as data cards; and it needed to be a population that the students would be familiar with.

Each student in the Karekare College data set was an actual student from the 2009 New Zealand CensusAtSchool database. Data cards were made to represent the 616 students at Karekare College and these were collected into a bag to represent the population (Figure 7-3).



**Figure 7-3. Karakare College population bag with data cards**

Each “student” had information pertaining to 13 variables (see Figure 7-4, next page): Row 1 – ethnicity, age in years, year level; Row 2 – transport to school, time to school in minutes, height in centimetres; Row 3 – how they carry their school bag, weight of school bag in grams, popliteal length in centimetres; Row 4 – fitness level, length of index finger in millimetres, length of ring finger in millimetres; and gender (which is shown by the different colour of the cards – girls are pink and boys are blue; see Figure 7-3 and Figure 7-4).

|           |      |     |           |      |     |          |      |     |           |      |     |
|-----------|------|-----|-----------|------|-----|----------|------|-----|-----------|------|-----|
| Indian    | 12   | 9   | NZE       | 12   | 12  | NZE      | 14   | 10  | NZE       | 16   | 12  |
| motor     | 15   | 163 | bus       | 15   | 166 | motor    | 20   | 176 | bus       | 30   | 159 |
| one       | 9300 | 39  | two       | 8800 | 43  | two      | 6000 | 44  | two       | 6000 | 43  |
| quitefit  | 80   | 2   | littlefit | 93   | 84  | quitefit | 79   | 80  | veryfit   | 85   | 94  |
| Samoan    | 14   | 10  | NZE       | 15   | 11  | Chinese  | 16   | 12  | Israeli   | 15   | 11  |
| motor     | 15   | 153 | bus       | 30   | 172 | walk     | 20   | 172 | bus       | 30   | 170 |
| two       | 9000 | 39  | two       | 8300 | 46  | two      | 6000 | 47  | two       | 6000 | 53  |
| littlefit | 95   | 84  | quitefit  | 106  | 101 | veryfit  | 61   | 56  | quitefit  | 90   | 95  |
| Niuean    | 16   | 11  | NZE       | 16   | 12  | NZE      | 12   | 9   | NZE       | 15   | 10  |
| bus       | 30   | 161 | motor     | 5    | 167 | bus      | 15   | 159 | motor     | 3    | 152 |
| two       | 9000 | 42  | two       | 8000 | 45  | two      | 5400 | 41  | one       | 5100 | 41  |
| quitefit  | 80   | 90  | veryfit   | 85   | 90  | quitefit | 80   | 70  | littlefit | 70   | 67  |

Figure 7-4. Example of data cards

**What is a sample and why sample?** – The first investigative question that the students worked on using the Karekare College data was set up in such a way as to get the students to come up with the idea of using a sample rather than being told this was what they were to do.

Once it was established that a sample could be used, the sample size was restricted to about 30 for this level. This was for pragmatic reasons; namely, that a class is about 30 students, a sample size of 30 gives a good-enough approximation to make inferences using the methods developed, and samples of size 30 do not take too long when drawing graphs by hand. Sample size and random sampling were two concepts that it was decided not to broach at this year level as it was felt that too many concepts could overwhelm the students, losing the main focus.

**Exploring sampling variability** – Sampling variability is “the variation in a sample statistic from sample to sample” (Ministry of Education, 2010a). In the teaching sequence, sampling variability was explored initially through the students making dot plots with their individual samples of data cards and informally finding the middle and describing their data distribution. Students then looked at all the other graphs that had been created around the class and noticed what was similar and what was different between the graphs. This activity was designed to support developing the idea that different samples from the same population give similar but different information about the population. This idea was further developed in the informal inferential reasoning block of lessons where the concept of sampling variability underpins the decisions about whether or not to make a call.

**(2) Informal inferential reasoning (lessons 12–15):** The development of this sequence of activities involved the researcher and her supervisor initially brainstorming a possible

teaching sequence for making the call based on the proposed developmental pathway for comparative situations (Wild et al., 2011). This was discussed with the research team and then teaching and learning activities were proposed, trialled, debated, reviewed, trialled and tweaked again. The debate was robust and, at times, a tense process, which culminated in a fluid teaching and learning sequence for teachers in the team to trial with their classes.

The teaching and learning sequence had five distinct parts to it. Firstly, in a previous lesson, the students had explored a number of different comparison investigative questions using technology. At the end of the lesson the teacher had asked them to record on prepared graph grids just the box part of the data for each of the two groups that were being compared for two different investigative questions. The whiskers were excluded as this information is extraneous to building inferential concepts and, as Pfannkuch (2008) noted, diverts attention to anomalies in the data. The idea was for students to use blue for the median and red for the box part. The colour coding was designed to focus the students' attention on the relevant structures in the representation and was the same colour cues used in the animations or "movies" in the fourth part of the sequence.

The two investigative questions used were: (1) Do the heights of Karekare College boys tend to be greater than the heights of Karekare College girls?; and (2) Do Karekare College students who walk to school tend to get there faster than Karekare College students who take the bus? These two investigative questions had been chosen as they demonstrate nicely across a number of samples the two different situations for making a call (discussed in section 7.5, making the call, page 161). Student graphs were collected in and these were copied in preparation for the next lesson and the teacher displayed the collated graphs on the wall in the classroom (Figure 7-5, next page).



**Figure 7-5. Collated student graphs on classroom wall.**

Secondly, the students, working together in pairs, sorted the graphs in the group of samples (either heights or time taken to get to school); they were asked to look for patterns across the samples, focusing on what was similar and what was different. The students looked at each set of graphs separately, i.e. they looked at all of the height graphs for patterns and then all of the time taken to get to school graphs for patterns. In this activity students were exploring the direction and amount of shift in the middle 50% of the graphs in comparison to one another, and they were also looking at the consistency (or not) across multiple samples of the same size from the same populations.

Thirdly, the students made a decision about what they thought was happening back in the populations, i.e. whether they could make the call that the heights of Karekare College boys tended to be greater than the heights of Karekare College girls, or the call that Karekare College students who walk to school tended to get there faster than Karekare College students who take the bus. Using these decisions and the patterns generated previously, ideas about how to make the call were established.

Fourthly, these ideas were tested through looking at many more samples using technology, both through repeated sampling using Fathom (Finzer, 2007) and through pre-prepared “movies” ([http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop2/heights\\_2samp\\_dots\\_30.pdf](http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop2/heights_2samp_dots_30.pdf) and [http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop2/times\\_2samp\\_dots\\_30.pdf](http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop2/times_2samp_dots_30.pdf)). During these movies the students raised their hands to indicate which median was higher, showing in one situation the constant swapping of the hand raised and in the other situation no swapping of the hand raised (Figure

7-6). The learning goal of these activities was to develop clarity around the making the call message.



**Figure 7-6. Students in class raising hands**

Finally, a decision rule was developed with the students as to when they could make the call, or not, about whether condition A tended to have bigger values than condition B (Wild, Horton, Pfannkuch, & Regan, 2008). The decision rule that was focused on was the one for curriculum level 5 (ages 13–15) (Figure 7-7, next page). In this final stage of developing the decision rule, both supporting evidence and reference to sampling variability were discussed with the students as being part of the “package” required to answer the investigative question (i.e. to be able to give the conclusion in the PPDAC cycle). The curriculum level 7 (ages 16–17) decision rule uses informal confidence intervals for the population median, and at curriculum level 8 (ages 17–18) students move on to formal inference.

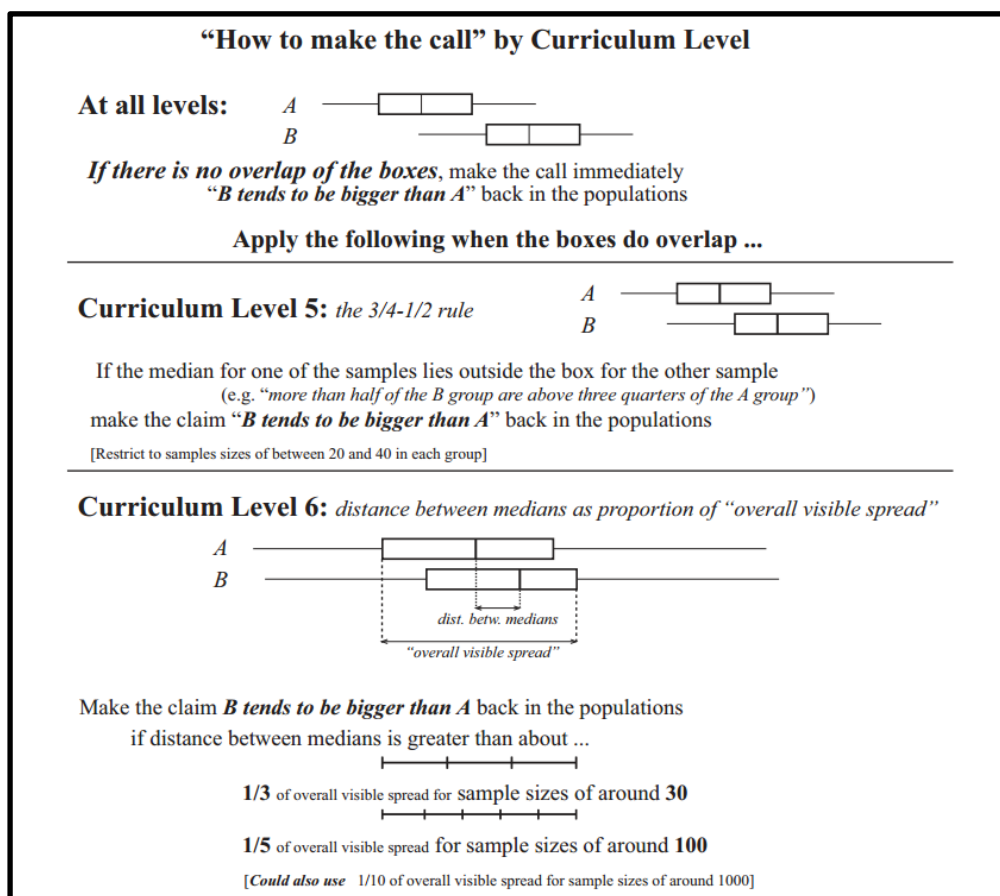


**Figure 7-7. How to make a call at curriculum levels 5 and 6**

*Note:* Reprinted from “Guidelines for “how to make the call”” by C. Wild, N. Horton, M. Pfannkuch and M. Regan, 2008. Retrieved from <http://www.censusatschool.org.nz/2010/documents/Shifts%20diagram.11.02.04.pdf>. Copyright 2008 by C. Wild, N. Horton, M. Pfannkuch and M. Regan. Reprinted with permission.

### ***Adaptations to the pre- and post-tests – teaching experiment 3***

The pre- and post-tests were changed for the third teaching experiment. In terms of answering the investigative question the changes were:

- A new task which specifically focused on: (a) students’ understanding of sample and population, (b) if they would support a claim (when the call couldn’t be made) and, in the post-test, (c) if they would support a claim (when the call could be made) – see task B 2009 pre-test and task C 2009 post-test in Appendix B.5 and B.6.
- Students completed an investigation only for a comparison situation, the contextual situation changed between the pre and post-test – see task C 2009 pre-test and task D 2009 post-test in Appendix B.5 and B.6.



### ***Interview questions – teaching experiment 3***

In the third teaching experiment, three students were interviewed following their pre-test and their post-test. The interviews were semi-structured with focused reflections on student responses in their tests (see Appendix D). In the pre-interview the students were asked, with reference to task B, about their understandings of sample, population, sampling variability and the basis for making a call (or not). They were asked in reference to the investigation (task C) why they may have (or have not) used the same justification for making the call as they did in task B and more about their understanding of sample.

In the post-interview, the same questions were asked for the questions that were the same (or similar in the case of completing the investigation). In addition the students were asked about specific learning from the unit of work focusing on the language that had been used in class; for example, Situation One or Situation Two (language students used to describe the two different situations for making a call), repeated sampling and overlap. They were also asked about the additional question (task C, part c) where they could make the call and their basis for making a call (or not). In the post-interview students were also shown two short “movies” similar to those that had been used in class and asked how the images helped them to make a call.

#### **7.4.2. Planning and preparation for teaching experiment 4 (2011)**

There was minimal adjustment to the teaching and learning sequence in 2011 around making the call as the material had had further refinement in the second year of the TLRI research project in 2010 (not included as part of this thesis).

### ***Adaptations to teaching and learning activities – teaching experiment 4***

The main adjustment was in the practice material used after the development of the decision rule for making the call. The practice material previously given to the students had each student using a different sample from the same population as they worked on the same investigative question. However, this had the effect of reinforcing the idea that they could use multiple samples to make the call – an unfortunate side effect that had not been anticipated. Therefore, in 2011, all the practice material involved the same single sample for all students in the class, reflecting what happens in reality, for each investigative question. The use of multiple samples from the same population was appropriate for developing the understanding

of making the call and sampling variability; however, it was not appropriate for subsequent practice as it created an unintended confusion for students.

#### ***Adaptations to the pre- and post-tests – teaching experiment 4***

The pre-test in 2011, in terms of answering the investigative question, had only the question that specifically focused on: (a) students' understanding of sample and population, (b) if they would support a claim (when the call couldn't be made), and (c) if they would support a claim (when the call could be made) – see task F, 2011 pre-test (in Appendix B.7).

In the post-test, this question remained but with a different context. Two additional sub questions were asked: one about whether there was anything else that could be investigated in the situation for part (b), and a second about whether the claim (for part (c), described above) made sense with what they personally knew about the situation – see task F, 2011 post-test (in Appendix B.8).

Students were not required to complete an investigation in either the pre- or the post-test.

#### ***Interview questions – teaching experiment 4***

The interviews for teaching experiment 4 (see Appendix D) were very similar to the interviews for teaching experiment 3. Six students were interviewed in pairs in both the pre- and post-interviews. In the pre- and post-interviews, with reference to task F, the students were asked about their understandings of sample, population, sampling variability and the basis for making a call (or not). They were also asked specifically for the definition of the terms *population* and *sample*.

### **7.5. Teaching experiments**

This section describes teaching moments that relate to answering the investigative question and are pertinent to the outcome of the research; for example, deliberate teacher acts, “aha” moments, or particular student insights or actions.

### 7.5.1. Teaching experiment 3

#### *Using the population bags*

As the “population” of Karekare College students was going to be used extensively over the next eight to ten lessons, it was important that the students in the class became familiar with the data that was available. This was done by getting the students to work out what the different variables were on the data cards through using the survey questions; the teacher also gave the students some hints about connections between variables in a row. It also helped that the students had participated in the census themselves. For more, see [www.censusatschool.org.nz/2009/informal-inference/teachers/workshop1/W1%20plan.pdf](http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop1/W1%20plan.pdf).

During the remaining lessons, whenever the teacher (T2, 2009) referred to Karekare College, she nearly always showed the population bag (see Figure 7-3, page 152), indicating that she was referring to the whole population, not just the data cards that the students had selected or the sample that had been selected when using technology. This action of showing the population bag was also used when students were posing investigative questions, reinforcing that the investigative question was about the population.

#### *Developing the idea of using a sample*

Having established the population and the variables that were available to use, the students posed a variety of investigative questions and one of these was “chosen” to be explored further. The students were to answer the question: “What are typical popliteal lengths of students at Karekare College?” The teacher asked them how they might go about answering this question, to which they ultimately replied that they would be “putting [the data] in a graph” (2009 student, lesson 5). There was then some discussion and the students started to graph **all** of the student data, using the data cards and a pre-prepared grid. After about 10 minutes some general discussion started about the “students” not all fitting onto the grid – “I’m not going to organise the whole college into this” (2009 student, lesson 5) – at which point the teacher asked, “Is there a better way than looking at the whole lot?” (2009 teacher, lesson 5). The ensuing discussion and action resulted in the students continuing until they had filled up their grid or felt that the shape of the graph was not changing despite adding more data cards, i.e. they did not use the whole population, just part of it. The teacher allowed the idea of using a sample to come from the students – she did not say to her class at the start, “Take a sample and use this to answer the investigative question.”

### ***Sampling variability***

Sampling variability was explored in a number of ways. In the lesson where sampling was first introduced, the students had created their graphs using the data cards, which provided a strong visual display. The teacher gave the students time to walk around the class and see how their graph compared with other graphs in the class. The students looked at features that were similar and features that were different. All of the groups gave an indication of where they felt the middle of their popliteal length data was, and across the class the middle popliteal lengths for the different groups lay within a 3–4-cm band. The students were able to see that the middle popliteal length was similar even though the samples were different.

Sampling variability was a focus again when students were looking at the patterns across the different samples of heights and time taken to get to school in the lesson about making the call. In particular, they noted that the samples of heights varied a lot from one sample to the next, noting that sometimes the boys' median was higher than the girls' median and sometimes it was the other way around. They also noted that for the data about time taken to get to school, although the position and size of the boxes and the position of the medians in the graphs varied from sample to sample, the location of the medians relative to one another remained the same, i.e. the median time to school by bus was always longer than the median time to school by walking.

### ***Making the call***

When the students were looking for patterns across the sets of graphs, additional prompts were required because information about the shift and the position of medians was not forthcoming from the students. According to Bodemer, Ploetzner, Feuerlein, and Spada (2004), leaving students to generate hypotheses about relationships on their own is very hard and they may not pay attention to salient features. Bodemer et al. (2004) suggest that learners' interactions with learning materials should be structured so that hypotheses are formulated only on one relevant aspect of the visualisation at a time, which the students did in this teaching experiment by first focusing on the distributional shift and then on which median was bigger.

After the students had sorted their samples for each question, the teacher and class reflected on the process. They described and abstracted the patterns and criteria for making a claim about what was happening back in the two populations. This allowed students an opportunity

to extract principles (Bakker & Gravemeijer, 2004). The students noticed that in the samples for the heights the boxes were close together, whereas in the samples for time taken to get to school the boxes were apart. They named these two situations about the relative location of the boxes Situation One and Situation Two, respectively. They also noticed that in Situation Two, there were consistent messages from the samples about the relative location of the two medians to one another back in the populations, allowing them to determine the larger of the two population medians. This was not the case in Situation One. Through recognising and reasoning from the patterns in the two situations they “discovered” collectively the criteria for making a call when two box plots are compared. In the following excerpt, the teacher explores the differences between the two situations.

*Teacher: So in our first situation we've got the boxes. They're all overlapping; some of them are going this way and some of them are going the other way. The medians are very close together and the medians are also within the overlap of the boxes. In the second situation, how is it different? What's different about the overlap here? Is there no difference between the overlap on these boxes and these boxes?*

*Student: They're not overlapped so much.*

*Teacher: They're not overlapped so much. No, they're not. Okay, do they all overlap?*

*Student: No.*

*Teacher: No, so when they do have an overlap they don't overlap much and otherwise they don't overlap at all. What can you tell us about the medians in this one?*

*Student: They're not overlapped.*

*Teacher: They're not in the overlap.*

Visually and verbally the students and teacher described the differences in the two situations in terms of shift, overlap, and the location of the medians. Students and teacher started to develop the criteria for making a call. Collectively they used hand gestures to describe the two situations, *close* (Figure 7-8a) and *apart* (Figure 7-8b), with vibrations, which according to Radford (2009) is a precursor to verbal conceptualisation.

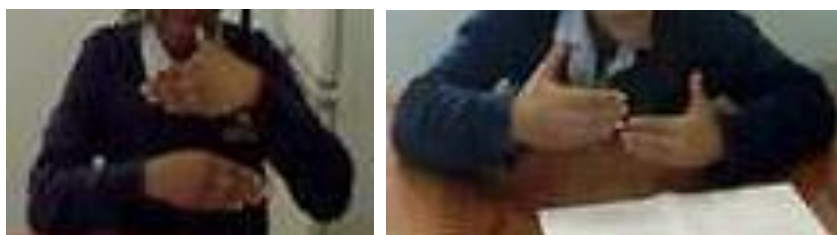


**Figure7-8. (a) hands close together (b) hands apart**

In the third teaching experiment, the language used with the students was around making a *claim*. In a handout (see <http://www.censusatschool.org.nz/2009/informal-inference/teachers/workshop2/IRON%20DATA%20Worksheet.pdf>) given at a workshop for teachers at the end of 2009, the structure for the conclusion (see Figure 7-9) included the phrase: “I would **claim** that ...” and this was based on the work that had been done in the class.

|                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Write a conclusion using the headings below.</p> <p><b>Answer the problem:</b></p> <p>“Do the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls?”</p> <p>I would <b>claim</b> that ...</p> <p>Explain why you have made this conclusion.</p> <p>Is my conclusion consistent with the literature?</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Figure 7-9. Structure of conclusion in 2009 worksheet

This language was used consistently throughout the 2009 work with the students encouraged to either state, for example, “I would claim that the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls” or “I am unable to claim that the iron levels of South Island urban boys tend to be lower than the iron levels of South Island urban girls.”

In discussion with the teachers involved in the TLRI project, it seemed that the students in their classes were more comfortable with the language “I would make the **call**” or “I cannot make the **call**” so this change from *claim* to *call* was adopted in the fourth teaching experiment.

#### 7.5.2. Teaching experiment 4

In the fourth teaching experiment there were no real aha moments regarding making the call. The teacher (T2, 2011) was teaching this material for the third year and was comfortable with and confident in the material to be covered. The idea of the overlap and position of the medians does take time to clarify as the following lesson transcript displays. This discussion

## Chapter 7 – Answering the Investigative Question: making the call

was in the third lesson (2011, lesson 15) on making the call, two lessons after the students had discovered the decision rule.

In the class transcript below, five statements regarding making the call have been highlighted as key statements about the decision rule; these statements will be discussed afterwards.

Teacher: *What about this one here?*

Student: *No.*

Teacher: *They're both within the overlap so we can't make a call can we. So sometimes there is an overlap. (1) Most of you got it – if there's no overlap you can definitely make a call but what happens when there is an overlap?*

Student: *It could change or it could be the other way around*

Teacher: *We know that if both the medians are inside the overlap then they could change quite a bit but if one of the medians is outside the overlap we'd expect that in another sample, wouldn't we?*

Student: *Yes.*

Teacher: *So we can make a call. If there is an overlap one of the things that helps us decide if we can make a call or not is whether both the medians are inside the overlap and if (2) **both the medians are inside the overlap we can't make the call and if one is then we can.** The average is higher than the middle 50% of the other group so we can make a call.*

Student: *So if one is in the overlap that means we can make a call?*

Teacher: *(3) If one median is outside the overlap.*

Student: *Oh, but if they're both ...?*

Teacher: *If they're both within the overlap then we can't make that call. So can we make a call?*

Student: *Yes.*

Teacher: *Okay, so can you see it now; okay, so we'll just write that up. (4) **When there is overlap we can make a call if one median is outside the overlap. One or more.** We've only been dealing with two populations but with box and whiskers and things, it's quite powerful, you can actually use it with more groups. So in our case (5) **if one or both medians are outside the overlap then we can make a call.***

In the first statement (1), the teacher clearly states that if there is no overlap then the call can be made. She then goes on to discuss situations where there is an overlap. In the second statement (2), she confirms that if both medians are inside the overlap, then the call cannot be made, but if one is then the call can be made. It is this second part of the statement that could

have potentially derailed the students’ understanding, changing the focus from at least one of the medians being outside the overlap which is the more obvious way to look at it. However, she corrected herself quickly (3) and confirmed this to be that you can make the call when one median is outside, and self-corrected that it could be one or more (4), which was reiterated in statement five (5).

It appeared that it was in the case of the overlap that students were confused and this short discussion served to help to clarify this misunderstanding. This was followed up with a note in students’ books: “When there is overlap, we can make a call if one or more medians are outside the overlap” (2011 student, class exercise book).

## 7.6. Retrospective analysis

Two of the three research questions for this chapter are explored in this retrospective analysis section: (1) What underpinning concepts do students need to support them to make a call at curriculum level 5 (ages 13–15)? and (3) What evidence do students use to make the call at curriculum level 5 (ages 13–15) given suitable learning experiences for developing criteria to make a call? The second question, (2) Can year 10 (ages 14–15) students consistently and coherently make a statistical inference?, is reflected upon in section 7.7 quantitative analysis of test questions.

Makar and Rubin’s (2009) modified framework for thinking about statistical inference, including sampling reasoning (Pfannkuch et al., 2012), as shown again in Figure 7-10 (a repeat of Figure 7-2), is used as a scaffold in answering both of the research questions in this section.

| Statistical inference for comparison of two samples of quantitative data<br>at New Zealand Curriculum level 5 |                                                               |                                                              |                                                                                            |
|---------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------------------------------------|
|                                                                                                               | Components                                                    |                                                              |                                                                                            |
|                                                                                                               | <i>Probabilistic</i>                                          | <i>Generalisation</i>                                        | <i>Evidence from data</i>                                                                  |
| <i>Verbalisations</i>                                                                                         | Articulating the uncertainty embedded in an inference         | Making a claim about the aggregate that goes beyond the data | Being explicit about the evidence used                                                     |
| <i>Underpinning reasoning concepts</i>                                                                        | Sampling variability<br>Uncertainty<br>Sample size (not @ L5) | Sample<br>Population<br>Distribution (Chapter 8)             | Shift, Overlap,<br>Position of medians<br>Decision guide (Figure 7-7)<br>Shape (Chapter 8) |

**Figure 7-10. Framework for thinking about statistical inference and sampling reasoning**

*Note:* Adapted from Makar & Rubin, 2009, p. 85.



The main sources of data were student pre- and post-test responses and student pre- and post-interviews. Field notes, class interactions and student class exercise books have been used to provide additional data. Findings from the TLRI project (Pfannkuch et al., 2011) have also been used to support the retrospective analysis in this chapter.

### **7.6.1. Underpinning concepts**

The underpinning reasoning concepts that students need to support them to make a call at curriculum level 5 (ages 13–15) include all three aspects of Makar and Rubin’s framework for thinking about statistical inference. In particular, at this level a focus was made on: the probabilistic concepts of sampling variability and uncertainty; the generalisation concepts of sample, population and distribution (see chapter 8); and the evidence from data concepts of shift, overlap, position of medians, decision guide and shape (see chapter 8). In this section, the notions of sample, population, sample to population, and sampling variability are discussed in more depth. It will be shown that students at curriculum level 5 (ages 13–15) are starting to develop these notions and use them in a statistical sense when answering investigative questions.

The evidence source in this section is primarily from the student interviews in 2009 and 2011. In 2009 three students (who are labelled S1, S2 and S3) were interviewed. Due to various circumstances only S1 has both a pre- and post-interview, while S2 has only a pre-interview and S3 has only a post-interview. In 2011 there were six students interviewed in pairs. These students are labelled S4, S5, S6, S7, S8 and S9. The pairs for the pre-interview were different to the pairs for the post-interview. In some instances both students in the pair have responded to a prompt and in other instances only one student has responded. This is by way of an explanation as to why the numbers reported in this section will not always give a total of nine.

#### ***Sample***

Initially students’ concepts about sample were able to be classified as either about a product sample (S2, S7), or part of a whole (S1, S4, S5, S6, S8, S9), similar to Watson’s (2006) findings. An example of the product sample was “like those shopping stores that like give you out free sample but you only get like a little bit” (2009 S2, pre-interview) and an example of part of a whole was “like a small portion of the larger thing so out of the whole of year eight she just took a sample of boys and girls, a small portion” (2011 S6, pre-interview). In

the post-interview all those students (S3, S5, S6, S8, S9) who responded to the question about what is a sample used a variation of part of a whole, improving this idea to the sample being part of the [whole] population; for example, “It’s a fraction of the total population which you can investigate with” (2011 S8, post-interview). The two students (S2, S7) who had the initial idea of a product sample did not have an individual response in the post-interview for this question so it is not possible to state if their position had moved.

Initially three of the students (S4, S6, S7) thought that a sample would tell you the average and made a tenuous link to the idea of this being of the whole [population]; for example “It will like the average length of year eight boys and girls” (2011 S6, pre-interview). This perspective was more clearly articulated in the post-interviews (S4, S5, S6, S8, S9); for example, “What the typical bag weights are and the distribution of year 11 boys’ bag weights and what they tend to be” (2011 S8, post-interview).

In addition, all the students in the pre-interview had a sense of what a random sample might be, though generally they (S1, S2, S4, S5, S6, S7, S9) were not able to articulate it clearly; for example “It means that they’re just chosen at random; they’re not selected in any sort of fashion” (2009 S1, pre-interview), and “Just take like any students for the group of people. No one in particular” (2011 S4, pre-interview). One student (S8) did articulate it more clearly, including an explanation of what it is not: “It means she may have picked them out of a hat or something. She didn’t go: okay, I want all of the foot samples of the Jack’s in New Zealand” (2011 S8, pre-interview). Again this perspective remained unchanged in the post-interviews. The concept of a random sample was not a feature of the teaching sequence as it was determined by the project team to leave the idea in abeyance at this level, but useful to note that the students did have some initial valid understanding of this concept.

### ***Discussion of sample***

Student responses in the post-interviews suggest that year 10 students (ages 14–15) can develop the notion of sample, an underpinning concept for generalisation (Figure 7-2/7-10) when making statistical inferences. In addition the students are starting to use the notion of sample correctly in a statistical sense. The students have developed the notion that a sample is a fraction or portion of the whole population and there are glimmerings of the concept of random sample. Technical statistical terms such as sample need to be used with students and regular discussion involving the definition may help to assist developing students’ understanding of sample to a deeper level.

### **Population**

Students' initial conceptions of population were mainly centred on the number of data values or the number of people, akin to thinking they were being asked: "What is the population of New Zealand?" Five of the students in the pre-interview referred initially to the population as the number of people, with four (S1, S2, S5, S6) responding with 60 (the size of the two samples combined) and a fifth (S4) "normally it's an amount of people" (2011 S4, pre-interview). With further prompting two students (S1, S4) amended their observations to reflect the actual population: "We're talking about all year eight students in New Zealand" (2009 S1, pre-interview), and "Year eight New Zealand boys and year eight New Zealand girls" (2011 S4, pre-interview).

Following the teaching experiment nearly all of the students (S1, S6, S7, S8, S9) who responded to the question about the population, responded with the actual population for the question. One student (S3) was still using the number idea – "30 year eight New Zealand boys" (2009 S3, post-interview) – while the remainder (S1, S6, S7, S8, S9) either used year 8 (or year 11) New Zealand students or year 8 (or year 11) New Zealand boys and girls. This suggests that the idea of population being the group that we are interested in investigating is becoming familiar to the students and they have moved beyond the notion of population being the number in a group. Also "full" population descriptors are being used (see Figure 6-10, page 124).

Only one pair of students was actually asked what was meant by population rather than just asking what the population was. The response suggests that they have a strong idea of the definition of population in the statistical sense and also linked this to the investigative question.

*The people who like you're taking the sample from, like if the sample was year nine to ten then your population would be the entire year nine to ten. It's the group you're concerned with in the investigative question (2011 S8, post-test).*

The definition of population was attended to in the course of lesson activity. In the class transcript below the teacher attended to both the description of the population and also the common misconception of population being the "number" of people. This attention was prompted by a reflection that the students had made at the end of the previous lesson.

*When I'm talking about Karekare College students that is my population. So the population is the description of the people that we are interested in, it's not the number of people in the*

*school, so a few of you are still, when we talked last night we asked you to write what was the population so remember when we talk about the population of New Zealand in that context we would say it's 4.3 million people but in this context when we're looking at a statistical investigation and we talk about the population we're describing the group of people that we're going to make statements about. So in this case the group of people that we want to make statements about is Karekare College students. So the population in this instance is Karekare College students (T2, 2011, lesson 7).*

### **Discussion of population**

Student responses in the post-interview suggest that their naïve understanding of population as the total number (of people) has matured into a deeper understanding and as a result students are able to identify the population correctly for the different situations. The students are developing an understanding of another of the underpinning concepts for generalisation (Figure 7-2/7-10), the notion of population.

The single response eliciting a definition of population made the link between the investigative question, the sample and the population, which is promising. As with sample, the notion of population needs to be attended to within the teaching and learning sequence frequently. Teachers need to repeat the definition, linked to the concept, throughout the teaching and learning sequence, as the example given has shown.

### **Sample-to-population ideas**

In the pre- and post-tests the students were required to complete the statement “Emma knows that her random sample of 30 year eight NZ boys can be used to find out ...”. The results discussed in this section include all student responses in the pre- and post-tests.

Initially, in the pre-test in 2009, approximately 65% of student responses to this question suggested that the students were thinking that Emma could find things out about her “random sample” (the sample), with the remaining 35% of students suggesting that Emma could find out about “year eight New Zealand boys” or find out about “boys and girls” (the population). In the 2011 pre-test, about 30% of the students suggested that Emma would find things out about her random sample (the sample), with the remainder (70%) giving similar responses to the 2009 students regarding year eight New Zealand boys and boys and girls (the population).

Therefore, initially in 2009 most students appear to be reasoning about sample distributions (the sample), whereas initially in 2011 most students appear to be reasoning about the population distribution (the population). The difference in these initial results could be due to

two factors: (1) the teaching in year 9 (ages 13–14) had changed at the school and the students had prior knowledge about samples and populations upon entering year 10 (ages 14–15), although no evidence was collected to clarify whether this could be a factor; and/or (2) the 2011 class was an above-average ability class, whereas the 2009 was an average-ability class, which may account for the difference in initial understandings.

The responses for the 2009 post-test showed a movement in the number of students who thought Emma could use her sample to find something out about the population. Two-thirds (67%) of the students wrote a response that could be classified as being about the population, approximately double the number in the pre-test. In 2011 the students who wrote a response that could be classified as being about the population also increased, from 70% in the pre-test to 81% in the post-test.

Sample-to-population ideas were dealt with directly in the class lessons. The teacher made a specific effort to remind the students about who the question was about, what the population was, and what a sample is used for.

*Do boys tend to be taller than girls? Now just an aside before we get to that, remember we were looking at our population, our Karekare College students, and all this that I've got is based on my sample so I could have said the evidence I've got is coming from my sample, so based on my sample in here. Based on my sample I'm picturing that Karekare College is going to be pretty much the same as I got from my sample. I've seen some of your comments, a few of you when you've been writing them up have actually been talking as though the sample was your population, and it's not. This is just a reminder that the population is the school, the times taken to get to school by the bus and the walking students and it's not just your sample. Your sample is giving you a picture of what's happening in the population (T2, 2011, lesson 14).*

### ***Sample size – linked to sample-to-population ideas***

Sample size was not a concept that was addressed in curriculum level 5 (ages 13–15), as it was perceived as slightly more complex and was deliberately left as a new concept for higher curriculum levels. However, analysis of the interview responses revealed untutored notions from the students that add to the current discussion.

In the pre-interviews, some students had initial ideas about how large a sample needed to be to enable them to make statements about all New Zealand year eight boys: students offered suggestions of adequate sample sizes that ranged from “more than 30” to “everyone”. One student even suggested that there needed to be “30 in each town because you can't judge them by 30 random people and expect everybody to be the same” (2009 S2, pre-interview).

In the post-test, student responses regarding the size of sample needed to enable them to make statements about all New Zealand year eight boys ranged from “just over 20” (being “better than taking 10”), through to “30 to 50 is good” and “5,000, like 10% of the population maybe”. One student (S3) in 2009 felt that a sample size 30 was fine but would like more than one sample. This also highlights the issue that arose in 2009 when students wanted to take multiple samples before they felt willing to make the call – an issue that resulted in an adjustment to the teaching and learning activities in 2011 (see page 158).

### ***Discussion of sample-to-population ideas***

Students are developing the notion that they can use a sample to make statements about the population; so in terms of the theoretical framework (Figure 7-2/7-10), they are starting to make claims about the aggregate that goes beyond the data, the generalisation component of making a statistical inference (Makar & Rubin, 2009). They are not yet convinced as to how small that sample can be but have tended towards “around 30” being “okay”. It needs to be remembered that sample size issues are not a requirement of the achievement objectives at this level, and so were not specifically taught or addressed in the teaching and learning sequence.

### ***Sampling variability***

Sampling variability is a reasoning concept for the probabilistic component of the framework for thinking about statistical inference (Makar & Rubin, 2009), whereas the previous concepts – sample, population, and sample-to-population ideas – are reasoning concepts for the generalisation component of the framework. The probabilistic component looks at the uncertainty of statistical inference (Makar & Rubin, 2009), a key aspect of which is the concept of sampling variability.

**Pre-interview:** Students’ initial intuitions about sampling variability were explored by asking the students in the pre-interview if they thought the graphs would be the same (as Figure 7-11, next page) if another person was to take a random sample of 30 year eight New Zealand boys and a random sample of 30 year eight New Zealand girls.

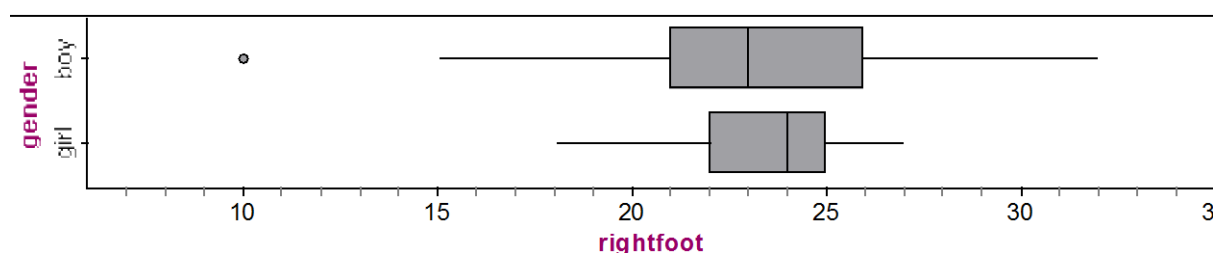


**Figure 7-11. Random sample of year eight New Zealand boys and year eight New Zealand girls (right foot lengths in cm)**

All the students responded to the effect that the graphs would be similar but not exactly the same; for example, “they would probably be similar, but not the same because they would have a different group of people obviously” (2011 S4, pre-interview).

After students had responded to the first question, they were then asked to sketch what they thought the graph might look like. The students again without exception drew graphs that had very similar features to the graphs in Figure 7-11: they drew the boys’ right-foot-length box wider and the girls’ right-foot-length box narrower and they put the girls’ right-foot-length median higher than the boys’ right-foot-length median (Figure 7-12). The medians were also within the overlap, though the significance of this would not have been considered.

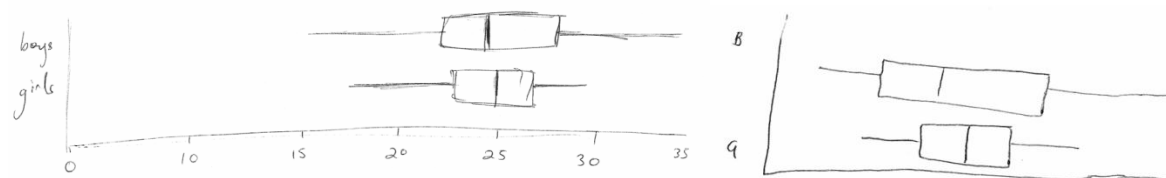


**Figure 7-12. Graph of repeated sample: (a) 2011 S8, pre-interview, and (b) 2009 S1, pre-interview**

They were also asked if they thought it was possible whether the person who took this other sample could claim that the right foot lengths of year eight New Zealand boys tend to be bigger than the right foot lengths of year eight New Zealand girls. The responses varied but the overall feeling was that it was possible; for example, “probably, probably. It just depends which random samples you happen to get and from where you got them” (2011 S8, pre-interview).

One student, S1, used a graph to show this (Figure 7-13, next page) and it is interesting to note how S1, the only one to do so, swapped over the interquartile range as well as reversing the medians. That is, S1 has made the graph with the boys’ right foot lengths having the

higher median and with a smaller interquartile range – as was the situation in the original graph for the girls' right foot lengths (Figure 7-11).

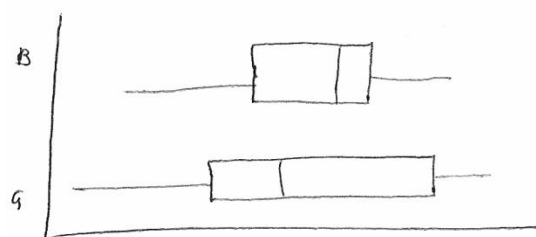


**Figure 7-13. Graph showing boys tending to have bigger right feet than girls (2009 S1, pre-interview)**

With regards to sampling variability, all students knew prior to the teaching intervention that another sample would produce slightly different plots. What they did not know was the extent of the sampling variability or really understand that the relative position of the medians in a situation such as in Figure 7-11 could swap around from sample to sample.

**Post-test:** In the post-test students articulated ideas around sampling variability in their responses to the reasons why they supported or not the claim made by Emma (2009) or Matt (2011). There were two questions; in the first (Q1), no call could be made, while in the second (Q2), the call could be made. The selected student responses are examples of responses to the two questions with a focus on the part of the response that related to sampling variability.

*(Q1) I know that another random sample could easily show the medians the other way round.*

*(Q2) I know that if another random sample was taken it most probably would show the medians the same way (2009 S1, post-test).*

*(Q1) The medians are within the overlap and could change position easily if another sample is taken.*

*(Q2) The medians are outside the non-existent overlap, which means that they are expected to remain in similar positions should Matt take other random samples (2011 S8, post-test).*

These two students have captured the essence of sampling variability with the idea that in the first question sampling variability could possibly see the medians swapped around and the second question would show the medians in relatively similar positions.

Some of the students had the idea that another sample would look different (or very similar) but did not articulate what they meant by that in their written response, i.e. they did not give a tangible link such as the position of the medians relative to one another, which the two examples above did. For example:



(Q1) *If we were to get another sample I would expect the results to be very similar (2011 S5, post-test).*

(Q2) *If another sample was taken, I would expect a similar result (2011 S4, post-test).*

**Post-interview:** In the post-interview students were once again asked about what it would look like if another person was to take a sample of 30 year eight New Zealand boys and 30 year eight New Zealand girls (or 30 year 11 New Zealand boys and 30 year 11 New Zealand girls – 2011). Without exception the students all sketched the graph with the medians the other way around to the situation in the given graph in the post-test (Figure 7-14).

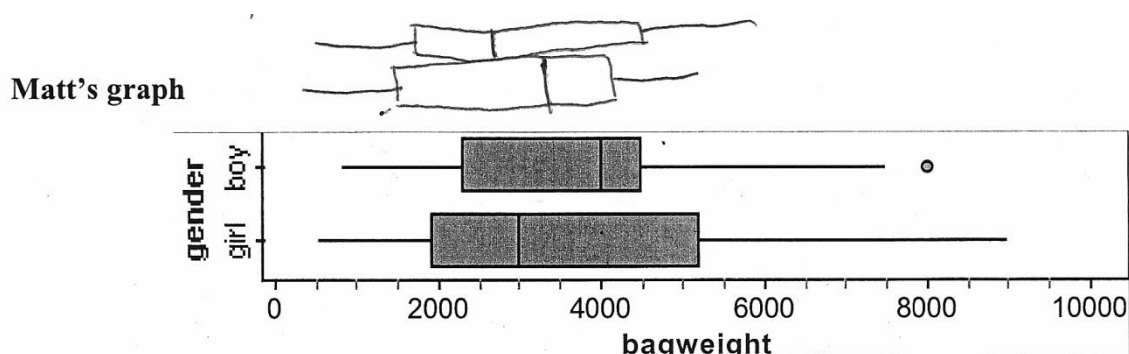


**Figure 7-14. Graph of repeated sample: 2011 S9, post-interview**

In 2009 in the post-interview, the students were specifically interviewed to see what enduring images they had retained. In response to the first situation, where Emma is comparing year eight New Zealand boys' and year eight New Zealand girls' right foot lengths, the following discussion ensued:

*Interviewer: So if you think that if this is Situation One, if just using your hands like the boxes, if I was to do the repeated sampling from the population, what would the graphs look like, what sort of image do you have of what would happen with the graphs? If I did a repeated sampling, so if there's like the boxes here?*

*S1: Okay, so maybe they would go like this and then maybe like this and the next one would be like this again and then maybe the next one would be like this [see Figure 7-8a]. There's not much difference but.*

*Interviewer: There'd be a tendency for them to move slightly backwards and forwards?*

*S1: Yeah.*

*Interviewer: Okay, and will the girls' median right foot length always be higher?*

*S1: No.*

*Interviewer: And would you expect the boxes to overlap?*

S1: Yes.

Similarly for the second situation, this time comparing Year 11 New Zealand boys' and Year 11 New Zealand girls' heights:

*Interviewer: Okay, and again just with your hands, if you can just show what a sample would look like, so where would the boxes sit relative to one another? If I was to take another sample what would you expect it to look like?*

S1: It would probably look something like this and this and say like this [see Figure 7-8b].

*Interviewer: So you'd expect the one that's to the right to stay to the right?*

S1: Yes.

### ***Discussion of sampling variability***

Student descriptive images combined with the student hand movements (Figure 7-8) suggest that the students are developing a fairly strong idea as to what sampling variability is. They can show with hand gestures and by drawing examples of possible graphs from another sample that they understand that different samples of the same size from the same population give similar indicators but are not exactly the same. They also can articulate that when there is a large overlap in the boxes of two groups being compared, and the medians are both within the overlap, the medians relative to one another can swap positions from sample to sample, whereas for the situation where there is little or no overlap and at least one of the medians is outside of the overlap it is unlikely that the medians will swap position so the signal is consistent from sample to sample.

In terms of the theoretical framework (Figure 7-2/7-10), the students are starting to articulate the uncertainty embedded in an inference (Makar & Rubin, 2009) by making statements about what another sample could look like; i.e. they are attending to concepts of sampling variability, which underpin the probabilistic component of making a statistical inference. They are also using the relationship between the overlap of the middle 50% and position of the medians to make statements about other samples. They are linking evidence from the data to this uncertainty.

### **7.6.2. Evidence used to make the call**

The evidence from data (Figure 7-2/7-10) that students can use to make the call at curriculum level 5 (ages 13–15) includes shift, overlap, position of medians, and the decision guide

(Figure 7-7, page 157). In this section, student pre- and post-test responses to questions requiring them to make the call are discussed in more depth. It will be shown that students at curriculum level 5 (ages 13–15) are making the call using mostly overlap, position of medians and the decision guide.

### ***Pre- and post-tests***

This section presents the responses of the nine students (S1–S3: 2009, and S4–S9: 2011) who were interviewed regarding two questions (Figure 7-15, next page) from the pre- and post-tests. Collectively these students represent a good range of responses to the pre- and post-test questions that are explored in this section.

Question one was in the pre- and post-tests in 2009 and the pre-test in 2011, question two was in the post-test in 2009 and the pre-test in 2011, and questions three and four were in the post-test in 2011.

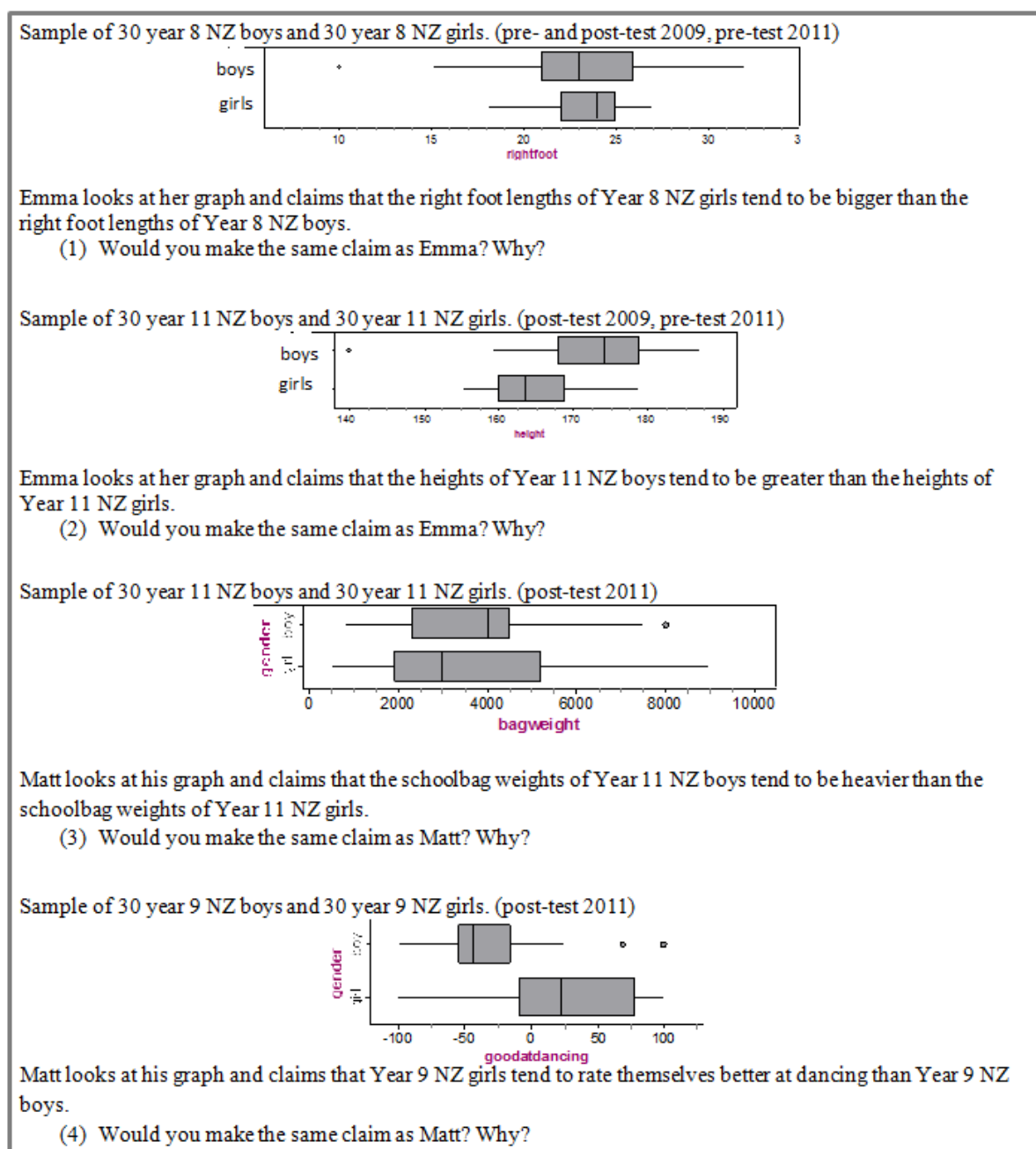


Figure 7-15. Abbreviated versions of pre- and post-test questions

**Pre-test:** In the pre-test (2009 and 2011), when making a call about the situation in question one (Figure 7-15), the students explained why they had agreed or not with the call using five different features: (1) using the size of the box or the middle 50% – three students; (2) using the centre, either average or median – four students; (3) calling on the maximum and/or minimum – two students; (4) using the median-to-upper-quartile range (incorrectly using most or majority) – one student; and (5) using the range – one student. Note the total is more than nine as some students used two pieces of evidence to justify their decision. The

## Chapter 7 – Answering the Investigative Question: making the call

following give examples of student responses regarding the five different features, which are numbered according to the feature.

(Note: in the 2009 pre-tests, boys' and girls' labels were missing on the test script due to printing error. Students assumed the top plot referred to girls.)

2009 responses:

- S2: (1) *Yes, because the box is bigger.*
- S3: (2) *No because the graph doesn't agree with Emma. It tells that year eight New Zealand boys have bigger average length with their right feet.*

2011 responses:

- S5: (3) *No, I would make the same claim as Emma because although boys have longer foot lengths, they also have shorter foot lengths.*
- S6: (4) *No, I wouldn't. Most of the boys' right foot sizes are between 23 cm and 26 cm, whereas the majority of the girls' feet sizes are either 24 cm or 25 cm.*
- S7: (5) *Maybe, the boys stretch over a bigger range whereas the girls are a tighter unit.*

For the second question in the pre-test (Figure 7-15; 2011 only, not given to 2009 students), the students used the following features to justify their decision: (1) using the centre, either average or median – two students; (2) calling on the maximum and/or minimum – one student; and (3) using the idea of shift, one graph moved further to the right than the other – three students. The following are examples of student responses to the three features, which are numbered according to the feature.

- S8: (1) *Yes, because the average height of a year 11 New Zealand boy is 176 cm, while the average for year 11 New Zealand girls is 163 cm.*
- S6: (2) *Yes, the tallest boys are between 173 cm and 178 cm, whereas the tallest girls are only between 162.5 cm and 168 cm.*
- S7: (3) *Yes, because the boys' results are higher in the table/graph than the girls'.*

Overall in the first question in the pre-test, most students used an incorrect feature (such as the size of the box, the range or the maximum) to make their claim (or not), while just under half used a correct feature (namely, measure of centre) to make their claim (or not). In the second question one student used an incorrect feature, two students a measure of centre, and three students the idea of shift. Therefore the basis for making a claim was not based strongly

on identified features that were used in the teaching units, such as shift and overlap – although a sense of shift is hinted at in 2011 in question two. This also reflects the identified problematic situation where there was no agreed criteria for making a call that were transparent to teacher and students (see page 143).

**Post-test:** In the post-test, one student (S2) had moved from using the size of the box to using the centre for making the call and also the idea of shift for question two.

S2: (1) *Yes the boys' box is bigger but the girls median seems to be higher than the boys.*

S2: (2) *Yes because the boys median is in a higher area than the girls.*

The remaining eight students had all moved to using the decision rule for question one (2009) and question three (2011). Of these eight students, seven (S1, S4, S5, S6, S7, S8, S9) also provided a second piece of evidence in their justification, referring to what would happen if another sample was taken. For example:

*No I would not make the same claim as Emma. I would not be prepared to make this claim because on Emma's box plots both the medians are in the overlap. This makes it hard to make an accurate claim because I know that another random sample could easily show the medians the other way round (2009 S1, post-test).*

*No, because the total ranges completely overlap each other, as do the middle 50% which means that the medians are within the overlap and could change position easily if another sample is taken (2011 S8, post-test).*

In question 2 (2009) and question 4 (2011), seven students (S1, S3, S4, S5, S7, S8, S9) used the decision rule, and of those, three (S1, S4, S8) also referred to taking another sample and two (S3, S9) included the idea of shift.

2011 – Example of decision rule and another sample

S4: *I would make the same claim as Matt as there is no overlap in the middle 50%. If another sample was taken, I would expect a similar result.*

2009 – Example of decision rule and shift

S3: *Yes because there is a tiny overlap and each medians are outside at the overlap the boys are further right than girls.*

The other student (S6) uses the idea of shift and taking another sample as their justification for supporting making the claim.

- S6: *Yes I would. Year 9 NZ boys only rated themselves between approx. -155 and -120 on the scale, while year 9 NZ girls rated themselves between approx. -10 and 75 on the scale if there was another sample taken I would expect it to be the same.*

Most of the students used two pieces of evidence to support their answer. Seven students used the decision rule in both question one/three and question two/four. This suggests that the students have an understanding of whether to make a call or not in comparison situations and what is required to justify this using the curriculum level 5 (ages 14–15) decision rule (Figure 7-7).

### ***Interviews***

Generally student responses to the interview questions reinforced the justifications that they had made in the pre- and post-tests so further retrospective analysis is not made. Of note, though, is the students' confidence in making the call for Situation One (overlap, both medians inside overlap) compared with Situation Two (little or no overlap, at least one median outside the overlap). In 2011 the students were specifically asked in which situation did they feel more confident with their call. Half of the students replied that they were more confident with Situation Two, because "they're really far apart" (2011 S5, post-interview). The other half of the students felt they had same confidence in both situations "'cause we're just following the steps that we were taught" (2011 S6, post-interview). Analysis of all students' post-test responses shows that they generally "scored" the same for both situations (see table in Figure 7-18b, page 185, where 18 of the 26 students had the same grade for both situations). In other words, they were as likely to disagree, with good justification, as they were to agree, with good justification.

### ***Discussion about making the call***

These students are developing ideas about making the call, inferring what is happening back in the two populations using samples. They are explicit about the evidence they used, in particular the degree of overlap and the position of the medians relative to the overlap. They are able to reason that another sample will be different and, depending on the situation, may give a similar or different signal (Situation Two and Situation One, respectively). In terms of the theoretical framework (Figure 7-2/7-10), these students are drawing on their developing reasoning concepts about overlap, position of the medians and the decision guide (Figure 7-7) which underpin the evidence from data component for making a statistical inference.

## 7.7. Quantitative analysis of test questions

One aspect of the pre- and post-tests is discussed in this results section in relation to the second research question: Can year 10 (ages 14–15) students consistently and coherently make a statistical inference?; namely, responses to the questions on whether or not the students would make the same claim as Emma and Matt (see Figure 7-15, page 177).

### 7.7.1. Answering the investigative question – focus on making the call

In the 2009 and 2011 teaching experiments, the students were asked in both their pre- and post-tests to agree or disagree with the claim(s) that Emma or Matt had made. These responses were analysed and used to inform a grading system based on SOLO (Biggs & Collis, 1982; Hook & Mills, 2011; Uniservices asTTle team, 2008; Watson, 2006), incorporating the Makar and Rubin framework (2009) and the categories proposed by Arnold et al. (2011, p. 21) for providing evidence for making a call (see Figure 7-16).

| SOLO taxonomy (categories for providing evidence) | Grade | Description                                                                                                                                                                                                                     | Example                                                                                                                                                                                                |
|---------------------------------------------------|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| No response or idiosyncratic                      | 0     | <ul style="list-style-type: none"> <li>No response</li> </ul> <b>OR</b>                                                                                                                                                         |                                                                                                                                                                                                        |
|                                                   |       | <ul style="list-style-type: none"> <li>Statement not based on the data or any feature of the data</li> </ul> <b>OR</b>                                                                                                          | No, because they are different. (2011 student, pre-test)                                                                                                                                               |
|                                                   |       | <ul style="list-style-type: none"> <li>Statement based on general knowledge of the situation; for example, some girls might have been tall and have long feet</li> </ul> <b>OR</b>                                              | Yes because boys in my family have bigger feet than mine, my cousins same age, boys have bigger feet than us. They are more strong and they grow faster when they grow older. (2011 student, pre-test) |
|                                                   |       | <ul style="list-style-type: none"> <li>Statement is as a given, but not reference to any evidence; for example, the graphs shows that right foot lengths of boys tend to be bigger than right foot lengths of girls.</li> </ul> | I would because the graph shows that generally NZ girls have a much longer foot length than boys do in NZ. (2009 student, pre-test)                                                                    |
| Pre-structural (irrelevant evidence)              | 1     | <ul style="list-style-type: none"> <li>Makes a call on any feature that appears to be bigger; for example, the maximum, upper quartiles, size of the box</li> </ul> <b>OR</b>                                                   | No, because the boys right feet reaches the length of 32 cm, and the girls right feet reaches up to 27 cm. DISAGREE. (2011 student, pre-test)                                                          |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a call by comparing centres but this is poorly verbalised; for example, the line for girls is higher</li> </ul> <b>OR</b>                                                          | Yes because the middle line for the girls graph is on 24 and the middle line for the boys graph is on 23. (2011 student, pre-test)                                                                     |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a call, but the evidence is contradictory</li> </ul>                                                                                                                               | Yes and no because the average right foot lengths for the girls is higher but the right foot lengths for the boys are more and could be                                                                |



## Chapter 7 – Answering the Investigative Question: making the call

| SOLO taxonomy (categories for providing evidence) | Grade | Description                                                                                                                                                                                                                                                                                                             | Example                                                                                                                                                                                                                                                                                                                                                                                                         |
|---------------------------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                   |       |                                                                                                                                                                                                                                                                                                                         | bigger, so I would probably agree with Emma? The boys right foot lengths are most spread out than the girls, most boys have bigger right foot lengths than girls. (2009 student, post-test)                                                                                                                                                                                                                     |
| Uni-structural (transitional)                     | 2     | <ul style="list-style-type: none"> <li>Makes a call by comparing the centres; for example, median, middle 50%</li> </ul> <b>OR</b>                                                                                                                                                                                      | <p>Yes because although the boys have a more varied right foot length the girls median is higher. (2009 student, post-test)</p> <p>Yes because the boys median is way past the girls median. (2011 student, post-test)</p>                                                                                                                                                                                      |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a correct call but poorly verbalises the evidence; for example, the data overlaps each other</li> </ul> <b>OR</b>                                                                                                                                                          | No, because both medians are inside the middle 50%. (2011 student, post-test)                                                                                                                                                                                                                                                                                                                                   |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes an incorrect call, but verbalises the evidence; for example, medians within the overlap/medians outside the overlap <b>or</b> the uncertainty; for example, sampling variability</li> </ul>                                                                                | Yes because the Year 8 NZ girls medians is more to the right than the Year 8 NZ boys but both medians are both in the overlap. (2011 student, post-test)                                                                                                                                                                                                                                                        |
| Multi-structural (towards relevant evidence)      | 3     | <i>Making a claim about the aggregate that goes beyond the data and being explicit about the evidence used or articulating the uncertainty embedded in an inference (Makar &amp; Rubin, 2009).</i>                                                                                                                      |                                                                                                                                                                                                                                                                                                                                                                                                                 |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a correct call <b>and</b> fully verbalises <ul style="list-style-type: none"> <li>the evidence; for example, medians within the overlap/medians outside the overlap <b>or</b></li> <li>the uncertainty; for example, sampling variability</li> </ul> </li> </ul> <b>OR</b> | <p>Yes, because there is no overlap meaning there is no median within that overlap (middle 50%). The median for boys is 30 and girls' median is 25. (2011 student, post-test)</p> <p>No because though the graph shows that girls have bigger foot length than boys if you take another sample it is possible that it could change to boys having bigger foot length than girls. (2009 student, post-test)</p>  |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a correct call but only partially verbalises the evidence and uncertainty</li> </ul>                                                                                                                                                                                       | No, I would not make the same claim as Matt because the medians of the middle 50% of the school bag weights of those Yr 11 NZ boys and the medians of those yr 11 NZ girls' schoolbag weights overlap. This means if Matt were to take a further amount of samples, the medians could change: it will either by yr 11 NZ boys who tend to have heavier school bags or yr 11 NZ girls. (2011 student, post-test) |
| Relational (relevant evidence)                    | 4     | <i>Making a claim about the aggregate that goes beyond the data and being explicit about the evidence used and articulating the uncertainty embedded in an inference (Makar &amp; Rubin, 2009).</i>                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                 |
|                                                   |       | <ul style="list-style-type: none"> <li>Makes a correct call (in context* and tendency*) <b>and</b> fully verbalises <ul style="list-style-type: none"> <li>the evidence; for example, medians within the overlap/medians outside the</li> </ul> </li> </ul>                                                             | No I would not make the same claim as Emma. I would not be prepared to make this claim because on Emma's box plots both the medians are in the overlap. This makes it hard to make an accurate claim because I know that another random sample could easily show the                                                                                                                                            |

| SOLO taxonomy (categories for providing evidence) | Grade | Description                                                                                                                                                                                                                                                                                                                                                                                               | Example                                                                                                                                                                                                                                                                                                                           |
|---------------------------------------------------|-------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                   |       | <p>overlap <b>and</b></p> <ul style="list-style-type: none"> <li>the uncertainty; for example, sampling variability</li> </ul> <p>*not possible for this assessment item</p>                                                                                                                                                                                                                              | <p>medians the other way round. (2009 student, post-test)</p> <p>Yes I would make the same claim as Matt because there is no overlap (of the boxes) and both the medians are outside the boxes. If I were to take another sample I would (be confident) expect to see the same based on my results. (2011 student, post-test)</p> |
| Extended abstract (full evidence)                 | 5     | <ul style="list-style-type: none"> <li>Makes a correct call (in context* and tendency*) <b>and</b> fully verbalises <ul style="list-style-type: none"> <li>the evidence; for example, medians within the overlap/medians outside the overlap <b>and</b></li> <li>the uncertainty; for example, sampling variability <b>and</b> the context is reflected throughout the conclusion.</li> </ul> </li> </ul> | No examples in student pre- and post-tests.                                                                                                                                                                                                                                                                                       |

Figure 7-16. SOLO criteria for grading making the call

### 2009 Results

In 2009 students in the pre-test answered one question where they were unable to make a call and in the post-test they had the same question where they were unable to make a call and a question where they were able to make a call. Thus for the question where the students were unable to make a call, there are both pre- and post-test responses. Their grades are summarised in the scatter plot and summary table in Figure 7-17 (next page). The line marked in Figure 7-17a is where the post-test grade equals the pre-test grade. Dots below the line signal where a student has a lower post-test grade compared with their pre-test grade and dots above the line signal where a student has a higher post-test grade compared with their pre-test grade. The numbers are confirmed in the table (Figure 7-17b).

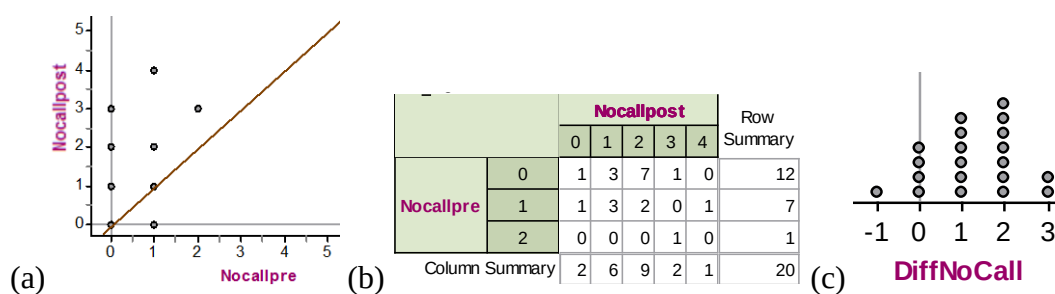


**Figure 7-17. Pre- and post-test responses for situation where no call could be made: (a) scatterplot showing pre- and post-test grade, (b) summary table showing pre- and post-test grade, and (c) dot plot showing difference in pre- and post-test grade**

Figure 7-17c shows the difference between students' pre-test grade and post-test grade for the situation where they could not make a call. A difference of zero indicates no change in their grade from the pre- to the post-test. Of the 20 students who sat both the pre- and post-tests, 15 improved their grade, four remained the same and one student lowered their grade. In the post-test, two students were working at a multi-structural level and one at a relational level. Nearly half of the students (nine) were using the centres to compare the two groups (year eight New Zealand boys and year eight New Zealand girls). Six students were still working at a pre-structural level and they were mainly making the call (or not) using the maximum or other similar irrelevant features of the data.

### 2011 results

Twenty-six students completed the pre- and post-test. As there were two questions in both the pre- and the post-test, students were given a final pre-test and a final post-test score; this was the mean of their two SOLO grades. These final scores have been analysed to look at the difference between pre- and post-test grades for these students. Their final scores for making the call are summarised in Figure 7-18 (next page). Figure 7-18a shows a scatterplot of students' pre-test mean score and their post-test mean score. The line marked is where the post-test mean score equals the pre-test mean score. The numbers are confirmed in the table (Figure 7-18b).

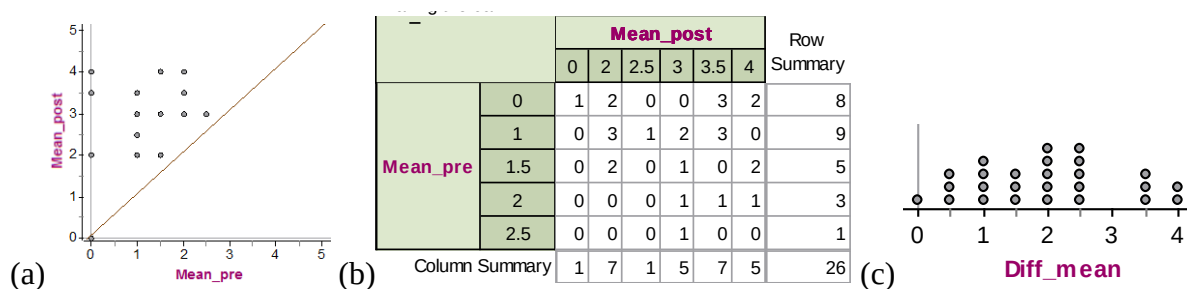


**Figure 7-18. Mean pre- and post-test scores for making a call: (a) scatterplot showing mean pre- and post-test scores, (b) summary table showing mean pre- and post-test scores, and (c) dot plot showing difference in mean pre- and post-test grades**

Figure 7-18c shows the difference between students' pre-test mean score and post-test mean score. A difference of two indicates a mean improvement of two points in their mean score from the pre- to the post-test. Of the 26 students who sat both the pre- and post-tests, all but one of the students improved their score. The student who remained as a grade zero made all of her justifications using her general knowledge rather than using statistical evidence:

*Yes, because girls do tend to exaggerate on what they bring to school where as boys just grab what they need for school. Sometimes boys also take an extra bag or tend to leave their things at home a lot more. Girls can usually bring extra things like make-up etc. for no good reason (2011 student, post-test).*

In the post-test, five students were working at a relational level overall and seven students had one response at multi-structural and the other at relational. Of these seven students, five had their “unable to make a call” at the relational level, while the other two had their “make the call” at the relational level. Five students were multi-structural for both situations and one was multi-structural for unable to make a call. The seven uni-structural students overall in the post-test tended to make the call (or not) using the medians. The students made significant improvement ( $p\text{-value} < 0.0001$ , paired  $t$ -test) in their mean scores from pre- to post-test in making the call and on average increased their mean score by 1.96 points (95% CI = [1.51, 2.41]).

In the post-test, students generally were able to agree or disagree with the claim made by Matt (17 students scored at multi-structural or above, which required a correct call). Twelve students were able to show in at least one of the situations a reasoned justification that included both explicit evidence (for example, there is an overlap and both medians are within the overlap) and articulation of the uncertainty (by referring to what would happen if another sample was taken).

The 2011 class, as mentioned previously, was an above-average-ability class compared with the 2009 class, which was of average ability. Factors that could account for the better results for the 2011 class may be the difference in class ability and/or that the teacher's pedagogical content knowledge has improved because of a deeper understanding of student misconceptions around making the call, which influenced the learning experience for the 2011 class.

### 7.7.2. Reflection on the lack of extended abstract responses

An extended abstract response was not shown in any of the student responses in the pre- and post-tests and this was probably due to the nature of the questions asked in the pre- and post-tests. The question asked them to agree or disagree with claims made, not to specifically make a claim. For students to give an extended abstract answer, it would be expected that their conclusion would include a correct contextual claim (or not) answering the investigative question. This claim would reference whether one group *tended* to have bigger values than the other group for the specific variable. They would also have sustained reference to the context (population groups and variable) throughout their response. A response of this nature is possible and was evidenced in the students' class assessment activity where they completed an investigation using the entire PPDAC cycle. In the conclusion, three students wrote an extended abstract response. An example of one is below.

*I cannot make a claim that yr 5–10 NZ students who have a cellphone tend to sleep less hours than yr 5–10 students who don't have a cellphone. This is because the middle 50% of students who have a cellphone and students who don't have a cellphone overlaps quite a bit. The median of students with a cellphone just fits into the overlap and the median of students without a cellphone is quite within the overlap. Because of these two factors I cannot say that yr 5–10 students with a cellphone tend to sleep less hours than yr 5–10 students without a cellphone.*

*I believe that if I took another sample of yr 5–10 students with a cellphone and yr 5–10 students without a cellphone I would get a similar result. The only thing that could differ could be the medians as they could change the other way around, but I don't believe that that would have a huge impact on the results. (2011 student, class assessment activity)*

In addition to the three students who gave an extended abstract response in the class assessment activity, nine students showed relational thinking, seven multi-structural thinking, five uni-structural and three pre-structural. Overall their grades were better than in the post-test, suggesting that they can show what they mean better in a free response situation, and/or they did further learning and clarification following the post-test which was three days before

the class assessment activity. What their assessment activity does suggest is that quality responses are possible at this level and they can be sustained and improved.

## **7.8. Discussion**

In chapter 6 the nature and elements of an investigative question were discussed. For students to understand how their investigative question could be answered, it was necessary to develop students' conceptual understandings of the relationship between sample and population, sampling variability, and making an inference, as well as building students' notions of decision criteria to make a call or judgement on whether one group tends to have bigger values than the other group.

In chapter 7 the design of the teaching sequence to develop the underpinning concepts of answering investigative comparative questions was discussed. Analysis of students' pre- and post-tests and pre- and post-interviews, classroom transcripts and field notes were all used to define and show how two groups of students in 2009 and 2011 progressed towards developing an understanding of sample, population, sample-to-population inference, sampling variability, and making a call. The three research questions for the study in this chapter were: (1) What underpinning concepts do students need to support them to make a call at curriculum level 5 (ages 13–15)? (2) Can year 10 (ages 14–15) students consistently and coherently make a statistical inference? and (3) What evidence do students use to make the call at curriculum level 5 (ages 13–15) given suitable learning experiences for developing criteria to make a call? Each of the questions is now addressed.

In response to the first question: the underpinning concepts that students needed to support their making a call at curriculum level 5 (ages 13–15) are summarised in the adapted theoretical framework (Figure 7-2/7-10) of Makar and Rubin (2009). According to the adapted theoretical framework, these students were beginning to understand how to make a statistical inference as they invoked all three components: the probabilistic component, by articulating the uncertainty embedded in an inference through ideas about sampling variability; the generalisation component, by making a claim about the population from the sample; and the evidence from data component, by articulating and being explicit about the evidence they used from the data such as the overlap, position of the medians, and the decision guide that enabled them to make a call.

In terms of the second question: the aim of building students' capability to answer the investigative question coherently and consistently using informal inferential reasoning seems to have been achieved with these students. The findings of Konold and Pollatsek (2002), Pratt et al., (2008), Meletiou-Mavrotheris et al., (2007) and other researchers that students in comparison situations did not know that means could be used or thought they were reasoning about the sample or were not aware of sampling variability seem to have been addressed with this teaching sequence and through the findings from the consequent responses from these students. Hence the teaching sequence, with its focus on underpinning concepts, using physical and dynamic visual simulations and allowing students to “reinvent” the decision rules for making the call, makes a contribution to the pedagogical content knowledge base for the wider statistics education community.

In this chapter, in answer to question three, it has been shown that students can attend to the appropriate evidence in the data to make a call. The two situations were designed to focus on one relevant aspect at a time (Bodemer et al., 2004) and to extract principles (Bakker & Gravemeijer, 2004) relevant to the two different situations. The evidence that the students used to make the call included the amount of overlap between the two boxes (in box plots) and the position of the medians relative to the overlap. They used these two pieces of evidence from the data in conjunction with the curriculum level 5 (ages 13–15) decision rule to make the call (or not) that condition A tended to have bigger/faster/longer values than condition B back in the populations. As a result of this research, the specific evidence that students at this age need and can access was realised, and the findings allowed the researchers to focus more clearly on pertinent features and concepts. Progress has been made towards resolving the problematic situation that there was no consensus on how to make a call. Students have moved from calling on summary statistics being higher to using an appreciation of sampling variability and shift, overlap, and position of medians to make their call.

The learning emphasis was not on the “rule” for the decision criteria; rather, it was on developing the underpinning concepts, such as sample-to-population and sampling variability ideas, that are needed to use the rule with understanding. The students in this study seemed to understand how and why the use of the overlap and position of the medians relative to the overlap informed their use of the rule to consistently and coherently answer their comparative investigative question.

## 7.9. Practical implications of the research

The research that this chapter is based on is part of a bigger TLRI research project. The collective resultant four journal articles, three conference papers, 14 key note presentations and more than 20 workshops from the TLRI project are outlined in Appendix G. These have been presented by all members of the TLRI team including the writer of this thesis. The associated teaching and learning activities that were developed for the TLRI project (as outlined in this chapter) and supporting teacher workshop material are all freely available to teachers in New Zealand and throughout the world on the New Zealand CensusAtSchool site. This material can be found at <http://www.censusatschool.org.nz/2009/informal-inference/>.

The work in this chapter on answering the investigative question through making a call in comparative situations has influenced the type of informal inference that students in New Zealand make at levels 5–8 (ages 13–18) of the curriculum (see Figure 7-7) and is reflected in the key mathematical ideas of the curriculum (see, for example, [http://www.nzmaths.co.nz/statistical-investigations-level-5?parent\\_node=](http://www.nzmaths.co.nz/statistical-investigations-level-5?parent_node=)), the *Senior Secondary Guides* (see, for example, <http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S6-1>), and the realigned achievement standards (see, for example, AS91035: Investigate a multivariate data set using the statistical enquiry cycle, <http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2011/as91035.pdf>).

This work has been foundational to the progression from informal inference through levels 5–7 (ages 13–17) to formal inference at level 8 (ages 17–18) through bootstrapping and randomisation techniques.



## **Chapter 8. Answering the Investigative Question: Describing Distributions**

### **8.1. Introduction**

Chapter 8 explores further aspects of answering the investigative question, and whereas chapter 7 focused on concepts underpinning inference and making the call in comparison situations, this chapter will concentrate on describing distributions with a focus on summary situations.

The problematic situation is described, making links back to chapter 6 and chapter 7. The rationale for the need for students to be able to describe the sample data distribution(s) using both descriptive and inferential thoughts and to conjecture the distributional shape is discussed. A review of the literature is given by way of setting the scene as the theme for this chapter was not identified as a focus in the initial literature review (chapter 2). The specific planning and preparation around describing shape, predicting distributions and describing distributions for the 2011 teaching experiment is detailed, providing context and rationale for the different decisions made about teaching and learning activities and assessment and interview tasks. The retrospective analysis that was undertaken is described and findings, including a distribution descriptive framework, are given.

### **8.2. Problematic situation**

Distributional shape and distribution were identified as two concepts of the statistical inference and sampling reasoning framework (Figure 7-2, page 148) that had not been attended to in the work that had been undertaken to date. Both of these were considered to be important, for the following reasons. Firstly, when verbalising a comparative investigative question, students should be encouraged to produce visual mental imagery of the situation, i.e. the location and shape of one population distribution relative to the other, since visual thinking is associated with high levels of functioning in mathematics (Clark & Paivio, 1991; Presmeg, 2006). Secondly, from a statistical perspective, visualising the expected distributional shape is part of statistical practice and thought, as any departure in the sample data from what is expected needs further investigation. Thirdly, the “story” in the data is revealed in the shape of the distribution (Pfannkuch & Wild, 2012). Statistics is about

unlocking the stories in the data and distributional shape is an important consideration in that quest. Both distributional shape and distribution, however, were in need of further attention. Based on student interviews, the Teaching and Learning Research Initiative (TLRI) project team in 2009 presented to teachers at a statistics day at the end of the year (Pfannkuch, Regan, & Arnold, 2009) their thoughts on why shape needed attention. In the presentation, inferential ideas that students have when they come to mathematics and statistics classes in year 10 (ages 14–15) were considered, including looking at the link between sample and population and what students are seeing when they look at sample distribution; for example, whether they are seeing the signal or the noise (Bakker, 2004a, 2004b; Bakker & Gravemeijer, 2004).

To start to address the issue, initial ideas were presented, based on student interview transcripts, as to what students might be seeing when they look at the distributions of samples. The relevant snippet from the presentation, given as an interaction between student and interviewer, is below and signals the starting point for the students when considering shape. The transcripts considered were recorded after the 2009 teaching experiment and specifically followed up on student responses to shape in the post-test, which was one of the descriptors of a sample distribution, and was set in the bigger context of a statistical investigation.

*Interviewer:* Sina, what do you see?

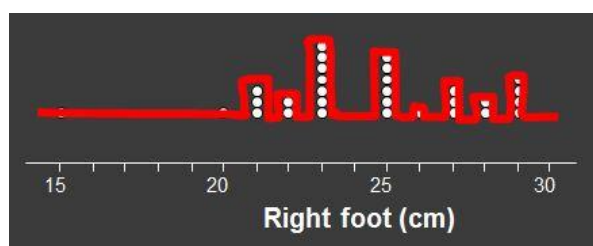


**Figure 8-1. What Sina sees**

*Sina:* It moves up, down, up, down, up, down [Figure 8-1]. It is like buildings in a city – if you outline the graphs it makes shapes of buildings.

*Commentator:* Sina is shape spotting, seeing pictures, like seeing pictures in clouds, a necessary precursor, according to Ainley, Nardi and Pratt's (2000) research, before she gets to see through the data.

*Interviewer:* Rita, what do you see?

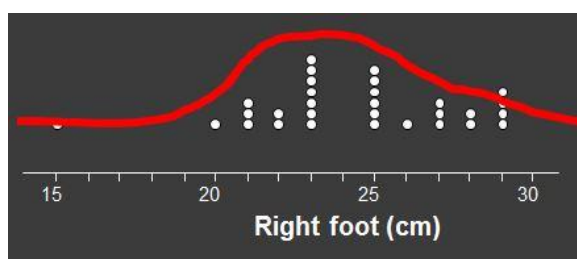


**Figure 8-2. What Rita sees**

*Rita: It's kinda shaped like a mountain [Figure 8-2]. It's like a mountain with the highest peak in the middle*

*Commentator: Rita is “seeing through the data”, seeing the “whole”; she is using her own language but she is seeing through. Compared to Sina she is on the next level, in fact a very important level in describing shape – she has gone beyond describing what she sees to inferring what she sees.*

*When we are just beginning to learn how to reason comparatively we have to keep the principle of statistical inference, the link between sample and population, to the forefront. We invoke two quite distinct kinds of thought processes: thoughts about what we see in the data and thoughts about what may be happening back in the populations. Through language we can clearly signal the difference between these two kinds of thoughts. We believe it is critical to stress this distinction with students at this level.<sup>1</sup>*

Before the 2009 teaching experiment, most students had no idea what shape or spread meant. There was limited progress on describing and inferring shape. Some were getting to the mountain as a descriptor but nothing beyond. The teaching programme in the 2009 teaching experiment had covered posing investigative questions, introducing students to box plots, doing complete investigations, and appreciating ideas of sample, population, sampling variability, and making a call. The teaching programme in the 2009 teaching experiment (first year of the TLRI) did not specifically address the issue of shape (Table 7-2, pages 150–151) – shape was just part of describing graphs.

In the second year of the TLRI (2010), some of the work focused on how to build students' ideas about shape and spread when they look at plots. Discussions were held with overseas experts (Bill Finzer, Cliff Konold and Sandy Madden) and a clear solution was not in existence. Following the discussion, some activities were developed for use in the second year of the TLRI and these seemed successful (Pfannkuch et al., 2011). When planning for teaching experiment 4, and making connections across the PPDAC cycle through the

---

<sup>1</sup> For the full presentation see: <http://www.censusatschool.org.nz/2009/informal-inference/>

investigative question, it was felt that the area of description of distributions was still very weak in terms of contextual and statistical knowledge and that the activities needed further development with an initial goal of developing the language of shape.

The problematic situation that was now identified was building students' capability to recognise and describe distributions. Distribution links to the investigative question. When students pose investigative questions they need to start to predict or hypothesise what the population distribution of the particular variable will look like. Students need to build the image of the population not just as the cards in the bag but as a visual picture of the distribution, i.e. the sketch of the shape of the distribution for the particular variable they are investigating. They need to think “whole population” and to predict or “suspect” the population distribution or shape of the variable of interest for the given population (Figure 8-3).

**Whenever I pose an investigative question...**

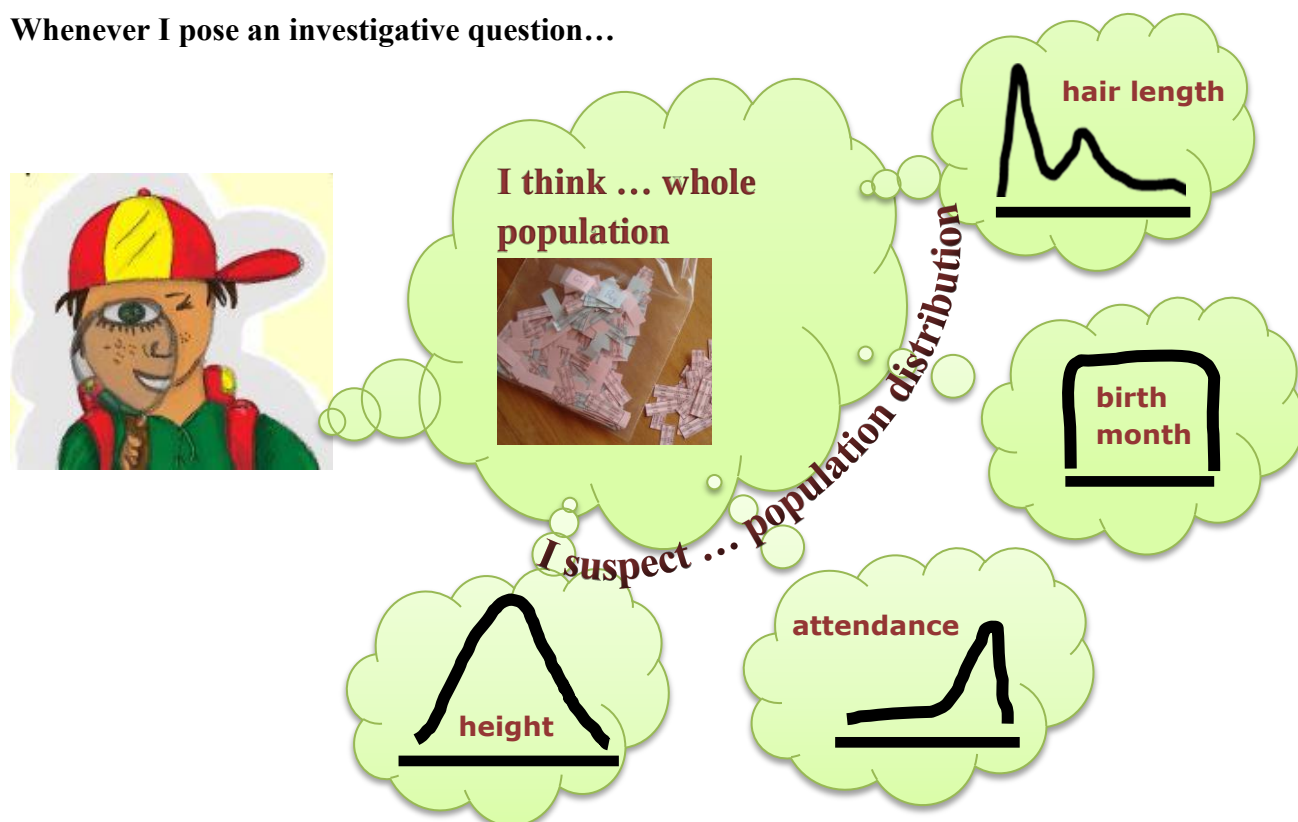


**Figure 8-3. Connection between posing an investigative question, the population data bag and the image of the distribution for a given variable**

In addition, students need to be able to describe the distribution of the data with sufficient detail that the description could be used to recreate a sketch of the graph of the data. Within

the description students need to be describing the overall shape of the data, and this in itself is a challenge as the shape is an inferred shape, like the “mountain” that Rita saw rather than the actual shape, or like the “city skyline” that Sina saw.

Put more simply, the problematic situation involves three research questions: (1) What descriptors do year 10 (ages 14–15) students intuitively use for distributional shape? (2) What makes a good distribution description at level 5 (ages 13–15) in the New Zealand curriculum? and (3) What distributional shapes and graphs do year 10 (ages 14–15) students predict when given the context?

### **8.3. Literature review**

Over the last decade there have been a number of in-depth research projects with a focus on distribution and students’ reasoning about distribution; for example, research has been done by teams at the Freudenthal Institute (Bakker, 2004a; Bakker & Gravemeijer, 2004) and Nashville (McClain, 2005; McClain & Cobb, 2001). And in 2005 the Fourth Statistical Reasoning, Thinking and Literacy (SRTL-4) research forum in Auckland had distribution as a focus with the forum theme being *Reasoning about Distribution*. From these and other literature, five themes have emerged: (1) the notion of distribution, (2) measures of centre, (3) shape of distributions, (4) predicting distributions, and (5) contextual knowledge.

#### **8.3.1. Notion of distribution**

Distribution is a multi-faceted notion involving centre, spread, skewness, shape and density (Bakker, 2004a; Ben-Zvi & Amir, 2005; Konold, Higgins, Russell, & Khalil, 2004; McClain, 2005; Pfannkuch, 2006; Reading & Reid, 2006). As they consider distribution, students will be considering all of the following together to get the “bigger picture”: measures of centre (mean and median); measures of spread (range, interquartile range and standard deviation); where the majority of data values are in relation to extreme values (skewness); and how density and skewness provide detail about shape. This simultaneous consideration of many aspects requires global reasoning, the coordination of these ideas that makes distribution a complex notion that students across the system find difficult (Ben-Zvi & Arcavi, 2001; delMas, Garfield, & Ooms, 2005; Hancock et al., 1992; McClain & Cobb, 2001; Pfannkuch, 2006).

### **8.3.2. Measures of centre**

How data are distributed and measures of centre are irrefutably linked. Symmetric distributions generally have the mean, median, mode and midrange all approximately in the same position, whereas for skewed distributions this is not the case. It is skewed distributions that lead to discussions about which “average” is more robust in the different situations (Bakker, 2004a). Medians are touted as more suited to asymmetric distributions as they are less influenced by extreme values and when linked with interquartile range, give a good sense of centre and spread in these situations. One generally cannot consider distribution without considering measures of centre.

### **8.3.3. Shape of distributions**

Describing shapes of distributions has had fleeting mention, with Bakker (2004a) in his dissertation providing the only real depth in work on shape. However, despite the relatively superficial exploration of shape, there are some starting points for consideration in this research. Firstly, the type of graph used to display the data has a major influence on students’ ability to perceive shape. For example, box plots and even histograms at earlier ages can prove a problem for students to use as they are too abstract and the actual data has disappeared (Bakker, 2004a; Friel et al., 2001). In fact, box plots offer minimal if any information on the shape of a distribution. Dot plots, on the other hand, provide an initial starting point for students to explore shape, along with simple case-value bar graphs (Bakker, 2004a; delMas et al., 2005). Pfannkuch (2006) suggests that dot plots and stem-and-leaf plots can provide a strong basis for interpreting and understanding distributions and students can transition from here to box plots. Most critically, displays used should allow sense to be made of the information with as much ease as is possible (Friel et al., 2001).

Bakker (2004a) suggests that single univariate distributions are a good starting point, but sounds a word of warning that students can initially assume that all distributions are symmetric if only this type of distribution is chosen. This initial misunderstanding can be challenged by deliberately choosing to use distributions that are skewed as well as symmetric ones (Bakker, 2004a; delMas et al., 2005; Makar & Confrey, 2005; Rubin, Hammerman, Puttick, & Campbell, 2005). Linked is the need to provide multiple opportunities for students to recognise and understand the direction of a skew (delMas et al., 2005) as even college-level students mix up the direction of the skew.

Descriptors of shape include uniform, normal, and skewed to the right or left (Bakker, 2004a), and normal, skewed, bimodal or uniform (delMas et al., 2005), while students new to statistics describe data in terms of low, average and high values, and name shapes using pyramid, semi-circle and bell shaped (Bakker, 2004a). Shape helps to develop meaning of mean, spread, density and skewness (Bakker, 2004a; Rubin et al., 2005) and connections between measures of centre and shape can be made (Konold & Higgins, 2003; Rubin et al., 2005). However, students can use the words in unconventional ways and need guidance, and Bakker (2004a) found that using too small a sample size was a problem when trying to identify shape of a distribution, as was unsuitable scaling and a lack of context.

### **8.3.4. Predicting distributions**

As researchers worked with students towards getting them to sketch the continuous “population” shape of a distribution, the researchers found that asking “What if ..?” questions or getting students to predict what a related but similar distribution might look like helped students to progress in their understanding (Bakker, 2004a, 2004b; Bakker & Gravemeijer, 2004; Ben-Zvi & Amir, 2005). The work on making predictions was also pre-empted by fostering students’ thinking towards an aggregate view of the data. For example, Bakker (2004a) in his dissertation work asked grade eight students to predict what the weights of all eighth graders in the city might look like, and Ben-Zvi and Amir (2005) asked second graders to predict how many milk teeth third graders would lose.

For students to be able to make predictions about the distribution of data for a particular context, the context that they are predicting about needs to be familiar so that the students can use personal knowledge and experience to make a prediction (Bakker, 2004a, 2004b; Ben-Zvi & Amir, 2005). The two examples cited above fit this criterion nicely. delMas et al. (2005) found that college students could match graphs to descriptions in contexts such as the set of scores for an easy quiz, but had difficulty with the distribution of a set of random digits. This may have been due to a misconception on the meaning of random, thinking it meant normal. Watson (2005) also cites examples of students making graphs in a variety of situations and shows differences between those situations where some data is provided and those where there is no data provided, and noted “the importance of choosing contexts where students have some intuition about the variation present” (p. 219).

### 8.3.5. Contextual knowledge

Distribution is most often realised in a display such as a graph. Friel et al. (2001) recognise the complexity of reading graphs and what the graph reader must attend to; for example, describing, representing, analysing and interpreting data as well as taking the context into account. Questions they ask such as “How does the learner’s understanding of the context contribute to his or her interpretation of data represented in a graph?” and “Can one interpret data accurately without having a significant level of understanding of the context?” (Friel et al., 2001, p. 152) are challenges to think deeply about when considering the role context plays in students’ understanding of distribution. Bakker (2004a) and Watson (2005) also note the importance of context both in recognising shape and in understanding how much variation from an expected distribution is realistic. Contextual knowledge is key when students are predicting distributions of data values for a given situation, as with no knowledge of the context it is extremely difficult to predict the distribution, and contextual knowledge is also key when looking at distributions and deciding if they are realistic (Ben-Zvi & Amir, 2005; Konold & Kazak, 2008; Watson, 2006).

A related matter is the inclusion of the context in the description of the distribution. Without the contextual references, the description becomes meaningless. The complete description of a distribution is “one that involves, shape, center, and spread, and in the context of the data” (delMas et al., 2005, p. 6). These features in the description are properties of the aggregate or collection as a whole and not of individual data values (Konold et al., 2004).

An end goal is that students are able to describe sample distributions as part of their using the statistical enquiry cycle to answer an investigative question about a population. The following three quotes encapsulate big ideas around distributions and why being able to understand the component parts and describe them in context is important.

The power of statistical data analysis lies in describing and predicting aggregate features of data sets that cannot be noted from individual cases. (Bakker, 2004a, p. 100)

A larger objective in terms of the goal of statistical literacy when students leave school is to be able to tell a story from a context with a graph that displays variation, clustering, middles and surprises. (Watson, 2005, p. 189)

What is relevant is building intuitions that sound warning bells if distributions look unusual or too many outliers appear. (Watson, 2006, p. 188)



### 8.3.6. Existing theoretical frameworks with links to distribution

Bakker (2004a) and Bakker and Gravemeijer (2004) proposed a structure (Figure 8-4) for analysing the relationship between data and distribution. Their proposed structure can be read both upward from a data (individual values) perspective or downward from a distribution (conceptual entity) perspective. They said that students as novices typically see individual values and use these to find values such as the median, range or quartiles, but that this does not mean that students are seeing the median, for example, as representative of a group.

Students need to develop a downward perspective as well: conceiving centre, spread, density and skewness as characteristics of a distribution, and looking at data with a notion of distribution as an organising structure or conceptual entity. (Bakker & Gravemeijer, 2004, p. 149)

| <b>distribution</b><br>(conceptual entity, pattern in variability; represented in diagram) |                                                        |                                              |                           |
|--------------------------------------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------|---------------------------|
| <b>centre</b>                                                                              | <b>spread</b>                                          | <b>density</b>                               | <b>skewness</b>           |
| mean, median, midrange,<br>...                                                             | range, standard deviation,<br>interquartile range, ... | (relative) frequency,<br>majority, quartiles | position majority of data |
| <b>data</b><br>(individual values; represented in a table)                                 |                                                        |                                              |                           |

**Figure 8-4. Between data and distribution**

*Note:* Adapted and reprinted from “Design research in statistics education: On symbolizing and computer tools,” by A. Bakker, 2004, p. 100. Copyright 2004 by A. Bakker. Reprinted with permission.

These characteristics of distribution – centre, spread, density and skewness – with their related features are all aspects of distribution, a multi-faceted notion.

Ben-Zvi, Gil and Apel’s (2007, p. 3) suggested theoretical framework for informal inferential reasoning (see Figure 8-5, next page) has cognitive aspects related to reasoning about chance and data and also socio-cultural aspects related to classroom and individual practices as well as discourse and dispositions. The framework is designed to form an emerging vision of informal inferential reasoning, and distribution is part of that picture. Distribution was also identified as part of the framework for thinking about statistical inference and sampling reasoning (see Figure 7-2).

| Informal inferential reasoning (IIR)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Cognitive aspects                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Socio-cultural aspects                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| <ul style="list-style-type: none"> <li>• <u>Reasoning about Variability</u><br/>Spread, density, variability from a variety of sources, ...</li> <li>• <u>Distributional Reasoning</u><br/>Aggregate views, pattern and trend, hypothesis and prediction, as well as local reasoning about individual cases, outliers, ...</li> <li>• <u>Reasoning about Signal and Noise</u><br/>Centre, measures, modal clumps, summary, ...</li> <li>• <u>Sampling Reasoning</u><br/>Sample size, randomness, sampling variability and behaviour, bias, representativeness, ...</li> <li>• <u>Contextual Reasoning</u><br/>Interpretation, alternative explanations, ...</li> <li>• <u>Graph Comprehension</u><br/>Creating and decoding visual shapes, ...</li> <li>• <u>Reasoning about Comparing Groups</u><br/>Comparison of centre, spread and shape, ...</li> <li>• <u>Probabilistic Reasoning</u><br/>Uncertainty, random events, chance, ...</li> <li>• <u>Inferential Reasoning</u><br/>Generalisations, limitations and strength of conclusions, ...</li> </ul> | <ul style="list-style-type: none"> <li>• <u>Instructional Context</u><br/>Learning environment design, teachers and students' awareness of purposes and utility, ...</li> <li>• <u>Language</u><br/>Discourse types and norms to discuss data, graphs, sampling, inferences, ...</li> <li>• <u>Culture and History</u><br/>Students' beliefs, dispositions, prior knowledge and background, ...</li> <li>• <u>Argumentation</u><br/>Arguing about inferences, claims and counterclaims, data-based evidence, ...</li> <li>• <u>Socio-Statistical Norms</u><br/>Classroom discourse norms, what a statistical claim is, what it takes to be convinced that claim is true or false, ...</li> <li>• <u>Evaluative Disposition</u><br/>Providing and assessing evidence, level of confidence, critical disposition to sampling and inference, ...</li> <li>• <u>Flexibility</u><br/>Transfer back and forth between local and global view of data, sample and population, data and context, reality and its representations, ...</li> </ul> |

**Figure 8-5. Suggested theoretical framework of IIR**

*Note:* Reprinted from “What is hidden beyond the data? Helping young students to reason and argue about some wider universe,” by D. Ben-Zvi, E. Gil and N. Apel, 2007, Paper presented at the Fifth International Research Forum on Statistical Reasoning, p. 3. Copyright 2007 by D. Ben-Zvi, E. Gil and N. Apel. Reprinted with permission.

In particular, reasoning about variability (spread, density), distributional reasoning (aggregate views, pattern and trend, hypothesis and prediction, local reasoning about individual cases, outliers), reasoning about signal and noise (centre, measures, modal clumps, summary), contextual reasoning (interpretation, alternative explanations), and graph comprehension (decoding visual shapes) would be considered as pertinent to distribution.

### 8.3.7. Visual imagery

Since the design of the learning experiences used in this research combine visual imagery, analysis, and verbalisation of what is seen and inferred, the theoretical framework of mental processes (Figure 8-6, next page) proposed by Aspinwall, Hacıomeroglu, and Presmeg (2008)

to describe successful calculus students' thinking may be useful for exploring concept development in this teaching experiment. Aspinwall et al. believe that a combination of visual, analytic, and verbal descriptive thinking sustains and supports students' mental processing. The first component of their framework, visual thinking, "includes processes of creating or changing visual mental images, a characterisation that includes the construction and interpretation of graphs" (Aspinwall et al., 2008, p. 98). Presmeg (2006) found that teaching that promotes visual thinking is characterised by aspects such as use of gesture, facilitating students' construction and use of imagery, use of colour, encouraging students to seek patterns, delayed use of symbolism, and deliberate use of cognitive dissonance. This type of learning is very similar to the learning experiences and learning theories that were drawn upon in this research (Arnold et al., 2011).

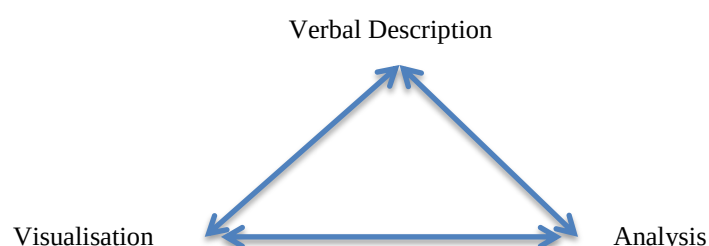


**Figure 8-6. The Mental Processes Framework**

*Note:* Reprinted from "Students' verbal descriptions that support visual and analytic thinking in calculus," by L. Aspinwall, El Haciomeroglu and N. Presmeg, 2008, Paper presented at the Joint Meeting of PME 32 and PME-NA 30, p. 100. Copyright 2008 by L. Aspinwall, E. Haciomeroglu and N. Presmeg. Reprinted with permission.

Dynamic visual imagery is associated with high levels of functioning in mathematics but needs to be supported by analytical thinking (Haciomeroglu, Aspinwall, & Presmeg, 2010). Imagery must be coupled with analytical thought processes, the second component of the framework, to be used effectively in mathematics. Logical reasoning, mathematical symbols and representations are part of the analytic process. For example, analytic thought processes do not necessarily mean translating a graph to an equation, then calculating the derivative; it can mean recognising a graph as a cubic function and that its derivative will result in a parabola (Aspinwall et al., 2008).

The third component of the framework, verbal description, is considered to be critical in supporting visual and analytical thinking. This internal talk seems to help with the interpretation of images and consequently helps in the translation between visual and analytical processes. Aspinwall et al. (2008) describe verbal-descriptive thinking as "the

linchpin sustaining the use of visual and analytical thinking” (p. 97). They believe their framework may be useful for instruction in mathematics, particularly for conceptual learning. The framework may also provide support for any progress that might be observed in the development of students’ reasoning in the fourth teaching experiment.

## **8.4. Planning and preparation**

### **8.4.1. Planning and preparation for teaching experiment 4 (2011)**

This section discusses the planning and preparation for the fourth teaching experiment with the lens placed on describing shape, predicting distributions, and building a contextual library with a view to describing distributions of data to support answering the investigative question. The learning trajectory designed was to specifically support students developing an idea of what is meant by: (1) shape in relationship to distributions, (2) sketching and describing the shape of a distribution, (3) predicting the shape of distributions, (4) identifying and describing features of data distributions, and (5) connecting all of these to the context, i.e. to both the population(s) and the variable(s). While some researchers, such as Bakker (2004a, 2004b), have reported on students’ responses regarding describing shape, no one has explored how to develop students’ language of shape for describing distributions and what describing data distributions encapsulates, in particular for 14- and 15-year old students.

#### ***Adaptations to teaching and learning activities***

The teaching and learning activities were designed to support students’ understanding of the language of shape used in description and making descriptions of distributions. These activities built on the work previously done in the TLRI project and through discussion with overseas experts. The activities also included additional thinking by the research team as the bigger picture of what was trying to be achieved was considered. The new and updated activities were based on key premises from the themes that emerged from the literature. Specifically, the activities focused on the notion of distribution, the language of shape of distributions, predicting distributions, and building contextual knowledge particularly related to the types of variables that tend to have symmetric, skewed or uniform distributions. Unpacking students’ existing contextual knowledge and misunderstandings were key ingredients in predicting distributions.

## Chapter 8 – Answering the Investigative Question: describing distributions

This chapter focuses on lessons 2–4 of the 16-lesson unit on statistics. Lesson 1 was a review of the PPDAC cycle and making dot plots using data cards, while lessons 2–4, the focus of this chapter, are summarised in Figure 8-7.

| Lesson number | Summary of lesson activity                                                                                                                                            |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2             | Developing the language for shape of distribution descriptors, sketching shapes from graphs, grouping similarly shaped graphs, matching language to groups of graphs. |
| 3*            | Making predictions of the graph from contexts, matching contexts to graphs, starting to develop a “library” of similarly shaped graphs.                               |
| 4*            | Sorting more graphs according to shape of distribution, starting to describe distributions.                                                                           |

**Figure 8-7. Lessons 2–4 in teaching experiment 4**

*Note:* \* means that the lesson was taught by the researcher as the teacher had laryngitis.

The material for the three lessons is detailed in Appendix H and was developed following the identification of areas that needed further attention. These areas were based on previous research findings and experiences and supported by the literature.

Wild and Pfannkuch’s (1999) paper on statistical thinking in empirical enquiry has in its introduction the idea that “statistical thinking is the touchstone at the core of the statistician’s art” and that “it has been much more a product of experience, war stories and intuition than it is of any formal instruction that we have been through” (p. 223). There are many aspects to statistical thinking, and Wild and Pfannkuch explore four dimensions in their paper, as were outlined in the literature review (see chapter 2, section 2.3.2, page 12). One aspect of the bigger picture is contextual knowledge that statisticians bring to statistical enquiry. Right from an inkling of an idea (see Figure 2-4, page 21) through to drawing conclusions, contextual knowledge plays a big role. For students to gain this experience their contextual knowledge base needs to be deliberately built, and one aspect of this is to develop a “library” of contexts (variable and population) that have similarly shaped distributions. In this fourth teaching experiment a library of contexts was built over the unit of work, starting with the contexts that were used in lesson 2 (Figure 8-7).

The three lessons, lessons 2–4 in the 16-lesson teaching unit, are described next to demonstrate how student-generated concepts, ideas and language were gradually transformed towards a statistical approach.

In lesson 2 the students sketched the shape of 15 data distributions. The data distributions were of very large samples. The idea of using very large samples was in the hope that the “shape” would be more obvious (compare this with “growing samples” (Bakker, 2004a)) and that students would attend to the signal more, developing a picture of the whole. The data distributions were shown briefly using a PowerPoint presentation. An added reason for showing the graphs briefly was to help focus the attention on the overall “smooth” shape or signal of the distribution, with the view to move students away from the “city skyline” representation where they attended to the noise. “The signal is the continuous shape with which they model the data, and the noise is the variation around the signal” (Bakker, 2004a, p. 234). The students then grouped the sketches of the graphs into similar shapes and used their own language to describe the shapes in each group. The teacher (T2, 2011) collated their responses and used these to come to a consensus regarding the different patterns in the shapes of the graphs. At this stage the statistical names were introduced and the students matched these to the shapes and names they had been using.

The aim of lesson 3 was to get students to think about how context and shape were linked. They were given 15 contexts without the graphs and in groups asked to sketch (predict) the graph for three of the contexts with possible values included for the particular variable and population. A class discussion justifying shapes for each of the 15 contexts followed. Students were then given the actual dot plots of the contexts and they matched these plots to the context. The 15 contexts were the same as the 15 graphs used in the second lesson. The graphs were sorted again into groups and each group was labelled using appropriate statistical terminology. From here the students started to build their context library, collating the graphs into similar shapes, gluing them into their books, and leaving room to add new contexts and graphs throughout the unit.

In the first activity of lesson 4 the students classified some more data graphs by shape and added these to their growing library of shapes and contexts (shown under “other examples” in Figure 8-10, page 207). The second activity involved starting to describe distributions.

Throughout the research a common theme was that teachers needed to model more – model posing investigative questions, model making the call, model writing descriptions. In this teaching experiment a specific focus was made on *active reflection* along with the modelling. The idea of active reflection was suggested to the teacher previously (teaching experiment 3, 2009) around posing investigative questions but the teacher had not used it very often. For

this teaching experiment the researcher felt that this was possibly a key component that was missing. The process of active reflection involves the teacher working with student-generated work and through questioning, prompting and pointing out errors, the student work is reflected on and corrected. This process is similar to what Franke et al. (2009) described where teachers take mathematical expressions and workings that are incomplete and/or incorrect and makes them complete and correct. Teachers can model all of the things mentioned above, but they also need to model the process for correcting statements made. The process of active reflection helps the students to “see” what the teacher is thinking, not just to witness the end result, the correct model. An example of this is given in section 8.5.2, Figure 8-12 (page 211).

In addition to the active reflection of student-generated work, the teacher was asked to try to always model using full statements rather than using bullet points. It was felt that unless the teacher only modelled good practice the students would perceive that bullet points were appropriate to use, when that was not the case.

### ***Adaptations to pre- and post-tests***

In the 2008 and 2009 pre- and post-tests the students had to “complete” an investigation including describing shape. For 2011 this aspect was dropped to make the focus tighter on the specific concepts that were the main focus of this thesis. The three aspects of the pre- and post-tests in 2011 (see Appendix B.7 and B.8) were: (1) posing investigative questions (tasks A & B); (2) shape, describing distributions and predicting distributions (tasks C, D & E); and (3) making the call (task F).

Three new tasks were developed to address the second aspect, i.e. shape, describing distributions and predicting distributions. The first task (task C) required students to sketch the shape of a distribution and then to describe the distribution for three different situations; the second task (task D) required students to match given statistical terms for shape to each of eight graphs; and the third task (task E) required students to sketch the predicted distribution for three different scenarios (similar to Bakker (2004a, p. 156), getting students to think about “If the context were to change, how would the graph change?”). The variable was the same for each scenario (age in years) in task E, just the population differed – year 10 class, CensusAtSchool participants (years 5–13), and New Zealand Association of Mathematics Teachers (NZAMT) conference participants. These questions were the same for both the pre- and post-test.

### ***Adaptations to pre- and post-interviews***

The 2011 interview was adjusted from the one used in previous years. Instead of asking students about their responses to tasks C and E, the students were given new situations to respond to in the interview. The task C equivalent required the students to sketch the shape of the distribution for three new contexts and to then describe the distribution; the task E equivalent asked them to sketch the distribution for a summary situation and then to sketch the distributions for a comparison situation and to describe what they had done and why. With the exception of the comparison situation, the questions remained the same from the pre- to post-interview; for the comparison situation the populations changed but the variable remained the same.

These changes were made to the interview schedule to try to capture student thinking as the students were doing the task rather than getting them to recall what they were thinking when they had sat the pre- or post-test. This gave immediacy to the situation with the aim to highlight good statistical thinking and reasoning and also student misconceptions.

## **8.5. Teaching experiments**

### **8.5.1. Teaching experiment 4**

This section outlines the key learning outcomes from the three lessons (lessons 2–4) where describing distributions was the main theme, reflecting on concepts and ideas the students generated, deliberate teaching acts and “aha” moments.

The two main outcomes from the three lessons were: (1) patterns in the shapes of distributions, and (2) developing the notion of distribution – describing distributions.

### ***Patterns in the shapes of distributions***

Lesson 2 started with the students sketching graphs from given distributions. After they had sketched their graphs, the teacher asked the students to sort the graphs into shapes that are similar and then, after allowing them time to do this, asked about the number of groups they had made. The responses ranged from three to five. The teacher sought feedback from a couple of the groups who had responded with four as this was the majority response. An example of a response from a class group to the question about how many groups they had split the graphs into:



## Chapter 8 – Answering the Investigative Question: describing distributions

*Teacher: Four groups. What were they based on?*

*Student: Sloped to the left, and sloped to the right, symmetric ones.*

*Teacher: So you have sloped to the left, sloped to the right, symmetric and what was your other group?*

*Student: You know [gestures with hand – up, across and down] ... it is even on the top.*

*Teacher: Even on the top. So let's see, symmetrical, some sloped to the left, slope to the right, other one was ...*

[Various student responses with “flat top” being the loudest.]

The teacher used these four group headings – symmetrical, sloped to the left, sloped to the right, and flat top – as a starting point. The class then sorted the graphs into one of the four groups (Figure 8-8). Finally the students were introduced to the statistical language used to describe shapes and were asked to match these words to their graphs. Intuitively the students re-grouped the graphs according to symmetry: symmetric or not symmetric, splitting the skewed (not symmetric) into two groups (left and right) and the symmetric into two groups (uniform and other).

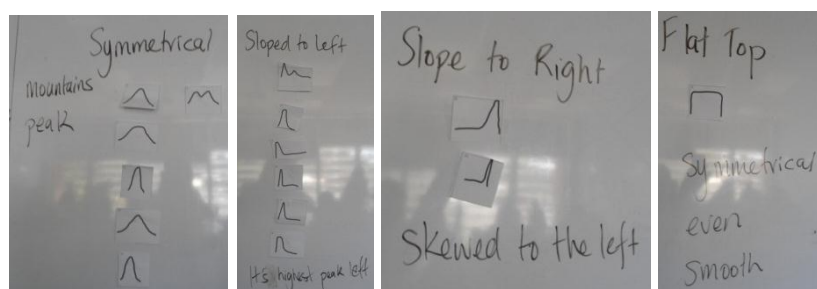


**Figure 8-8. Four shape groups with graphs and additional ideas**

Interestingly the students did not use modality for grouping, yet in the planning and preparation for the lesson bimodality was one of the suggested shapes, as is shown in the “solutions” (Figure 8-9) which was in the supporting notes for the teachers.

| <i>Description of shape</i> |                      |                               |
|-----------------------------|----------------------|-------------------------------|
| Symmetric – unimodal        | Uniform              | Bimodal                       |
| right skew                  | Symmetric – unimodal | left skew                     |
| bimodal – right skew        | right skew           | Symmetric – unimodal          |
| Symmetric – unimodal        | left skew            | bimodal – symmetric each part |
| right skew                  | Symmetric – unimodal | right skew                    |

**Figure 8-9. “Solutions” to description of shape from teacher’s activity sheet**

When the graphs were sorted again in lesson 3, following the matching of the graph to the context, the students again sorted them into the four groups (symmetric, sloped to the left, sloped to the right, and flat top) that had been identified in the previous lesson. Each group was labelled using appropriate statistical terminology – symmetrical, right skew, left skew and uniform (compare this with Bakker's (2004a, p. 235) terminology: uniform, normal, left-skewed, and right-skewed). A discussion was had around why and which way the skew was recognised. At this point the distinction between unimodal and bimodal was also made.

*Researcher/Teacher: These are the graphs yesterday that you said were symmetric, and I've moved this one out to the bottom. Why do you think I have done this?*

*Student: Because it's bimodal.*

*Researcher/Teacher: Because it's bimodal. So these are symmetric, and unimodal, which means that they have one bump or one peak. So they have one mode, or peak, and this one here is symmetric and bimodal because it has two peaks.*

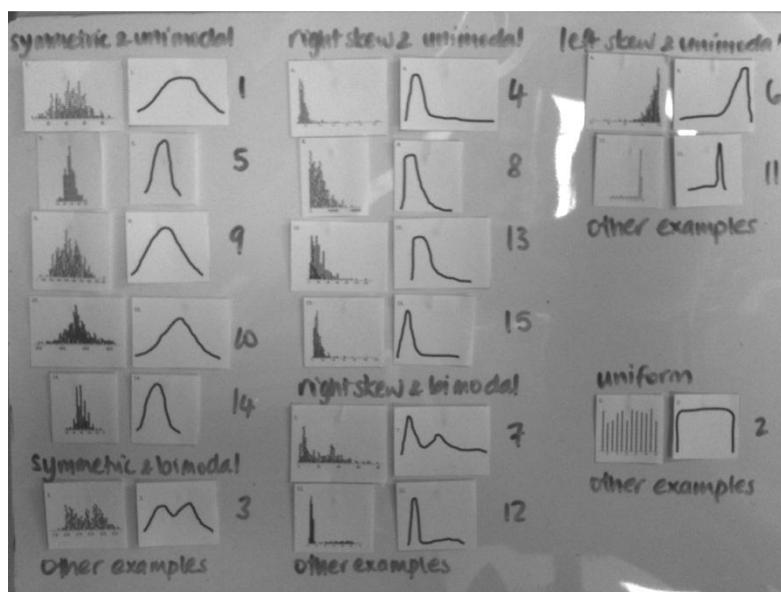


**Figure 8-10. Final collation of shapes into four groups with modality distinction**

From this brief conversation the way to sort the shapes became clear – sort by symmetry and then by modality (Figure 8-10). The shape descriptors developed from the way the students intuitively sorted the graphs. The students did not make a separate group for bimodal, as the research team did. They sorted into four groups and then split two of the groups by modality.

### ***Describing distributions***

In lesson 4 the researcher/teacher facilitated a class discussion on key features for describing graphs. The students were given the challenge that if they had to draw the graph from the description, what information would they need. Shape was a given in the discussion, and the students came up with range, peak(s) and averages and, with prompting, variable, population and units. In the following transcript of the discussion, the feature is in **bold** when it is first mentioned.

*Researcher/Teacher: Obviously one of the things that we think is pretty important about describing graphs is describing the **shape**. ... If I was going to describe this graph, what other things might I want to describe about it?*

*Student: The **range**.*

*Researcher/Teacher: I might want to describe the range. That would tell me the values it goes from and to. What other things would be important?*

*Student: The height.*

*Researcher/Teacher: The height. So what do you mean by the height?*

*Student: Its highest point.*

*Researcher/Teacher: What are we calling that highest point?*

*Student: The **peak**.*

*Researcher/Teacher: The peak. So we might want to talk about what the peak is. What else might we want to talk about?*

*Student: The **average**.*

*Researcher/Teacher: The average, or the middle. Look at whereabouts the middle is. What else might we want to talk about? What makes that graph different to, say, number 14?*

*Student: The **amount of peaks**.*

*Researcher/Teacher: The amount of peaks. We might want to talk about the peak and the number.*

*Student: The amounts.*

*Researcher/Teacher: But how do I know what these two graphs are about? What is number 9 about?*

*Student: Height.*

*Researcher/Teacher: And what is number 14 about? [see Figure 8-10]*

*Student: Index finger.*

*Researcher/Teacher: Index finger. Do you think that is important to put in the description?*

*Student: Yes.*

*Researcher/Teacher: What are they called? Height and index finger, what do we call those?*

*Student: Context.*

*Researcher/Teacher: They are part of context.*

*Student: **Variable.***

*Researcher/Teacher: The context is the variable and the **group** that it's about. One of the things that we need to talk about is the variable. ... The variable we are talking about here is height. What's it in?*

*Student: Centimetres.*

In the discussion the actual features that the students suggested included: target population, variable, units, general shape sketched, overall shape, modality, peaks, range, median and mode.

Later in the same lesson there was a discussion focused on describing the right skew graph of reaction times (see Figure 8-11a, next page). The following additional features surfaced: clustering density, majority, modal group, and describing shape in terms of parts of the whole. The transcript below related to this discussion also highlights how the teacher used active reflection to move from student language to statistical terms, modelling good practice. The final description for the reaction time graph is shown in Figure 8-11b (next page). The conversation starts part way through the description, following the first two statements:

*Researcher/Teacher: Okay, so that gives me kind of an overall picture, but it doesn't really tell me about where the bulk of the data is does it?*

*Student: It **bunches** at zero point five and it drastically drops to ...*

*Researcher/Teacher: Let's work on that bit. The reaction times. You said they're bunched up at 0.5. So I'm just going to put it into better words. The reaction times are tightly bunched or **tightly grouped** around 0.5 seconds. Okay and I would say that they are approximately symmetrical from 0.2 to about 0.7 seconds. If I just look at this bit of the graph here, [indicates using hand gestures from 0.2–0.7 on the graph] that's all I'm talking about is approximately symmetrical. So they are tightly grouped around that 0.5 seconds and they are **approximately symmetrical in that small range**. That helps me quite a bit, doesn't it, in defining.*

[pause]

Where do they peak?

Student: 0.3, 0.4.

Researcher/Teacher: The peak is here. So it's about 0.38.

[pause]

And we could say that **nearly all the reaction times are between 0.2 and 0.7 seconds with a tail to 3 seconds**. Something like that. Okay. Some graphs you'll write a lot more about than others. It just depends on what's in them.

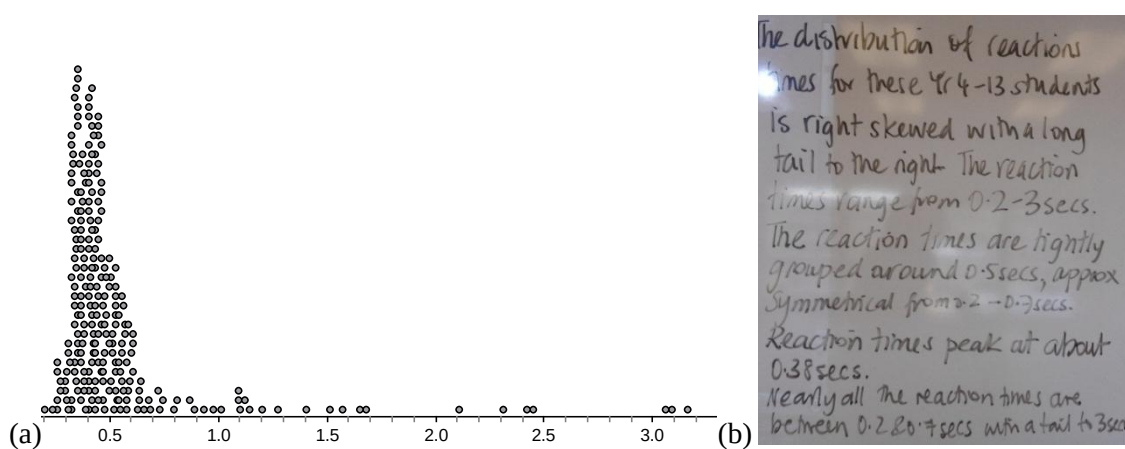


Figure 8-11. (a) Reaction times graph, (b) Final description for reaction times graph

Initially the plan was to do a couple of descriptions together as a class and then to give the students most of the descriptions leaving only a few for them to do themselves. In the end it was decided that giving the students the descriptions was not the best strategy. Instead they did one each night for homework and at the beginning of the next lesson (and every lesson) a student wrote their description up on the board and the teacher with the class used active reflection to make the description correct and complete. This process was modelled at the start of every lesson. Figure 8-12 (next page) shows two descriptions students have written that the teacher has then used active reflection with the class to improve and tidy up the language and to complete the description as necessary.

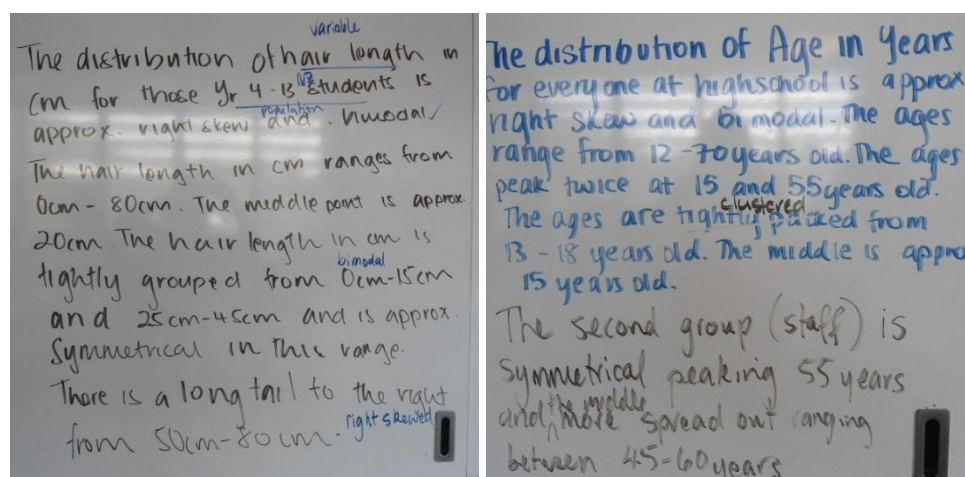


Figure 8-12. Two examples of student descriptions that have been “actively reflected” on

### ***Students using words in unconventional ways***

Bakker (2004a, p. 225) refers to students using words in “rather unconventional ways” in relationship to terms such as centre, spread and density. In this research students were found to be using words in unconventional ways as well. Initially the students used phrases such as “approximately right skewed” (Figure 8-12, second example) and “slightly symmetrical” to describe shape. They were starting to use the language that the teacher had been using but had not yet realised the subtleties of that language; i.e. the teacher had referred to graphs as being “approximately symmetrical” and “slightly right skewed”. The teacher was very reflective and picked up on these unconventional phrases and explained the more common usage. This unconventional usage did persist with some students; for example, “It tends to be unimodal and slightly symmetrical” (2011 student, post-test).

### ***Building a contextual shape library***

Once students had sorted the initial graphs from lesson 2, they glued the graphs into their books in the form of a reference library. As additional activities were completed they added graphs and contexts to each of the different shapes. Figure 8-13 (next page) is an example from one student’s book and shows symmetrical and unimodal contexts.

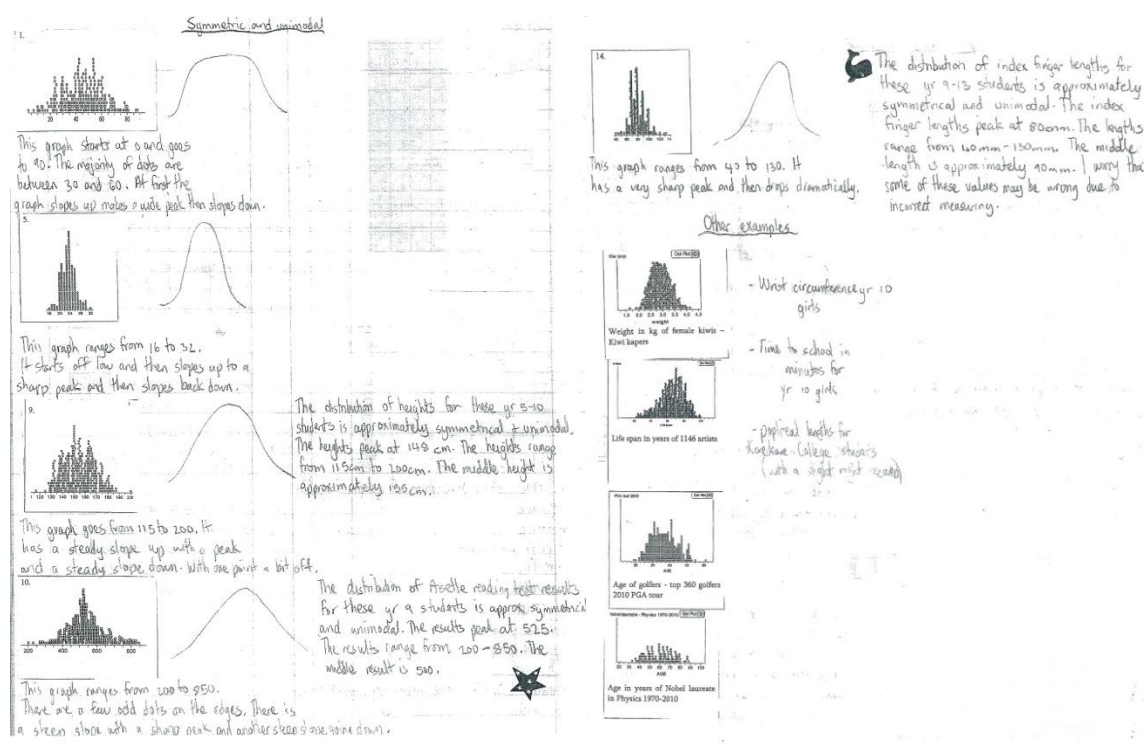


Figure 8-13. Example of symmetrical contexts from 2011 student class exercise book

Note that as well as the graph and the sketch of the shape of the graph, below each of the first five graphs is a first attempt at describing the graph and to the right is their updated descriptions (where the descriptions have been done). In addition, the student has added in the graphs and contexts from the matching activity, the contexts from the first lesson, and one of the contexts from the work with Karekare College students. This process is repeated for the other shapes.

## 8.6. Retrospective Analysis

Two of the three research questions for this chapter are explored in this retrospective analysis section: (1) What descriptors do year 10 (ages 14–15) students intuitively use for distributional shape? and (2) What makes a good distribution description at level 5 (ages 13–15) in the New Zealand curriculum. These two questions along with the third question, (3) What distributional shapes and graphs do year 10 (ages 14–15) students predict when given the context?, are also discussed in section 8.7 Quantitative analysis of test questions.

The main sources of data were student pre- and post-test responses and student pre- and post-interviews. Field notes and student exercise books have been used to provide additional data.

### 8.6.1. Shape descriptors

The students intuitively sorted the graphs into four groups. These four groups were initially described as symmetrical, sloped to the left, sloped to the right, and flat top (see section 8.5.1). In preparation for the teaching and learning and from the literature, a potential fifth category, bimodal, was identified, but this was not used as a category for sorting by the students. Bakker (2004a) had named the four types of distributions as “uniform, normal, left-skewed, and right-skewed” (p. 235). It is not clear if these were realised from student findings or given by the research team. What is interesting is that these four distributions reflect where the students in this research ultimately got to, but do not quite give the bigger picture of shape and how one might go about classifying the shape of the distribution. An important aspect that is missing from these descriptors is modality, which this research has shown is as important as the “type of distribution” in describing the shape of distribution. In addition, these four descriptors do not recognise other symmetric graphs that are not uniform or normal – uniform and normal being two special cases of symmetric graphs.

When it comes to describing the shape of a distribution, the first aspect that is attended to is the symmetry of the graph: the graph is either symmetric or asymmetric. If the graph is symmetric it can be uniform, approximately bell shaped (normal), or some other symmetric shape; if the graph is asymmetric then it either has a tail to the right (right skew) or a tail to the left (left skew). By attending to symmetry the four types of distribution given by Bakker (2004a) are attended to, but there is a fifth type which is the symmetric graph that is neither uniform or normal. This is the first part of the shape descriptor, and with this a clear picture of the distribution is starting to form in students’ heads. If that was all of the description that was given, one would assume that the distribution is also unimodal, i.e. it has only one peak. This is the case with many contexts but there are sufficient contexts that are at least bimodal that the modality should be considered a key aspect of the shape descriptor. Once the idea of modality was established, the students automatically used both aspects (symmetry and modality) to describe the shape of the different distributions.

Therefore, the finding of this research is that a descriptor of the shape of a distribution should include both the symmetry and the modality of the distribution (with the exception of uniform and bell-shaped (normal) distributions, which assume the modality (zero and one, respectively) within their description). Figure 8-14 (next page) shows the progression from the graph of the distribution to the shape descriptor.



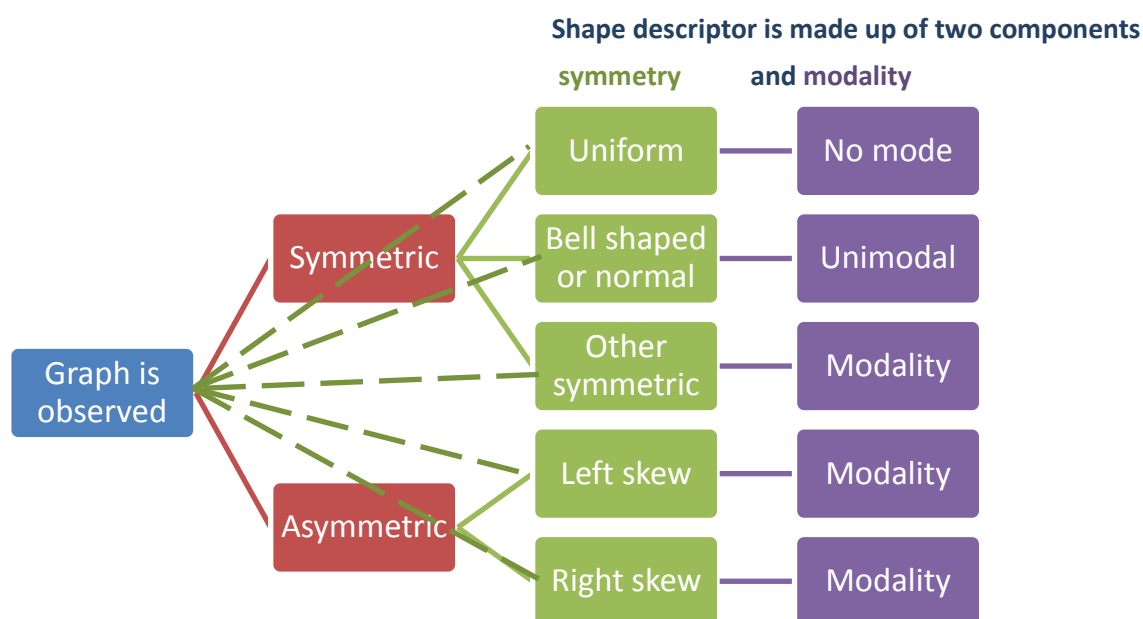


**Figure 8-14. Components of shape descriptors**

The split by symmetric and asymmetric is only obvious in the formal classification. In practice one would go straight to the symmetry descriptor and combine this with the modality (except for uniform and perhaps bell-shaped (normal) distributions). This usual practice is indicated by the dashed green lines. In the post-test the vast majority of students used this structure of symmetry and modality when describing the shape of the various distributions. Bell shaped or normal was not used, mostly as this had yet to come up as a particular case. An example from each of the situations in task C is given below:

*The distribution of ABs scores is right skewed and unimodal (2011 student, post-test).*

*The distribution of NZ Yr5–10 students' heights is approx symmetrical and unimodal (2011 student, post-test).*

*The distribution for heights of Tokoeka kiwis is approx symmetrical and bimodal (2011 student, post-test).*

There were some variations on the theme:

- (1) *The distribution of heights in cm for Tokoeka Kiwis is approx symmetric with a slight right skew and unimodal (2011 student, post-test).*
- (2) *The distribution of heights of Tokoeka kiwi is bimodal (2011 student, post-test).*
- (3) *The distribution of heights for NZ Yr5–10 students is approx symmetric (2011 student, post-test).*
- (4) *The distribution of NZ Yr5–10 students heights is unimodal and symmetric (2011 student, post-test).*

The variations included: seeing symmetry and skewness within one graph (1); only one aspect of the descriptor (2) and (3); and the descriptor written the other way around, modality then symmetry (4).

### 8.6.2. Prior conceptions on describing distributions

When looking at the descriptions for distributions it was decided initially that shape, range, median/centre, middle group and peak(s) (Figure 8-15) would be the starting point for describing distributions. In partial alignment with Bakker and Gravemeijer's (2004) aspects (Figure 8-4), it attends to centre through the median, spread through the range, density through looking at the middle group, and skewness through the description of the shape.

Discuss with students what key features of a graph to describe are.

- Put the challenge out if they had to draw the graph from the description only what info would they need. Collect ideas from the class.
  - Suggest may be: shape, description of range, median/centre, middle group, and peak(s) – there may be other features, discuss as a department first.

Figure 8-15. From lesson notes re: descriptions

On reflection this list does cover many ideas, but does not get at the heart of distribution nor really address the overarching statistical concepts that underpin understanding of distribution. This is true also of Bakker and Gravemeijer's (2004) list of aspects: they cover some of the characteristics of distribution, but again do not fully encompass the overarching statistical concepts; rather they assume knowledge of them.

### 8.6.3. What makes a good distribution description

Distribution is a complex notion and what makes a good distribution description is complex as well. Before any criteria could be developed as to what make a good description, distribution needed to be unpacked fully to appreciate all of the relevant components. This process involved thorough analysis of student responses in their pre- and post-tests by looking for patterns in their responses and aligning them to the two frameworks (Bakker & Gravemeijer, 2004; Ben-Zvi et al., 2007) outlined in section 8.3.

### ***Distribution description framework***

During the process of analysis of student responses, the two frameworks (Bakker & Gravemeijer, 2004; Ben-Zvi et al., 2007) were found to only provide part of the picture. The frameworks needed to be linked with a specific focus on the underlying conceptual structure of distribution. Hence, the distribution description framework (DDF) for thinking about, exploring and describing distribution was developed (Figure 8-16, next page). The DDF was developed as a result of needing to connect the two frameworks (Bakker & Gravemeijer, 2004; Ben-Zvi et al., 2007) and to connect them to the in-class observations, and pre- and post-test responses. Collectively these sources of data and ideas built a richer picture of the possible features that may be present in a particular distribution. Some aspects, such as variable and overall shape, will be true and relevant in all descriptions, while other aspects (for example, clustering density or mode) may or may not be relevant in the description because these aspects depend on the data and therefore the type of distribution. The DDF is organised by: (1) *overarching statistical concepts* that underpin distribution, (2) *characteristics of distribution*, and (3) the *specific features* that are used when describing distributions. Examples of specific features that a student might give to describe a particular distribution are the range (a feature of spread, which is a characteristic of variability) or the median (a feature of centre, which is a characteristic of signal and noise).

Ben-Zvi, Gil and Apel's (2007) cognitive aspects from their IIR theoretical framework – reasoning about variability, distributional reasoning, reasoning about signal and noise, contextual reasoning and graph comprehension – were used to inform the overarching statistical concepts for distribution descriptions. Bakker and Gravemeijer's (2004) characteristics of distribution – centre, spread, density and skewness – formed the backbone, with Pfannkuch, Regan, Wild and Horton's (2010) ideal data-dialogue providing further characteristics and features to supplement those listed in the IIR theoretical framework. The result of the analysis of student pre- and post-test responses and in-class observations provided additional characteristics and features, noted in the DDF (Figure 8-16) in italics.

| Overarching statistical concepts | Characteristics of distribution | Specific features measures/depictions/descriptors                                                                                                                      |
|----------------------------------|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Contextual knowledge             | Population                      | <i>Target population</i> (e.g. New Zealand year 5–10 students)<br><i>Other acceptable population</i> (e.g. year 5–10 students)                                         |
|                                  | Variable                        | Variable<br>Units                                                                                                                                                      |
|                                  | Interpretation                  | <i>Statistical feature described in contextual setting</i> (e.g. interpreting right skew as very few high test scores, with most test scores between 20 and 50 points) |
|                                  | Explanation                     | <i>Possible reason for a feature</i> (e.g. bimodal due to gender for kiwi data)                                                                                        |
| Distributional                   | Aggregate view                  | General shape sketched<br>Hypothesis and prediction                                                                                                                    |
|                                  | Symmetry                        | Overall shape                                                                                                                                                          |
|                                  | <i>Modality</i>                 | <i>Modality</i>                                                                                                                                                        |
|                                  | Skewness                        | Position of majority of the data (to the left or the right)                                                                                                            |
|                                  | Individual cases                | <i>Highest and lowest values</i>                                                                                                                                       |
| Graph Comprehension              | Decoding visual shape           | Overall shape<br><i>*Parts of the whole</i> (splitting the distribution into parts and describing the parts as well as the whole)<br><i>Modality</i>                   |
|                                  | Unusual features                | Gaps<br>Outliers                                                                                                                                                       |
| Variability                      | Spread                          | Range, interquartile range<br><i>*Interval for high and/or low values</i> (may be describing a tail)                                                                   |
|                                  | Density                         | Clustering density<br>Majority (mostly, many)<br>Relative frequency                                                                                                    |
| Signal and noise                 | Centre                          | Median, mean                                                                                                                                                           |
|                                  | Modal clumps                    | <i>Peak(s) (local mode)</i><br><i>Modal group(s)</i>                                                                                                                   |

Figure 8-16. Distribution description framework for curriculum level 5 (ages 13–15)

Note: \* indicates part of feature listed.

When students are describing statistical distributions they need: (1) to invoke contextual knowledge, (2) to know what relevant characteristics of distributions they can actually see in the plots and therefore describe, and (3) to be explicit about the evidence for specific features. In other words, students need to be able to identify which features are evident in a particular plot, name and provide evidence (values) for the features, and to interlace these with contextual information such as the population, variable and units.

### ***Criteria for what makes a good description at level 5 (ages 13–15) in the New Zealand curriculum***

Multiple reviews of student responses in pre- and post-tests along with the concepts and ideas summarised in the distribution description framework (Figure 8-16) were used to establish the criteria for a good description of a distribution.

In a nutshell a good description has the context correctly identified including both the variable and the target population. The context is connected throughout the description using both the variable and units and the population more than once. The overall shape is described and at least three other features. A very good description would also include either some explanation or interpretation of the results (Figure 8-16).

An analogy of a good description is to think of the body. The body in a simplistic form is made up of three components: the skeleton; the connecting muscles, tissues and vessels; and the skin. With just the skeleton we have an outline of the form of the body (Figure 8-17a); with the addition of the connecting muscles, tissues and vessels we start to get a better picture of the final form (Figure 8-17b); but it is not until we add the skin (and fur) do we get the real picture (Figure 8-17c). Giving a good description of a distribution is like layering on each component of the body.

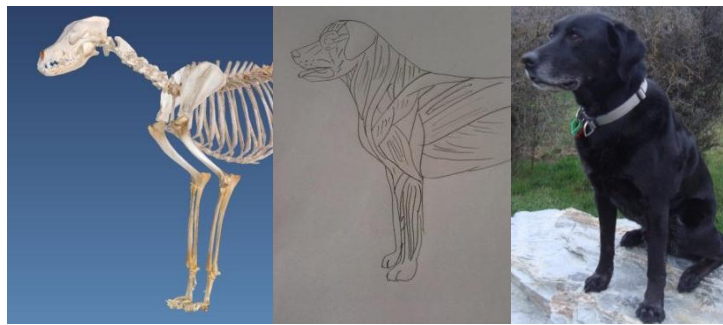


**Figure 8-17. (a) skeleton of a dog (University of Bristol & Creative Dimension Software Ltd, 2011)  
(b) dog showing muscular structure (c) the whole dog**

*Note:* (a) Reprinted from “Canine Skeleton,” by University of Bristol and Creative Dimension Software Ltd, 2011, retrieved from <http://www.real3danatomy.com/bones/dog-skeleton-3d.html>. Copyright 2011 by the University of Bristol and Creative Dimension Software Ltd. Reprinted with permission.

(b) Author’s sketch.

(c) Author’s photograph.

This analogy can be seen in the two graphs and the written statements that make up Figure 8-18. If the shape is described first, a mental visual image of the distribution (Figure 8-18a) is already forming as the description is read. As different features are described (for example, values added onto the x-axis of Figure 8-18b), detail is added to the skeleton of the description. However, it is the context in the form of the variable, population and units that confirms what the description is about – in this example, the first statement in Figure 8-18a establishes the context, which is then continually weaved throughout the description in Figure

8-18b. Without these three components – description, detail, and context – only a partially reconstructed visual image is possible.

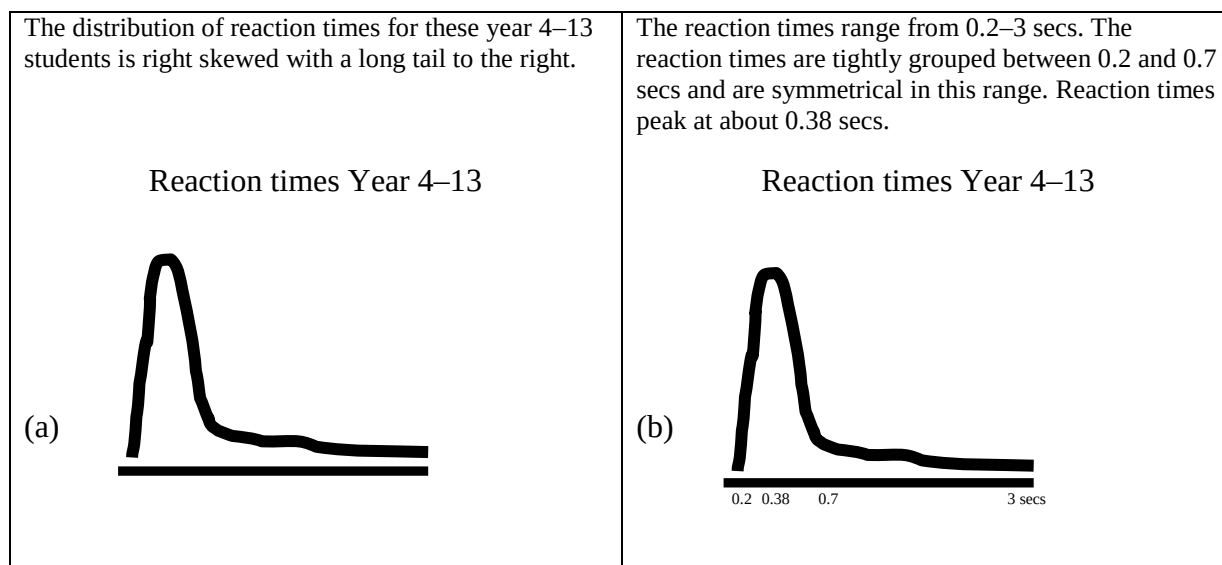


Figure 8-18. (a) drawing the shape (b) filling in the features

This type of description was modelled throughout the teaching and learning sequence, although at the beginning of the sequence what made a good description was not articulated. Through the retrospective analysis, however, a good description was able to be articulated based on student responses and the literature. It appeared that “experience, war stories and intuition” (Wild & Pfannkuch, 1999, p. 223) of the teacher, researcher and researcher’s supervisors were sufficient to set the students on the right path in terms of describing distributions. The retrospective analysis has allowed the statistical practice of describing distributions to be made overt and provided an opportunity for formal instruction to support students as they gained their own experience.

### 8.7. Quantitative analysis of test questions

All three research questions are discussed in this results section by looking at student responses in the pre- and post-tests and in the pre- and post-interviews to questions pertaining to distribution. Specifically, these are tasks C, D and E in the pre- and post-tests, and the associated activities for tasks C and E in the pre- and post-interviews. Three aspects relating to the three questions will be discussed: (1) *shape descriptors* – What descriptors do year 10 (ages 14–15) students intuitively use for distributional shape?, (2) *describing distributions* – What makes a good distribution description at level 5 (ages 13–15) in the New Zealand

curriculum?, and (3) *predicting distributions* – What distributional shapes and graphs do year 10 (ages 14–15) students predict when given the context?

### 8.7.1. Shape descriptors

In the pre- and post-test in task C the students were asked for each of the three situations given (Figure 8-19) to sketch the shape of the distribution of the variable and to write two statements about the distribution of the variable.

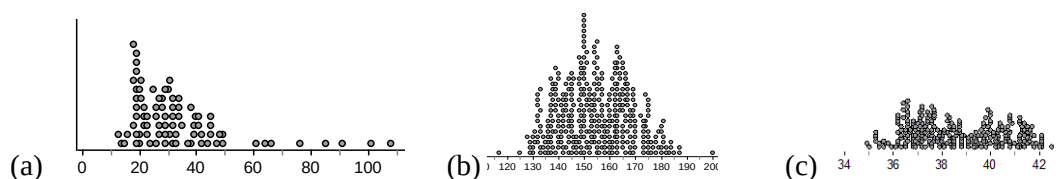


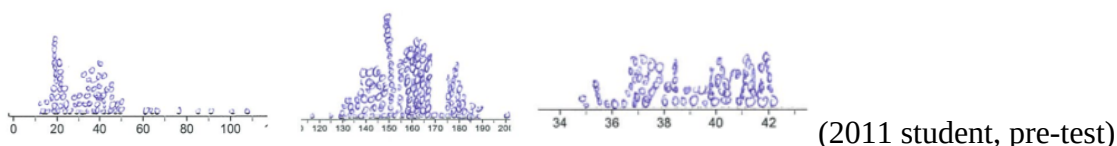
**Figure 8-19. (a) All Blacks' (NZ rugby team) scores in test matches 2005-2010, (b) heights of NZ Year 5-10 students, (c) heights of Tokoeka Kiwis (NZ native bird)**

#### *Sketching the shape of the distribution*

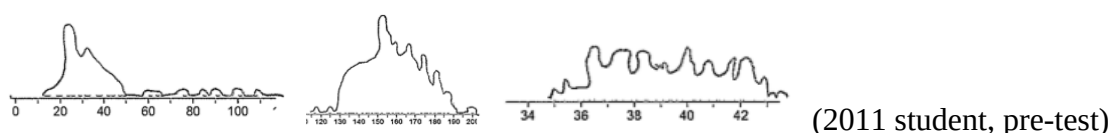
The quality of the shapes (curves) sketched fell into four broad categories. Each of the curves was scored according to the criteria below. With the criteria are examples for each of the three contexts.

Score 0 – No shape drawn

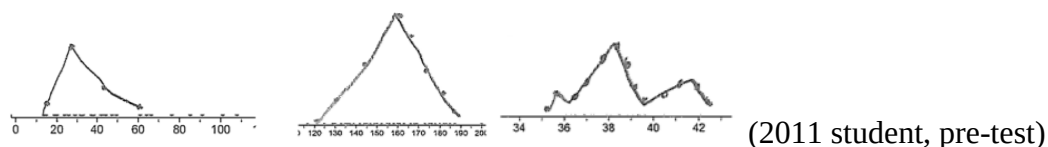
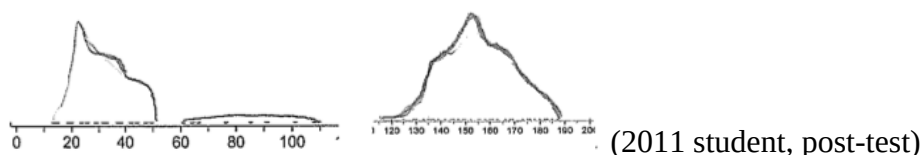
Score 1 – have redrawn the graph with the points again



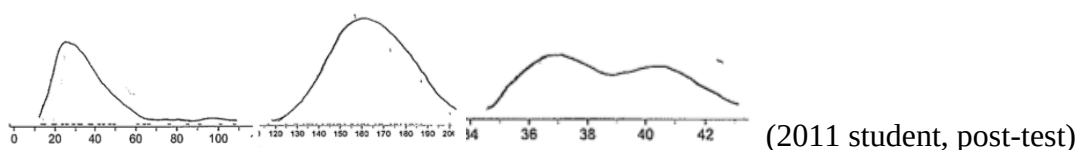
Score 2 – have drawn a curve with multiple bumps; basically drawing the city skyline; drawing literally around the data



Score 3 – smoother curve, but additional bumps OR connected the dots to make a polygon shape rather than a smooth curve



Score 4 – smooth representative curve



The scores for curve sketching in the pre-test ranged from 0 to 4 with a median of 2. Only three of the sketched graphs scored 4. The scores for curve sketching in the post-test ranged from 2 to 4 with a median of 4. This means that nearly all of the students in the post-test were sketching curves that reasonably well represented the data and were no longer doing the detail of the city skyline, and most were doing a smooth representative curve of the data.

In the pre-test, 4% of the student responses across the three situations showed a smooth and appropriate curve for the shape (score 4); in the post-test, 62% of the student responses showed a smooth and appropriate curve for the shape across the three situations.

### ***Describing the shape of the distribution***

In terms of describing the shape of the distribution this was elicited from the two statements about the distribution of the variable in task C.

**Pre-test:** In the pre- test, only one instance of a statement that might possibly be describing shape was found: “They are very evenly distributed between 36 cm and 42 cm with no big spikes or drops” (2011 student, pre-test).

The “evenly distributed” could also refer to spread, but the inclusion of “no big spikes or drops” suggests that in fact this student is attempting to describe the shape. The language used is not overly sophisticated, though it does have hints of statistical terminology – evenly, distributed, spikes. So only one weak piece of evidence for describing shapes could be found in all the pre-test responses.



**Post-test:** The findings were very different, however, for the post-test responses. With the exception of one description, across the three situations and the 29 students who sat the post-test, all remaining 86 descriptions included a statement about shape. Student responses were analysed and the descriptors they used for shape were coded by symmetry and by modality. The three situations are discussed separately as the responses varied.

***All Blacks' scores in test matches 2005–2010*** – The graph of All Blacks' scores in test matches (2005–2010) can be best described in terms of shape as right skewed and unimodal (Figure 8-19a). Bimodality was not accepted as appropriate for this data. All the students described the distribution of All Blacks' scores as right skewed. Of the 29 students who responded, 26 of them also said that the distribution of All Blacks' scores was unimodal. One student made no comment about modality and the remaining two students said that the distribution of All Blacks' scores was bimodal. These two students' descriptions matched their sketches and while their description is consistent with their sketch, their sketch did not represent the data. This means that in the end their description was not consistent with an image of the population data, a smoothing out of what is seen. To summarise: 26 students gave a full description with both components, one student had a partial description with only one component, and two students had a partial description with an error.

***Heights of New Zealand Year 5–10 students*** – The graph of heights of New Zealand year 5–10 students can be best described in terms of shape as symmetrical and unimodal. All the students described the distribution of heights of New Zealand year 5–10 students as symmetrical. Of the 29 students who responded, 25 of them also said that the distribution of heights of New Zealand year 5–10 students was unimodal, giving a full description. Four students made no comment about modality, only giving a partial description.

***Heights of Tokoeka kiwis*** – The graph of heights of Tokoeka kiwis can be described in a number of ways depending on the sketch that the students have done and their interpretation of that sketch (Figures 8-20b, c, and d (next page)). Acceptable descriptions include symmetrical and bimodal, symmetrical and unimodal, right skewed and unimodal, right skewed and bimodal.

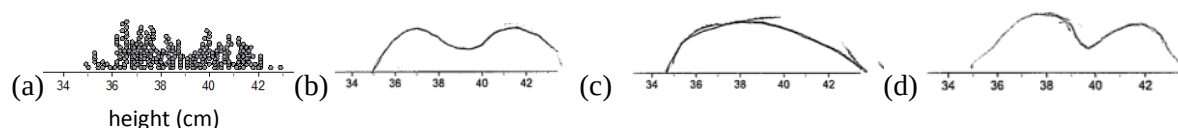


**Figure 8-20. Heights of Tokoeka kiwis: (a) provided graph  
(b) example of sketched shape approximately symmetrical and bimodal  
(c) example of sketched shape slightly right skewed and unimodal  
(d) example of sketched shape slightly right skewed and bimodal**

There were 13 different categories of response to the Tokoeka kiwi height data. Fifteen of the 29 students gave a full description of the shape of the distribution of heights of Tokoeka kiwis. The most common response (seven responses) was that the distribution of heights for Tokoeka kiwis was approximately symmetrical and bimodal (Figure 8-20b), followed closely by those students who responded with the distribution of Tokoeka kiwi heights being right skew and bimodal (six responses) (Figure 8-20d). Eight students gave a partial description and this was a mixture of the two components, some describing the symmetry and some describing the modality. Five students gave an incorrect description of the shape and one student did not describe the shape.

**Overall post-test** – Of the 86 descriptions about shape given by students in the post-test, 66 responses gave a full description (25 in each of the first two situations and 16 in the last situation). Thirteen responses had a correct partial description and the remaining seven responses were either an incorrect response or had an error in the response. The 86 responses (92%) that were either full or partial descriptions of shape compares with just one response (1.2%) in the pre-test.

### ***Discussion on shape descriptors***

Introducing the students to the correct statistical terms for describing shape seems to have improved their ability to describe the shape of distributions. In the post-test, 92% of the descriptions made by students gave an indication of the shape of the distribution, with 74% being full descriptions. Students at ages 14–15 appear to be able to use the correct statistical terminology once they have been introduced to it.

Ability to sketch the distributional shape improved from only 4% of the original responses being suitable to 62% being suitable by the post-test. Modelling sketching shape with the students appears to have been beneficial in improving their ability to sketch the shape of a distribution. This is in regard to reasonably large sample sizes ( $n > 80$ ) and to graphs that have sensible scales so the shape is more easily seen. There was no assessment of situations

with small sample sizes (around 30) or where students were working from their own constructed graphs.

### 8.7.2. Describing distributions

Student pre- and post-test responses were analysed to see if their ability to describe distributions had improved over the course of the statistics unit. In task C students were asked for each of three situations (Figure 8-19) to sketch the shape of the distribution of the variable and to write two statements about the distribution of the variable.

#### *Overall description ranking*

The SOLO taxonomy (Biggs & Collis, 1982; Hook & Mills, 2011; Uniservices asTTle team, 2008; Watson, 2006) was used as a basis for grading student responses. The particular descriptors (Figure 8-21) aligned to each question were developed through a process of moving between the literature, in-class observations and student responses, and were based on what makes a good distribution description (see pages 217–219).

| SOLO taxonomy level          | Grade | Description of evidence                                                                                                                                                                                                                                                 | Example of student responses with commentary                                                                                                                                                                                                                          |
|------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| No response or idiosyncratic | 0     | <ul style="list-style-type: none"> <li>No response.</li> </ul>                                                                                                                                                                                                          | No examples.                                                                                                                                                                                                                                                          |
| Pre-structural               | 1     | <ul style="list-style-type: none"> <li>Context and/or evidence missing.</li> </ul>                                                                                                                                                                                      | Bunched up between 20 and 50? (2011 student, pre-test). <ul style="list-style-type: none"> <li><i>no context</i></li> </ul>                                                                                                                                           |
| Uni-structural               | 2     | <ul style="list-style-type: none"> <li>Students give one correct* piece of evidence in simple context (usually variable only; e.g. heights, scores).</li> </ul> OR <ul style="list-style-type: none"> <li>Evidence for multi-structural without any context.</li> </ul> | The results are quite spread out. There is a high from 36–38 and 40–42. Then there are fewer heights from 35–36 and 42–43 (2011 student, pre-test). <ul style="list-style-type: none"> <li><i>one piece of evidence density, and simple context (just)</i></li> </ul> |

| SOLO taxonomy level                                                                                                  | Grade | Description of evidence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | Example of student responses with commentary                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----------------------------------------------------------------------------------------------------------------------|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Multi-structural                                                                                                     | 3     | <ul style="list-style-type: none"> <li>Students have identified a simple context.</li> <li>They have correctly described two features.</li> </ul> OR <ul style="list-style-type: none"> <li>Evidence for relational without any context.</li> </ul>                                                                                                                                                                                                                                                                                                                     | <p>The most common height is between 36 cm–37 cm. The heights range from 35 cm–43 cm (2011 student, pre-test).</p> <ul style="list-style-type: none"> <li><i>two correct features (mode and range) and simple context</i></li> </ul> <p>The distribution for the scores of the AB is approx right skewed with a tail to the right. It peaks at approx 20. It seems like the 2 groups, one tightly together from 20–45 and another from 60–100. It ranges from 0–100 (2011 student, post-test).</p> <ul style="list-style-type: none"> <li><i>three correct features (shape, mode, modal clump) although range incorrect; acceptable close population and variable</i></li> <li><i>missing the context connected throughout, i.e. apart from first sentence there is no context in rest of description, no units given; therefore is not a relational level description</i></li> </ul> |
| Relational<br>(see significance of parts of the whole, can sketch the graph with shape and two pieces of evidence)   | 4     | <ul style="list-style-type: none"> <li>Students have identified the context including the variable and either the target population or acceptable close population (see Figure 8-16).</li> <li>They have connected the context throughout most of the description through use of the variable and units.</li> <li>They have correctly described the overall shape and at least two other features.</li> </ul>                                                                                                                                                           | <p>The distribution of the heights (cm) of NZ Yr 5–10 students is approx symmetrical and unimodal. The heights range from 110–200. There is a peak at 150. A large group is located from 130–180 cm. If another sample was taken, these plots could change (2011 student, post-test).</p> <ul style="list-style-type: none"> <li><i>four correct features (shape, range, mode, modal clump); context (variable and units) connected throughout most of the description</i></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                 |
| Extended abstract<br>(integrating the statistical and contextual, also can be seeking explanation or interpretation) | 5     | <ul style="list-style-type: none"> <li>Students have identified the context including the variable and the target population.</li> <li>They have connected the context throughout the description through use of the variable and units. They have included the population more than once.</li> <li>They have correctly described the overall shape and at least three other features.</li> <li>They may have included some explanation or interpretation of data to the context (e.g. explanation – these two groups might mean the two different genders).</li> </ul> | <p>The distribution of the heights of Tokoeka kiwis is approx symmetrical and bimodal. The heights range from 35–43 cm. The middle Tokoeka kiwi height is 39 cm. The heights peak at around 36.5 and 40 cm. The heights are tightly grouped in two groups one between 36.1–39 cm and another between 39–42 cm. These two groups might mean the two different genders (2011 student, post-test).</p> <ul style="list-style-type: none"> <li><i>five correct features (shape, range, median, modes, modal clumps); context connected throughout description (variable, units and population); possible explanation given</i></li> </ul>                                                                                                                                                                                                                                                 |

Figure 8-21. Criteria for a good description of distributions

Note: \*means the value(s) is given for the feature and it is the correct value(s)).

The median grade across the three situations was used to represent the students' overall grade for describing distributions. These are summarised in Figure 8-22a.

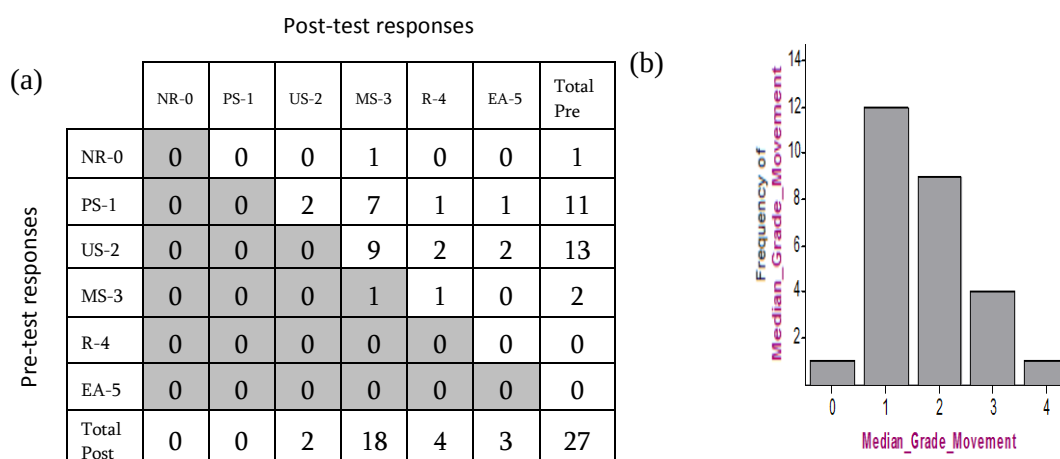


**Figure 8-22. (a) Pre- and post-test results for task C (b) Median grade movement from pre- to post-test**

In the pre-test the highest median grade was multi-structural (MS-3) with two students achieving this grade. In the post-test three students achieved at extended abstract (EA-5) and all but two students reached at least a multi-structural level. This means that the students could identify the context and describe at least two features of the distribution. Many of these students actually described more than two features, but they failed to make the broader link to the context, which was required to show relational thinking (R-4). The biggest movements were from students who scored 0–2 in the pre-test, perhaps indicating that acquisition of language and knowledge for describing distributions assists students. Figure 8-22b shows the median difference between students' pre- and post-test scores for describing distributions. The students made a significant improvement ( $p$ -value  $\approx 0$ , paired  $t$ -test) in their median scores for describing distributions from the pre- to post-test, and on average increased their median grade by 1.7 points (95% C.I.= [1.34, 2.07]).

### ***Discussion on describing distributions***

At level 5 (ages 13–15), students are grappling with the different features that they need to describe. Most of the students seemed to have a reasonable grasp on describing distributions, since in the post-test 80 of the 89 descriptions reached at least multi-structural level. Of the 89 descriptions, 15 displayed relational thinking, meaning the description incorporated the context beyond the opening statement and described at least three features including shape, and 10 descriptions displayed extended abstract thinking (one description with an

explanation), integrating context throughout the description and connecting with the context (as might be generated by the investigative question).

### 8.7.3. Predicting distributions

The pre- and post-test task E required the students to sketch what they suspected might be the distribution of data for the variable given (age), and for each of the three populations (a class of year 10 students, CensusAtSchool year 5–13 students, and New Zealand mathematics teachers at a conference).

#### *Pre-test*

In the pre-test, only seven of the 27 students who sat the test had a sensible sketch for at least one of the three populations. All seven had a reasonable idea about what the graph of ages for year 10 students would look like with the ages ranging from around 13 years to around 16 years with the peak around 14 or 15. The sketch in Figure 8-23 is one of the better examples, especially as the number of students on the vertical scale is commensurable with the population size of 28.



Figure 8-23. Student response to task E, year 10 students (2011 student, pre-test)

Four of these students also had the idea that the year 5–13 student ages would range from 8 to 18 years, but only one of them gave a sensible justification for their graph – “There are roughly the same number of students in each year until year 12 and 13, when some students drop out” (2011 student, pre-test).

Only two students had a possible valid sketch for the age of New Zealand mathematics teachers at a conference. One had the ages ranging from 22 to 60 with the justification that “Teachers finish college after a couple of years, and then they can start teaching” (2011 student, pre-test) and the other had the ages ranging from 35 to 65 because “no one in their right mind will teach maths unless they are older” (2011 student, pre-test).

Of the remainder of the responses, nine students had made no attempt at any of the questions, four students had attempted one question and left the other two blank, and the rest had drawn something but it was not correct or relevant to the situations described.

### ***Post-test***

In the post-test all the students attempted the task and had a sketch for each of the three populations. Seventeen students had a reasonable shape (reflecting the possible distribution), with sensible values for the variable (age) and a justification that made sense and aligned with their sketch for all three populations, and three students gave a satisfactory response for two of the three populations. The examples in Figure 8-24 are indicative of the responses that the students gave for the predicted distributions.

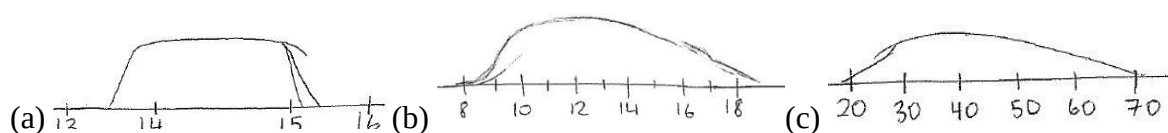


**Figure 8-24. Sketch of predicted distributions for age (a) year 10 class, (b) year 5–13 students C@S, (c) teachers at New Zealand Maths conference (2011 student, post-test)**

Two students had a sensible age range for their predictions, but the shapes of the graphs were not consistent with the actual situation. One student had reasonable shapes, with sensible comments, but their statements and graphs were missing possible values for the variable. The remaining six students had sketched possible distributions, but these were either incorrect, or they lacked any labelling or description that might have put them into one of the groups described previously.

### ***Discussion on predicting distributions***

The students seem to be able to predict the distributions for the given scenarios. Twenty students, in the class of 29, had reasonably sketched and justified at least two of the scenarios described in task E in the post-test. In the pre-test, only four students had achieved this, with two of the four barely meeting the criteria for all three scenarios.

The students appear to have used visual imagery (Figure 8-6, page 200) and used their knowledge on the basic distribution shapes to sketch their predicted distribution, as many of the students used these shapes (Figure 8-14, page 214). To be able to select the shape to begin with, students need to have a reasonable understanding of the context (contextual

knowledge), including possible values for the variable, and use this contextual knowledge with their statistical knowledge about distributional shapes to sketch the predicted distribution graphs.

## 8.8. Discussion

In chapter 7 the underpinning concepts of answering comparative investigative questions was discussed. Students need to develop conceptual understanding of sample, population, sampling variability and sample-to-population inference to support their using decision criteria to make a call consistently and coherently to answer their comparative investigative question.

In chapter 8 the design of the teaching sequence to develop the notion of distribution as a key concept for answering summary investigative questions was discussed. Analysis of students' pre- and post-tests and pre- and post-interviews, classroom transcripts, field notes and students' classwork were all used to define and show how the students in 2011 progressed towards describing distributional shape, describing distributions and predicting distributions. This chapter has now addressed the three research questions: (1) What descriptors do year 10 (ages 14–15) students intuitively use for distributional shape? (2) What makes a good distribution description at level 5 (ages 13–15) in the New Zealand curriculum? and (3) What distributional shapes and graphs do year 10 (ages 14–15) students predict when given the context?

In response to the first question: the distribution shape descriptors that the students intuitively use have two components to them (Figure 8-14, page 214). The first component is symmetry of the graph, i.e. whether the is graph symmetric or asymmetric, and this gives five categories: uniform, bell shaped or normal, other symmetric graphs, left skew and right skew. The second component is modality, i.e. whether the graph is unimodal, bimodal or has some other modality. The use of modality as part of the distributional shape descriptor is a new idea as previous descriptors such as uniform, normal, skew and bimodal (Bakker, 2004a; delMas et al., 2005) did not necessarily attend to both components. Skew as a descriptor only attends to symmetry and bimodal only attends to modality, whereas the descriptors uniform and bell-shaped (i.e. normal) – special cases of symmetrical graphs – attend to both components. (By definition a uniform distribution has no mode and a bell-shaped or normal distribution has one mode.) Previous shape descriptors given (Bakker, 2004a; delMas et al.,



2005) missed other types of symmetric graphs and the addition of the modality component allows for a more efficient picture to be formed in one's mind as the descriptor is read and more accurately describes the distributional shape.

For students to successfully classify distributional shape from samples they need to be able to infer the shape. This is more easily done from dot plots (Bakker, 2004a; Pfannkuch, 2006), but also becomes easier with experience. Experience is acquired over time, but in this research a deliberate act was made to build students' experiences faster. This was through the development of a contextual library of variable and population shapes (see pages 211– 212). Connected to this is predicting distributional shapes before the data are looked at or sourced. The prediction, too, is based on knowledge of the variable and population and this knowledge is built using previous experiences. These students are beginning to understand how to draw and describe distributional shape, inferring the shape through sketching, describing their sketched shape using symmetry and modality, and, in some instances, they are also able to articulate explanations or interpretations of the data relative to the shape.

In terms of the second question: a good distribution description at level 5 (ages 13–15) includes a description of the overall shape of the distribution and at least two other features (Figure 8-16, page 217) and links these features to the context through the variable, units and an acceptable population descriptor, while a very good distribution description describes at least three features of the distribution in addition to the overall shape, connects the context throughout the description and may include some explanation or interpretation of the data in context. The features of distribution at level 5 (ages 13–15) in the New Zealand curriculum are specified in the distribution description framework (DDF) in Figure 8-16 (page 217). These features are organised by the overarching statistical concepts: contextual knowledge, distributional, graph comprehension, variability, and signal and noise. By attending to the different features, a good picture of the distribution can be build up.

The DDF was designed to support the development of the notion of distribution. The DDF for level 5 (ages 13–15) reflects the statistical knowledge related to distribution that students at this level have access to. Students at this level should be able to call upon any of the features of the DDF, depending on the context and the data available. The DDF can be expanded to include more features as the curriculum level increases; for example, at curriculum level 6 (ages 15–16) this would include features around comparing distributions, and as the distribution description framework extends further up the curriculum levels, it would include

distributions of sample statistics. Similarly, the DDF can be modified to support student progressions at lower curriculum levels. The DDF has the potential to inform curriculum developers, researchers and teachers as they introduce students to the conceptual structure underlying distribution. Further research is needed both above and below the curriculum level reported here to ascertain what is appropriate for students at the different curriculum levels.

In answer to question three: the distributional shapes and graphs that the students predict fall into the categories of distributional shapes as discussed previously (section 8.6.1, page 213). The students appear to have an image of the distribution (Figure 8-3, page 193) in mind and have sketched this with justification. The students have connected their visual image, analytical thinking and verbal description (Figure 8-6, page 200) together to sketch a sensible graph with appropriate values for each of the scenarios. As students build their contextual shape library, they build the experiences they can call on to predict distributions, and their growing knowledge of the different features of distributions supports the detail in their prediction; for example, giving maximum and minimum values, where the peak(s) or mode(s) is, and an idea of the modal cluster location. The process of drawing a predicted distribution calls on the mental processes of visual, analytic and verbal descriptive thinking (Aspinwall et al., 2008). The visual thinking is captured by the construction of the basic distribution sketch; adding detail to the sketch (for example, age ranges or what the peak value might be) supports the analytical thinking; and justifying, either verbally or in written form, connects the image to analytic thinking and helps to sustain student thinking.

In summary, these students' results suggest that they are able to produce a visual mental image of the situation that is given, and they are able to locate the distribution on a scale sensibly. The students appear to have used both their new statistical knowledge about distributional shape and their growing contextual knowledge library. The students are using both contextual and statistical knowledge when they are describing distributions. They have been able to infer the distributional shape of the data and use this and other features to describe the distribution of the data. They are starting to think like a statistician, looking at the whole and starting to unlock the stories in the data.

### **8.9. Practical implications**

The associated teaching and learning activities have been shared with teachers at various workshops in 2011–2013, across New Zealand. See, for example: <http://nzstatsedn.wikispaces.com/Gisborne>.

A paper focusing on distributional shape was presented at the International Congress on Mathematical Education (ICME-12) in Korea (Arnold, P., & Pfannkuch, M. 2012).

## **Chapter 9. Theoretical Frameworks, Implications, and Where to from Here**

In the final chapter the grand themes are pulled together. The findings from the research are reflected upon in a much larger arena, linking back to the literature and connecting with the research questions. Implications for policy and practice and areas for further research are discussed within the broader context of the grand themes. Limitations of the research complete the chapter.

### **9.1. Overarching themes and frameworks**

This research thesis started by looking at teachers' knowledge of statistics (adapted from Simon, 1995) and in particular the specialised statistical content knowledge needed by teachers to meet the requirements of the statistical investigation thread in the new New Zealand curriculum at level 5 (ages 13–15) (Ministry of Education, 2007). As will be shown, through the overarching themes and a stocktake of the frameworks and associated findings, this research thesis has completed a full cycle, starting with exploring teachers' knowledge of statistics and the statistics teaching cycle (Figure 3-2 and Figure 9-6) and finishing with a clearer picture of the knowledge of statistics needed for teaching and learning at curriculum level 5 (ages 13–15).

The initial problematic situation that was identified was posing investigative questions within the statistical enquiry cycle (the PPDAC cycle) by both teachers and students. The exploration and identification of underpinning concepts of posing good investigative questions realised the second problematic situation: that students and teachers did not have shared criteria with which to answer a comparative investigative question, a situation which required the use of informal inferential reasoning to draw conclusions. As with the second problematic situation, the third identified problematic situation arose from looking at the underpinning concepts that support posing and answering investigative questions. In this third problematic situation, two aspects were identified as being in need of attention: (1) describing distributions and building the concept of distributional shape; and (2) making predictions and connecting this to analysis as part of the process of answering investigative questions.

The four focus areas that were addressed in this thesis were: the investigative question, making predictions, analysis, and conclusions (Figure 9-1).

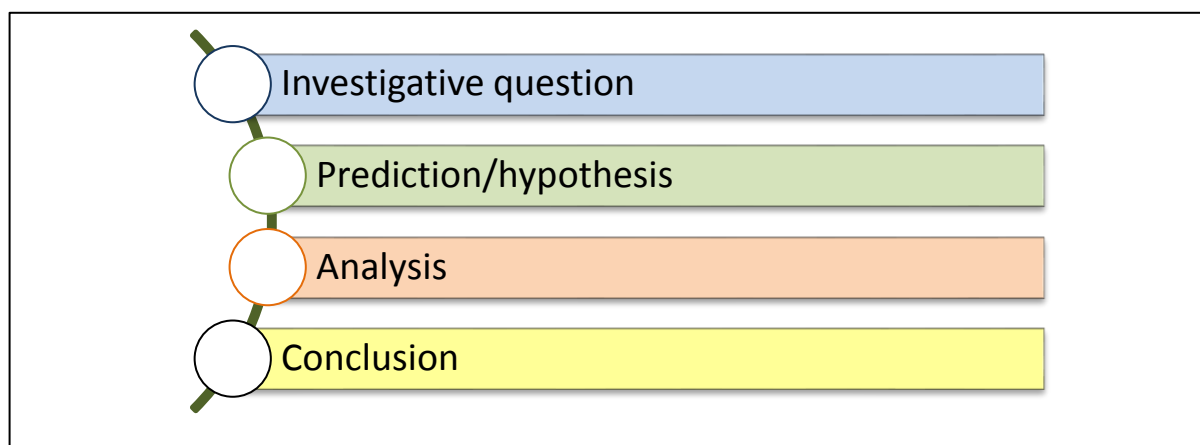


**Figure 9-1. The four focus areas arising from the problematic situations**

The research questions identified for each of the three problematic situations fit within the four focus areas, as shown in Figure 9-2.

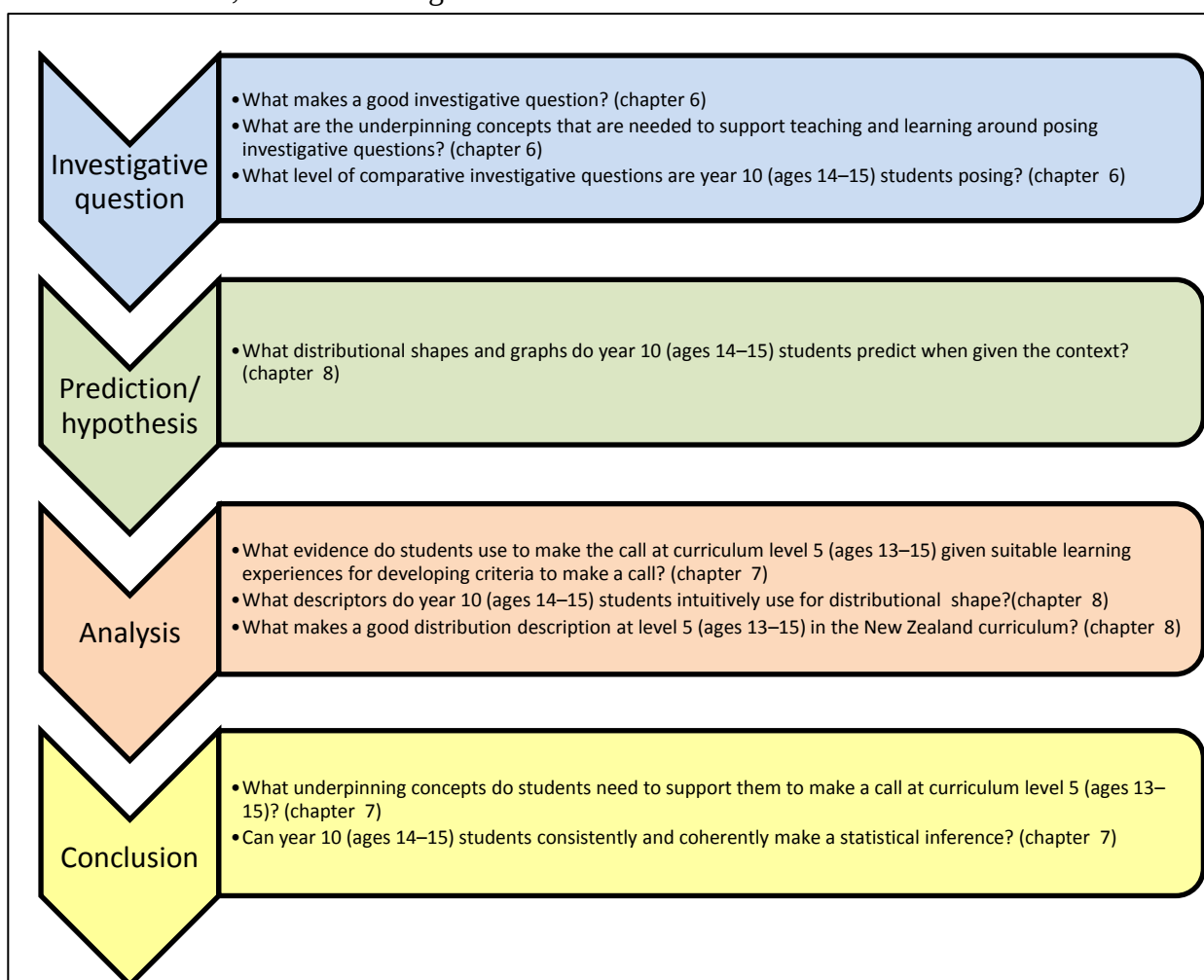


**Figure 9-2. Research questions within the four focus areas**

The answering and discussion of these research questions were detailed in chapters 6, 7 and 8. The purpose of this chapter is to draw out overarching themes from the specifics of the results and produce some big “take home” messages.

### 9.1.1. Overarching themes

In this research the overarching themes look at the art of teaching statistics. The overarching themes (Figure 9-3) support the teaching of the big ideas in statistics; for example, variation and distribution.

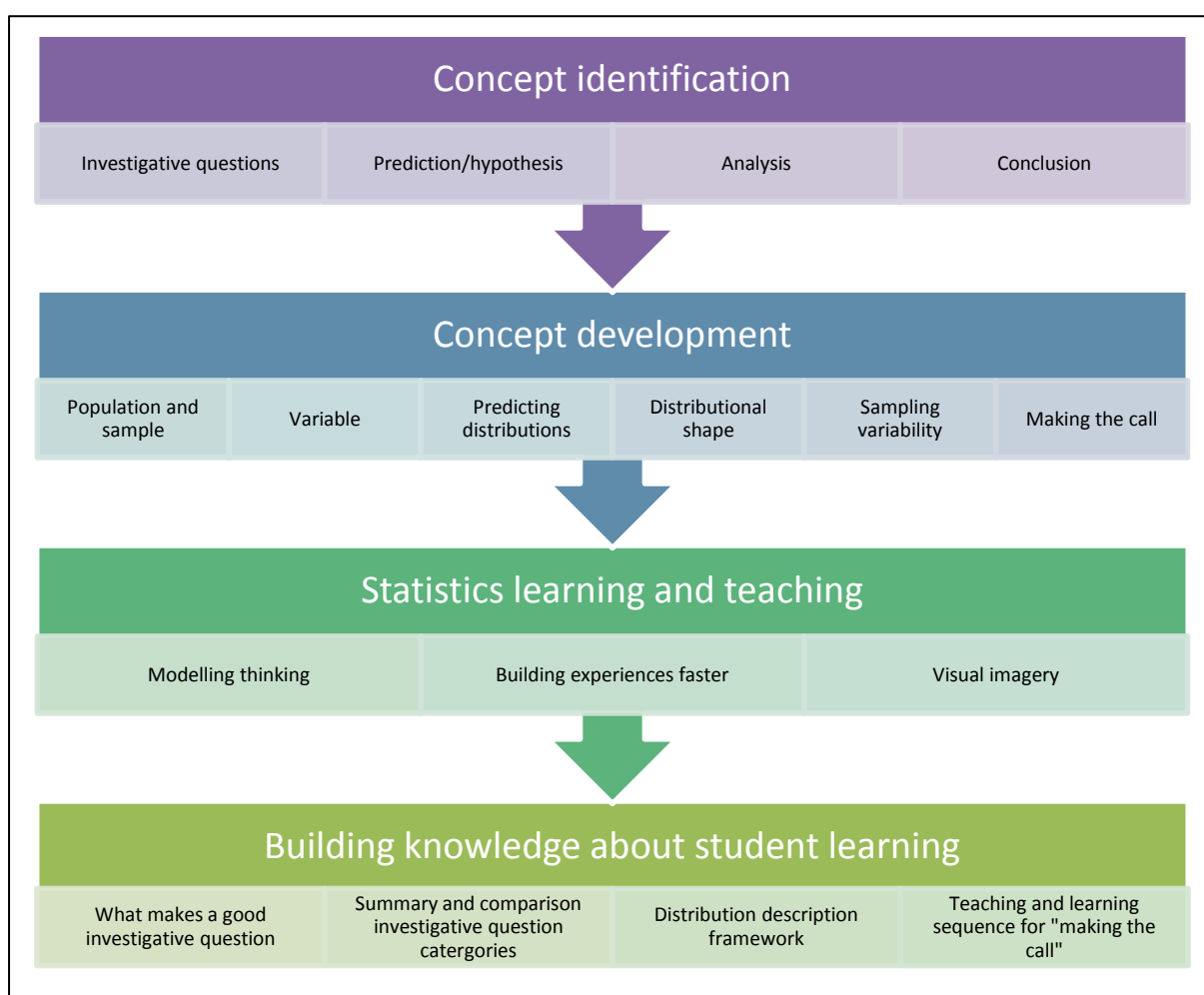


**Figure 9-3. Framework for overarching themes**

The overarching themes encompass: (1) concept identification – What concepts are needed to develop the four focus areas? (2) concept development – How are the concepts developed to support the particular focus area? (3) statistics learning and teaching – What teaching pedagogies support the concept development? and (4) building knowledge about student learning – proposed frameworks and criteria to support student learning. These overarching themes are now discussed.

### Concept identification

As each new problematic situation was identified, the concepts that needed to be taught were discussed in depth with both the teacher and the researcher's supervisor. The process of identifying the concepts and deliberately teaching them is new to many teachers in New Zealand. Previously, statistical calculations and graph-drawing skills were the focus of the teaching and learning, but this research set out to intentionally improve the quality of the content of statistics that is taught, and explicating the conceptual foundations was an overarching theme. When posing investigative questions was first identified as an issue, the concepts of population and sample were quickly identified as key concepts that needed developing.

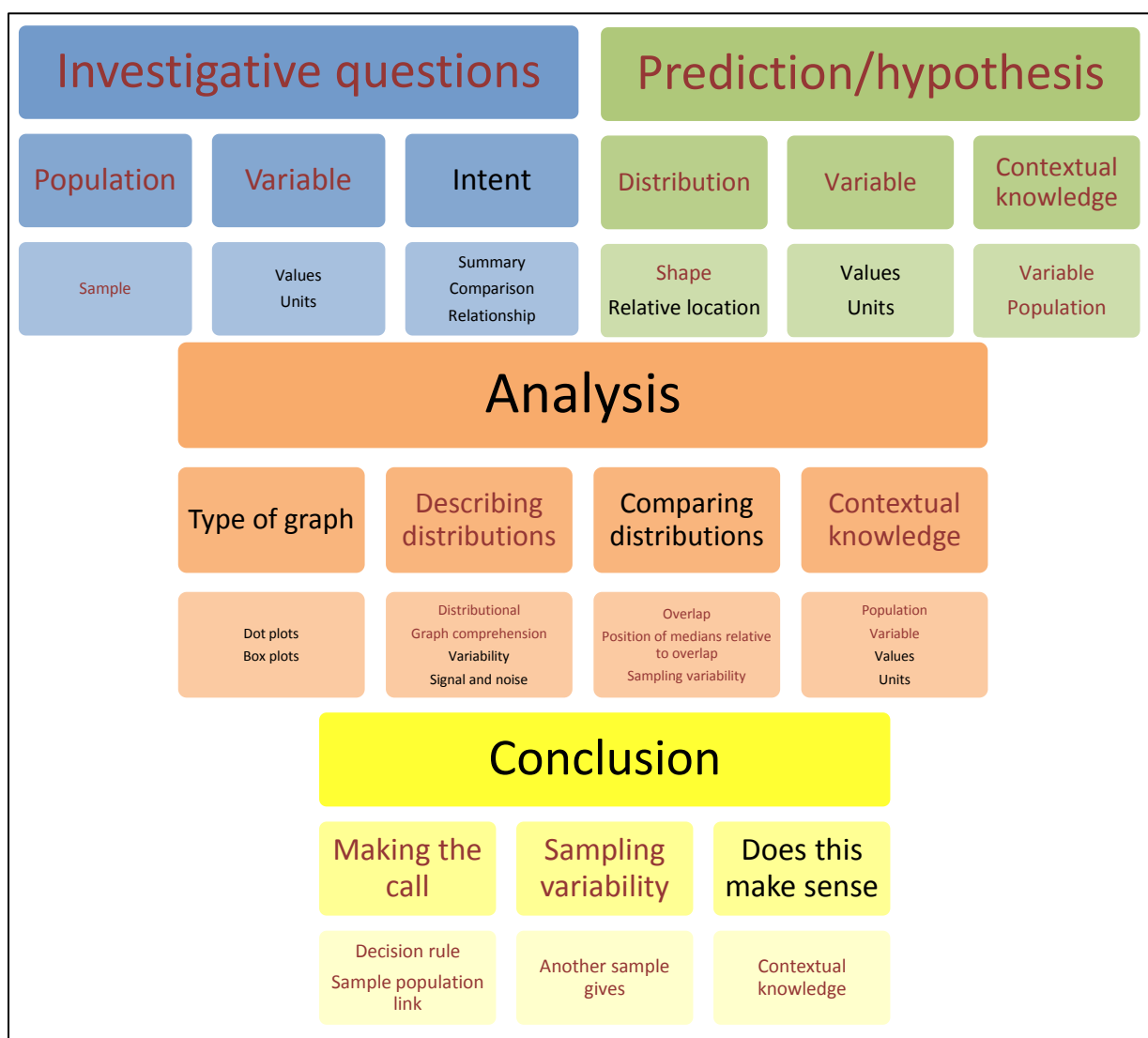


Figure 9-4. Conceptual framework for statistical investigations thread, level 5 of the New Zealand curriculum

Each of the four focus areas – investigative questions, prediction/hypothesis, analysis and conclusion – generated identification of their own “set” of concepts to be considered in the teaching and learning sequence. The infrastructure signalled in Figure 9-4 (previous page) is a suggested conceptual framework to support the teaching and learning of statistics at curriculum level 5 (ages 13–15) for the statistical investigations thread. These concepts have been identified as being of particular importance and the specific concepts that were focused on in this research are coloured red in Figure 9-4.

### ***Concept development***

Throughout this research a constant theme has been the development of the identified concepts to support the particular problematic situation under investigation, from posing investigative questions through making the call to describing distributions. Prototypical instructional materials were created to support this concept development and connections were established across the four focus areas.

On reflection of the findings in this research, concept development was involved in every teaching experiment and in every new problematic situation. The conceptual foundations (Figure 9-4) had to be determined for each situation as the conceptual foundations for particular focus areas had not yet been mapped out, either in the literature or in text books. As each problematic situation was realised, concepts for development were decided and innovative teaching and learning materials were developed as the teacher and researcher envisioned the hypothetical learning trajectory (Bakker, 2004a; Schwartz et al., 2008; Wittman, 1998).

This focus on teaching the underpinning concepts in statistics was about changing the paradigm of how to teach statistics. In the past many of these concepts were not taught; students were expected to use the concepts in descriptions of distributions, for example, yet the students were not specifically taught these concepts. The change also involved shifting the emphasis from calculating statistics and graphing data to assisting students to reason from data. In a sense, teachers need to “look under the bonnet” and see just what is driving statistics at curriculum level 5 (ages 13–15), i.e. identify the concepts that underpin investigative questions, prediction/hypothesis, analysis and conclusion, and then deliberately teach these concepts, not leave it to chance or osmosis that the students will understand the concepts.



Throughout this thesis instructional materials (Bakker, 2004a; Brown, 1992; Cobb, 2000a, 2000b; Cobb & Gravemeijer, 2008; Gravemeijer, 1998; Hjalmarson & Lesh, 2008; Lesh & Kelly, 2000; Roth, 2005; Steffe & Thompson, 2000) have been used to engineer learning around the highlighted concepts. Table 9-1 summarises the teaching and learning instructional materials and activities that were used for the different concepts.

**Table 9-1. Concepts: instructional materials and activities**

| Concept                                                                                                                | Specific learning materials/activities produced                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Variable                                                                                                               | Teaching experiment 1 – data cards were used to introduce the data sets for the teaching experiments. Variables are numbers with a context (Moore, 1990) and when the students were given the data cards to use, instead of the teacher telling the students what the variables were, the students had to “unpack” the variables by exploring the responses on the data cards alongside the survey questions that were used. A key aspect of this was that the students had been involved in collecting similar data themselves previously and used this knowledge as well to help them to decide what the different variables were. This activity supported students to think about “cleaning” the data, too, and what were sensible responses and what were not. This process was continued throughout the four teaching experiments with any new data set the students were given; for example, the Karekare College data cards. |
| Posing investigative questions<br>See Adaptions to teaching and learning activities – teaching experiment 3 (page 100) | Teaching experiment 3 – introduced an activity where students critiqued investigative questions that had been posed by others. The purpose of this activity was to elicit criteria from students as to what made a good investigative question. The concept of population (see next concept) to support posing of investigative questions was introduced.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Population<br>See Identifying and clarifying the population (page 151)                                                 | Teaching experiment 3 – introduced the population “Karekare College”. This population was created to help students to visualise the whole population and to stimulate the concept of sample. The population is big enough that there are too many to use all members for graphs or calculations (by hand), but small enough that it can be “made up” using data cards. The population is contained within a plastic bag and this plastic bag contains all the people in the population of interest.                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Sample<br>See What is a sample and why sample? (page 153)                                                              | Teaching experiment 3 – students were introduced to the idea of sample through needing to answer an investigative question about the Karekare College population. In making a dot plot using data cards, the students got to a point of saturation – either they were tired of placing cards, or they were seeing no changes – where they felt what they had represented the whole population. They recognised that a sample would be sufficient.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Sampling variability<br>See Exploring sampling variability (page 153)                                                  | Teaching experiment 3 – a number of activities worked on building the idea of sampling variability, starting with the activity described for sample above. Students observed other graphs made by students in their class and looked for similarities and differences across the samples. They also explored sampling variability further in the instructional material developed for making the call.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

| Concept                                                                                           | Specific learning materials/activities produced                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Making the call<br><br>See Informal inferential reasoning (pages 153–157)                         | Teaching experiment 3 – an extensive instructional sequence was developed to build the concept of making the call (informal inferential reasoning). The materials and activities allowed the students to develop criteria (the decision rule) for making the call for two different situations, one where they could make the call and one where they were unable to make the call. The Karekare College data set was used throughout this sequence. This data set does develop the concepts that are required and for that reason becomes a special data set and prototypical instructional material. There is no guarantee that other data sets will develop the same concepts in the same way. |
| Distributional shape<br><br>See Adaptions to teaching and learning activities (pages 201–204)     | Teaching experiment 4 – the teaching and learning materials created to support the development of the concept of distributional shape started with large samples to support students visualising the smooth or inferred shape. Through a carefully structured sequence of activities, students established that both symmetry and modality were important in describing distributional shape.                                                                                                                                                                                                                                                                                                     |
| Predicting distributions<br><br>See Adaptions to teaching and learning activities (pages 201–204) | Teaching experiment 4 – as part of the overall development of the concept of distribution and building contextual knowledge students engaged in instructional activities that required them to predict distributions of a number of different variables. This activity required students to use contextual knowledge and other prior experiences. Predicting distributions is linked to posing investigative questions.                                                                                                                                                                                                                                                                           |
| Describing distributions<br><br>See Adaptions to teaching and learning activities (pages 201–204) | Teaching experiment 4 – while describing distributions as an overall concept was not fully realised until after the retrospective analysis, in the teaching and learning materials the use of active reflection and modelling thinking was a key aspect of the teacher actions around describing distributions. In addition, the students were given many examples to practise describing distributions and the students practised this daily.                                                                                                                                                                                                                                                    |

Concept development is a relatively new idea in the teaching of statistics. In doing this research it was noticed that what appeared simple was not necessarily so. A number of ideas needed to be crystallised including identifying what concepts were needed to build on previously taught skills. Teachers were required to teach concepts that they had not necessarily previously taught and this may have been because they felt that there was “nothing to teach, it’s completely obvious, how could anyone *not* get it?!” (Wild, 2006, p. 23), or they were unaware of the rich conceptual repertoire underpinning reasoning from seemingly simple data plots.

In addition to building concepts, clarification and crystallisation of specific concepts was also an output of the research. The research developed criteria for what makes a good investigative question at the school level and also provided a suggested classification system for summary and comparison investigative questions. Alongside this classification of questions sits a clarification of what a good population descriptor would look like. All of this

knowledge about what makes a good investigative question has helped to unpack the concepts underpinning posing investigative questions.

Another outcome of this research, in conjunction with the Teaching and Learning Research Initiative (TLRI) project (Pfannkuch et al., 2011), has been a teaching and learning sequence with instructional material to develop the concept of making a call. Developing informal inferential reasoning at curriculum level 5 (ages 13–15) is a building block towards formal inferential reasoning. Understanding, developing and articulating the concepts of population, sample, sampling variability and sample-to-population inference is foundational.

Distribution as an overarching concept, with its many underlying concepts, is a multi-faceted notion (Bakker, 2004a; Ben-Zvi & Amir, 2005; Konold et al., 2004; McClain, 2005; Pfannkuch, 2006; Reading & Reid, 2006). As a result of this research a possible distribution description framework was created to pull together all of the characteristics and features of distribution, some of which were identified in the literature and some of which were identified as new. These characteristics and features sit under five overarching statistical concepts: contextual knowledge, distributional, graph comprehension, variability, and signal and noise. Underlying concepts of the first three – contextual knowledge, distributional and graph comprehension – were explored more deeply in this research.

### ***Statistics learning and teaching***

**Modelling thinking:** Wild (2006) talks about “aspects of statistics that are so basic to the way we think in the subject that no one abstracts, enunciates and examines them” (p. 10), and while this statement is true for the concepts discussed in the previous section, it also links unequivocally to ideas around modelling thinking. The initiated have experience, allowing them to engage in statistical thinking; for example, considering which features might be worth describing in a particular distribution or making a call (or not) for a comparative investigation. To aid students’ development of conceptual ideas, teachers can support their students by actively and deliberately modelling their thinking. Modelling thinking can be done in several ways; for example, by active reflection. The process involves the teacher (and/or students) taking an investigative question, a description or a conclusion and actively reflecting on what was given. Active reflection means writing the investigative question, description or conclusion on the board (or similar) and making it correct and complete. It is through this process of modelling how to improve a response that students start to develop

their own statistical thinking processes. In addition, teachers can think out loud as they are considering a situation, be it posing an investigative question or making a call. Both active reflection and thinking out loud appear to support concept development and the development of statistical thinking routines.

In New Zealand a number of thinking prompts have been used to aid thinking in certain aspects of the statistical enquiry cycle. Since the early 2000s “I notice ...” has been used as a thinking prompt for describing distributions, “I wonder ...” has been extensively used to generate investigative questions (Shaughnessy, 1997), and “These data suggest ...” has been used as a thinking starter for conclusions. These three prompts align with three of the four focus areas, and a fourth thinking prompt, “I suspect ...”, is suggested for prediction/hypothesis, as shown in Figure 9-5.

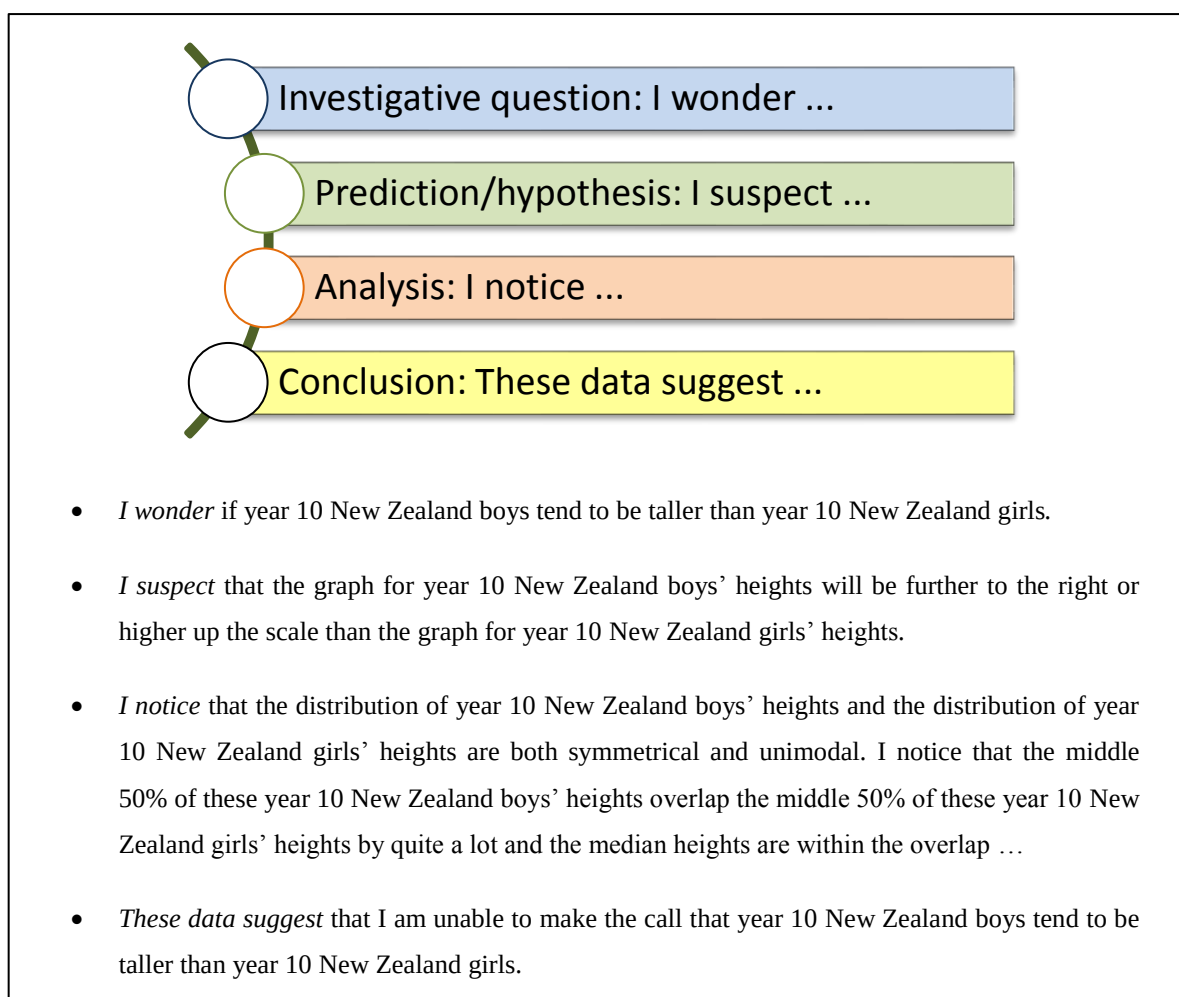


**Figure 9-5. Thinking prompts for the four focus areas, with examples**

The use of precise wording is also essential (Biehler, 1997; Pfannkuch, Wild, et al., 2009); teachers need to be modelling the statistical ideas correctly and this includes using the right

language and using it correctly. In some instances new language was “invented” to describe the new situations that arose; for example, the use of the term *overlap* in the description of box plots to describe the location of the boxes relative to one another and the position of the medians relative to the overlap.

Providing thinking prompts and modelling thinking through active reflection and thinking out loud are also an important part of the next big idea for teaching: building experiences faster.

**Building experiences faster:** Wild and Pfannkuch (1999) discuss statistical thinking and state that it “is the touchstone at the core of the statistician’s art” (p. 223). To get to the point of building experiences faster, it is necessary to quote Wild and Pfannkuch further from their introduction.

However, rather than being a precisely understood idea or set of ideas, the term “statistical thinking” is more like a mantra that evokes things understood at a vague, intuitive level, but largely unexamined. Statistical thinking is the statistical incarnation of “common sense”. “We know it when we see it”, or perhaps more truthfully its absence is often glaringly obvious. And, for most of us, it has been much more a product of experience, war stories and intuition than it is of any formal instruction that we have been through (p. 223).

Developing statistical thinking should not be left to experience alone. Alongside concept development and useful thinking prompts, educators can enculturate students into a statistical way of thinking. Similarly, posing investigative questions and reasoning inferentially should not be left to experience in a haphazard manner; rather educators need to be proactively building experiences faster. So what does this mean in terms of this research? It means building a structured set of experiences deliberately, modelling thinking, making connections explicit, using the technical language of statistics consistently and systematically, and building a context library.

Educators need to be deliberate about teaching the concepts that underpin posing investigative questions, that underpin making the call, and that underpin describing distributions. In addition to being deliberate about teaching the concepts, prototypical instructional materials were needed. These instructional materials were designed to build conceptual understanding, so it was deliberately choosing instructional materials that build conceptual understanding as well. Through being deliberate about teaching the concepts and being deliberate about the instructional materials used, there was an attempt to build the

experiences faster for the students. It was not left to chance or osmosis that students will “get” the concept of sample, or population, or sampling variability.

Modelling thinking, including providing thinking prompts and using models such as the PPDAC cycle, seems to support students to reason about statistical situations. When teachers think out loud, students are exposed to more experiences than just their own. In this way students build experiences faster because they are now exposed to the teacher’s experiences as well as actively reflecting on other students’ experiences.

The PPDAC cycle provided a model to help students and teachers undertake a statistical investigation. It was the interconnectedness of the PPDAC cycle, the moving backwards and forwards through the cycle, that provided rich experiences for students to support the development of statistical thinking. Making connections between concept development and how the concepts are realised within the PPDAC cycle also supported the building of experiences faster and supported concept development by providing a purpose. The “surfacing” of different concepts within the teaching and learning programme was motivated by a need to understand a part of the PPDAC cycle. Once the concept was developed it was reconnected back into the PPDAC cycle by establishing its role within a statistical investigation.

Building students’ intuitive ideas about statistical concepts is part of the process of developing statistical thinking. Within this process students use their own language to describe what they are seeing. Often the descriptors the students use vary from student to student and, if left unchecked, can cause confusion later on. Once an idea is established (for example, the idea that there are a group of graphs that look the same), it is useful to get agreement that the various terms such as “flat top” and “even” are referring to the same types of graphs, in this case uniform graphs. The movement from student language to statistical language is a necessary step in the process of developing statistical thinking; the term uniform has a specialised meaning in statistics and students should be moving to using statistical terminology with understanding as soon as possible. This is not to say that they should be given the statistical terms for the shapes of graphs without first developing the concept of distributional shape. Students should find the patterns across many graphs, label them first with their own descriptors, and then come to a shared understanding of the different types and finally to the statistical terms.

Contextual knowledge is a key element in building experiences faster. The investigative question defines the context, but understanding of the context is critical in posing investigative questions. For students to predict what they think the distribution of the variable might look like, they need to have contextual knowledge of the same variable in different populations, similar variables in similar populations, and some knowledge about distributions for different variables. When describing distributions students need to talk about the data in context because data are numbers with a context, to paraphrase Moore (1990). Understanding the context that they are exploring helps students to not only describe in context but to think about interpretations and explanations for what they are seeing and to “sound warning bells if distributions look unusual” (Watson, 2006, p. 188). Deep contextual knowledge comes from experiences – experiences with similar data, and experiences in life. One way of building contextual knowledge faster is through building a context library with students (Figure 8-13). This allows for connections between the types of variables that have right-skewed distributions or bell-shaped distributions, for example, to be made at an earlier stage than experience alone allows. In this way there is a deliberate building of students’ contextual knowledge base, i.e. building experiences faster.

**Visual imagery:** The final aspect of statistics teaching and learning that is realised in this research is the use of visual imagery to support conceptual development. During the planning and preparation phases of each teaching experiment, instruction materials were developed or sourced that used visual imagery. Visual imagery promotes visual thinking, and encouraging students to generate mental images enhances their learning (Arnold et al., 2011; Clark & Paivio, 1991). Examples using visual imagery in the research include: the use of dynamic representations to focus students’ attention on the properties and structure of sampling variability (Wild et al., 2011), building visual images of distribution by looking for and establishing the patterns in the shapes of distribution (Figure 8-10), and establishing the visual image of a population distribution (Figure 8-3). Also linked to visualising the population distribution is students being able to visualise the predicted outcome for a comparative investigative question. For example, the investigative question “Do year 10 New Zealand boys tend to be taller than year 10 New Zealand girls?” should prompt the visual image of two population distributions with the boys’ distribution being further to the right, indicating that the investigative question poser suspects that year 10 New Zealand boys tend to be taller than year 10 New Zealand girls.

Students may have had the idea that samples vary, but the instruction materials deliberately aimed to establish visual images to support that thinking. This visual image is reinforced by the dynamic representations and is often realised physically by students moving their hands back and forth (Figure 7-8) to show sampling variability. Students may have known the word distribution, but now they experience visually the distributional shape as a representation of distribution. Students did know the word population and thought of it as a number; they now seem to have the visual image of the population bag as representing the whole population and many students also “see” the population distribution. Most of the students are able to pose comparative investigative questions and now also seem to have a visual image to support their investigation. These visual representations, which are realised either physically through using hands to show the distribution in the air or to show sampling variability, or as a visual image in the student’s head, are important in concept development (Clark & Paivio, 1991; Presmeg, 2006).

### ***Building knowledge about student learning – specific frameworks derived from research***

Design research involves intervention, researching the intervention, and developing theories based on findings from the intervention. As a result of this research thesis, multiple frameworks or categorising structures have been developed and these were discussed in the results chapters 6, 7 and 8. To summarise:

1. Criteria for posing investigative questions were developed: Figure 6-3.
2. Hierarchical categories for summary and comparison investigative questions were developed that are cognisant of intent, variable and population: Figure 6-5 and Figure 6-8.
3. Population description categories to support posing investigative questions were defined: Figure 6-10.
4. A framework for describing distributions at curriculum level 5 (ages 13–15) was proposed: Figure 8-16.
5. A teaching and learning sequence for making the call was developed: Table 7-2, lessons 12–15.
6. Makar and Rubin’s (2009) framework for thinking about statistical inference was adapted and modified for sampling reasoning: Figure 7-2.
7. Criteria based on the SOLO taxonomy were created and used to grade student responses in pre- and post-tests: Figure 6-16, Figure 7-16 and Figure 8-21. These



criteria also provide support for teaching and learning of the related concepts as the criteria suggest the next steps when a specific blockage is reached by students.

### **9.1.2. Connections to original thesis focus**

This research thesis started by looking at statistical content knowledge needed by teachers to meet the requirements of the statistical investigation thread in the new New Zealand curriculum at level 5 (ages 13–15). From the initial problematic situation around posing investigative questions, through making the call and on to describing distributions, the research has come full circle. On reflection it was not possible to develop the statistical content knowledge needed by teachers because the relevant statistical context knowledge was not known, the concepts were not explicated, and students' reasoning processes were not identified for the new curriculum. However, the findings from this research have afforded an identification of the statistical content knowledge needed by teachers. The findings of the research and the overarching themes all combine to give detail to the statistics teaching cycle (adapted from Simon, 1995) at curriculum level 5 (ages 13–15). Figure 9-6 (next page) shows the connection between the statistics teaching cycle and the findings from this research thesis.

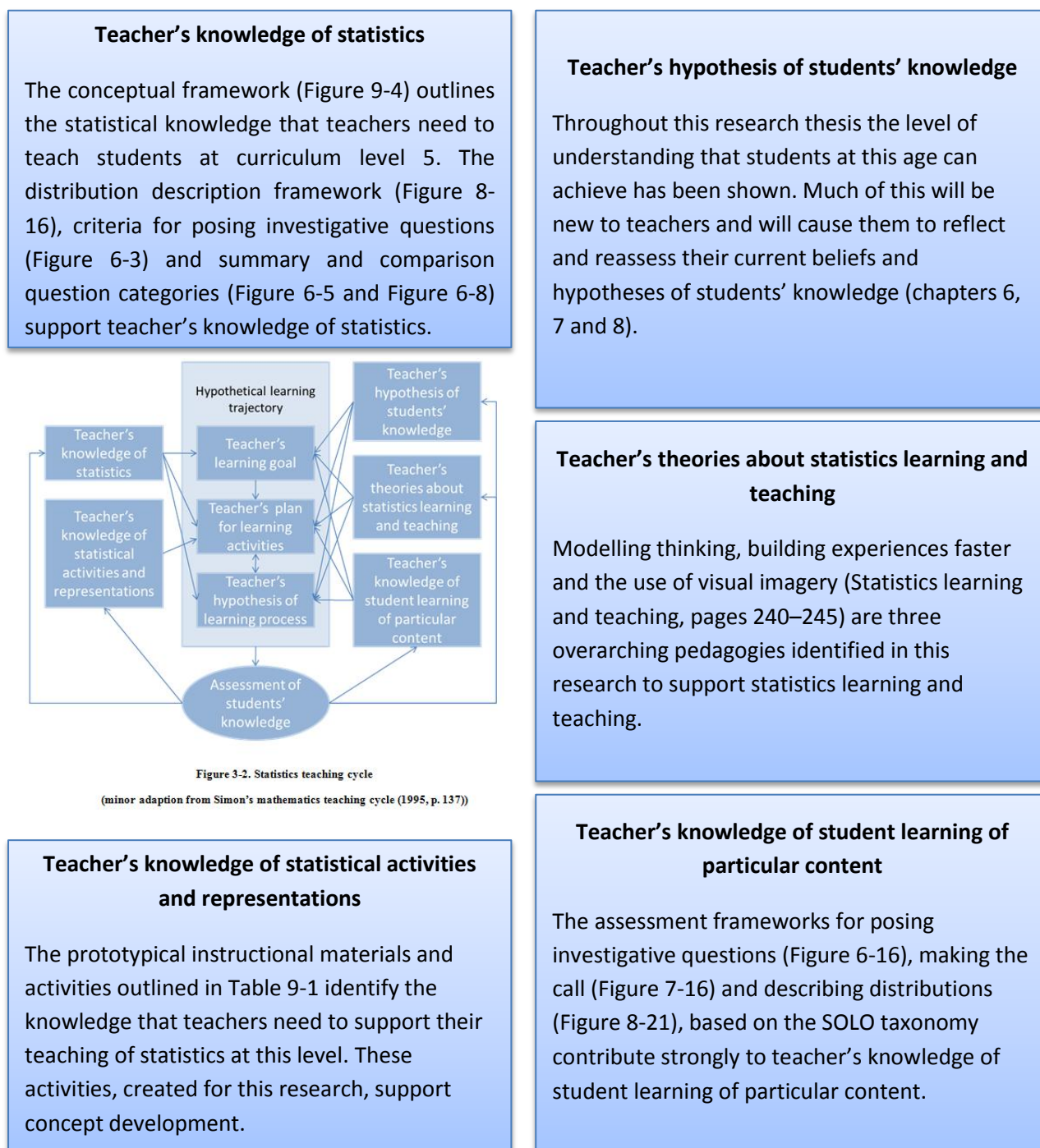


**Figure 9-6. Connecting research outcomes to the statistics teaching cycle**

Therefore the journey into posing investigative questions appears to have identified and described the statistical content knowledge needed by teachers and the journey into the development of this knowledge in teachers can now be an avenue of further research.

## **9.2. Implications for policy and practice**

This research has identified the gaps and consequently the big concepts needed for teaching and learning statistics at curriculum level 5 (ages 13–15) in New Zealand. New Zealand is one of the world leaders in statistics education and the New Zealand statistics curriculum reflects the needs of society in the 21st century. The concepts identified in this research would also be applicable to other countries at the appropriate level with similar curricula.

### **9.2.1. What has already happened**

This research into posing investigative questions has already had a huge impact in New Zealand classrooms and not just at year 10, curriculum level 5 (ages 13–15). Posing investigative questions are a key aspect of many of the statistics achievement standards in the national assessments and the term investigative question is now widely used. Criteria for what makes a good investigative question, along with summary and comparison question categories, are already available online as a support for teachers.

The TLRI research (Pfannkuch et al., 2011) into making the call and the associated teaching and learning sequence that was developed was trialled over three years and was shared extensively with teachers throughout New Zealand. The material was used in workshops with more than 400 teachers and is available online through the CensusAtSchool site for teachers to access from anywhere in New Zealand and the world. The prototypical instruction materials and activities that were developed as part of this research thesis are being used in classrooms across the country. It continues to be an ongoing challenge to provide opportunities for teachers to engage with the material as learners first and to experience the activities and then to discuss the reasons for the development of the activities and materials and what the underpinning concepts are that the activities are supporting.

### **9.2.2. What needs to happen**

The research for this thesis was working with the new statistics curriculum in New Zealand and the researcher has developed a conceptual framework specifically for the concept development needs for students in curriculum level 5 (ages 13–15). The findings from this research need to be shared with teachers. In particular, teachers need to have the opportunity to experience the new teaching and learning material in order to support their understandings of the research findings, before they take the material into their classrooms to use with their students. The sharing of the findings can support the teachers in the same way as it is hoped

they will help their students – by building experiences faster, by modelling thinking, and by using visual imagery. Using the overarching themes as the way of working with teachers will help in the same way it helps with students.

There is urgency to upskill teachers in this area as they lack knowledge of the conceptual foundations at this level. Many mathematics and statistics teachers are mathematics – not statistics – trained, or trained years ago. Either way, the statistics of today is not the statistics of their schooling or university days. It requires new knowledge and new ways of thinking. It also requires new ways of teaching, from a focus on the skills and calculations of the old statistics curriculum to a focus on the statistical reasoning and thinking of the new statistics curriculum.

Professional learning and development needs to be cognisant of the new statistics curriculum and the implications of the findings from this thesis. There are a number of instructional materials and activities that need to become common place within level 5 (ages 13–15) statistics classrooms as students are learning the conceptual foundations of the discipline. These conceptual foundations need to be started at this level in the curriculum as they will provide solid building blocks for higher-level statistical concepts that students need as they move up through the curriculum levels.

### **9.3. Looking ahead – suggestions for further research**

In the literature review summary (section 2.7) a number of research areas were highlighted as being shallow in their depth of coverage. These included the investigative question, interrogating the enquiry cycle, students asking analysis questions, the link between the investigative question and the analysis, the link between the investigative question and the conclusion, and the nature of data for which there is no consensus on definitions and classifications.

The first problematic situation addressed the investigative question, including the underpinning concepts that are needed to support its teaching and learning (see chapter 6). Aspects of the link between the investigative question and the analysis (see chapter 8) and conclusion (see chapter 7 and chapter 8) have also been addressed. Coincidentally, but not deliberately, aspects of interrogating the enquiry cycle have arisen within the research; for example, students were using the seek stage of the interrogative cycle when they were seeking information about the situation, or querying the data in hand, and they were using the

criticise stage when making the call, asking questions such as “Is this right?”, “Does this make sense?” (Wild & Pfannkuch, 1999, p. 232). Further work on interrogating the enquiry cycle is needed to find out what aspects should be a focus at curriculum level 5 (ages 13–15) or other curriculum levels. Another suggestion for future research would be to explore students asking analysis questions. For example, what thinking prompts do students need to have when they are making the call or describing distributions? This would have strong connections with conceptual development, modelling thinking and building experiences faster.

The decision was made to focus on summary and comparison investigative questions in this research thesis. Therefore, two further areas of research would be posing relationship and time-series investigative questions. Another area for possible research would be to look at what differences there might be between posing investigative questions from given data sets (secondary data), as was the case in this research, and posing investigative questions when the data is yet to be collected (primary data).

In chapter 6 a third finding from the retrospective analysis was a teaching and learning framework showing the connection between the different types of knowledge needed at the various stages of posing an investigative question (Arnold, 2009) from an inkling through to a precise question (Wild & Pfannkuch, 1999). These different types of knowledge and associated framework were not discussed in this thesis but are identified as an area for further analysis and research.

The distribution description framework (Figure 8-16) was developed following the retrospective analysis of the final teaching experiment. The framework would benefit from being used in classrooms to see if there are any additional features that can be realised at this level, in other words to reach saturation. This would also provide an opportunity to develop instructional material that aimed to introduce the different features of distribution that are relevant at this level. The distribution description framework is only developed for curriculum level 5 (ages 13–15) and further research could focus on developing the framework to include levels 1–4 (ages 5–13) below, and levels 6–8 (ages 15–18) above. Furthermore, the distribution description framework only addresses summary and comparison situations; it does not address relationship or time-series situations, both of which also have distributions that need describing, and both of which need urgent development. Linked to the

distribution description framework is the need for a more in-depth look at spread and other features of the distribution description framework that have not yet been researched fully.

#### **9.4. Limitations**

The findings from this research material may be limited in their applicability for the following reasons. Firstly, the research involved only one researcher working with the one teacher in the school and while other teachers had access to the material they did not have the benefit of the in-depth clarification discussions with the researcher which may be important for general applicability. Secondly, the students' data that were analysed were limited to those in the two teachers' classes. Data from other students in the school may have realised different findings. Thirdly, only two schools were involved and while these were both multicultural and mid-decile, a reflection of typical Auckland schools, they do not necessarily reflect New Zealand schools. One of the schools was also a single-sex (girls only) school. Two of the four classes were above-average ability and the other two of average to below-average ability. Fourthly, the technology was limited to the use of one computer, the teacher's, and a data projector for most of the time, a situation that is typical for New Zealand year 10 (ages 14–15) classes. For a few lessons the students were in the computer room, but they did not have daily access to the computer. Fifthly, both of the teachers working with the researcher had very similar philosophical positions as the researcher. Sixthly, the analysis was undertaken by the researcher only and while multiple sources of data were used to triangulate, there was no independent analysis of the data.

Any conclusions drawn from these findings are limited by the reasons given above. There is no guarantee that the same outcomes would occur if the material was used by another teacher in a similar or different situation. A mitigating factor is that the researcher has extensive experience in this field, more than 16 years' teaching experience at the level researched and seven years working with and observing teachers in their classrooms at the same level.

#### **9.5. Conclusion**

This enquiry into posing and answering questions from existing data spanned four teaching experiments and five years. The initial problematic situation, arising out of the first teaching experiment, was identified as: What makes a good investigative question? The second teaching experiment focused more deeply on this problem and subsequent retrospective analysis from the second teaching experiment gave rise to the second problematic situation

which looked at making the call in comparison situations. The third teaching experiment involved the development of a teaching and learning sequence, using innovative prototypical instruction material, to address the second problematic situation. From the ensuing retrospective analysis a third problematic situation arose around describing distributions and this was addressed in the fourth and final teaching experiment. These three problematic situations all contributed towards the identification of key concepts for statistical investigations at level 5 (ages 13–15) of the New Zealand curriculum and motivated the need to pay more attention to the conceptual development of these key concepts.

Enculturating students into new ways of thinking statistically requires a paradigm shift for teachers and students as they move away from the calculating and graph drawing of the past into the statistical thinking and reasoning of the future. Explication of the conceptual infrastructure students need was a major outcome of the research, which can be used to support teachers and curriculum developers as they navigate through this new paradigm. Frameworks to describe concepts were developed and can be used to support concept development; for example, criteria for what makes a good investigative question, hierarchical categories for summary and comparison investigative questions, population description categories, and a framework for describing distributions. In addition a number of frameworks to assess the level of student reasoning for investigative questions, making the call and describing distributions have also been developed as a result of this research.

The conceptual infrastructure and supporting frameworks should be useful tools for teachers as they navigate their way through statistical thinking and reasoning in the 21st century. The new material developed through this research has the potential to support posing and answering questions from existing data alongside developing statistical thinking and reasoning.

## Publications associated with this research

- Arnold, P. (2008a). Developing new statistical content knowledge with secondary school mathematics teachers. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference (July, 2008), Monterrey, Mexico* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Arnold, P. (2008b). *What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation*. Paper presented at the 11th International Congress on Mathematical Education (ICME-11, July, 2008) Monterrey, Mexico. Retrieved from <http://tsg.icme11.org/document/get/481>
- Arnold, P. (2009). Context and its role in posing investigative questions. In K. Makar (Ed.), *The Role of Context and Evidence in Informal Inferential Reasoning. Proceedings of the Sixth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-6, July, 2009), Brisbane, Australia* (pp. 101–115). Brisbane, Australia: University of Queensland.
- Arnold, P., & Pfannkuch, M. (2010). Enhancing students' inferential reasoning: from hands on to "movie snapshots". In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8, July, 2010), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Arnold, P., & Pfannkuch, M. (2012). The language of shape. *Proceedings of the 12th International Congress on Mathematical Education (ICME-12, July, 2012), Seoul, South Korea* (pp. 2446–2455). Seoul, South Korea: ICME-12.
- Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education*, 19(2), 1–32.
- Pfannkuch, M., Arnold, P., & Wild, C. J. (2011). *Statistics: It's reasoning not calculating*. (Summary research report on Building students' inferential reasoning: Levels 5 and 6) Retrieved from <http://www.tlri.org.nz/tlri-research/research-completed/school-sector/building-students-inferential-reasoning-statistics>
- Pfannkuch, M., Arnold, P., & Wild, C. J. (2012). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics* (accepted).
- Pfannkuch, M., Regan, M., & Arnold, P. (2009). *NCEA 2011: Laying foundations for inference in 2010*. Paper presented at the Statistics Teachers Day, Auckland, New Zealand. Retrieved from <http://new.censusatschool.org.nz/resource/laying-the-foundations-for-inference/>



## **Appendix A – Statistical investigations achievement objectives**

*Source:* Ministry of Education (2007)

Level 1: Conduct investigations using the statistical enquiry cycle: posing and answering questions; gathering, sorting and counting, and displaying category data; discussing the results (p. 47).

Level 2: Conduct investigations using the statistical enquiry cycle: posing and answering questions; gathering, sorting, and displaying category and whole-number data; communicating findings based on the data (p. 49).

Level 3: Conduct investigations using the statistical enquiry cycle: gathering, sorting, and displaying multivariate category and whole-number data and simple time-series data to answer questions; identifying patterns and trends in context, within and between data sets; communicating findings, using data displays (p. 51).

Level 4: Plan and conduct investigations using the statistical enquiry cycle: determining appropriate variables and data collection methods; gathering, sorting, and displaying multivariate category, measurement, and time-series data to detect patterns, variations, relationships, and trends; comparing distributions visually; communicating findings, using appropriate displays (p. 53).

Level 5: Plan and conduct surveys and experiments using the statistical enquiry cycle: determining appropriate variables and measures; considering sources of variation; gathering and cleaning data; using multiple displays, and re-categorising data to find patterns, variations, relationships, and trends in multivariate data sets; comparing sample distributions visually, using measures of centre, spread, and proportion; presenting a report of findings (p. 55).

Level 6: Plan and conduct investigations using the statistical enquiry cycle: justifying the variables and measures used; managing sources of variation, including through the use of random sampling; identifying and communicating features in context (trends, relationships between variables, and differences within and between distributions), using multiple displays; making informal inferences about populations from sample data; justifying findings, using displays and measures (p. 57).

## Appendices

Level 7: Carry out investigations of phenomena, using the statistical enquiry cycle: conducting surveys that require random sampling techniques, conducting experiments, and using existing data sets; evaluating the choice of measures for variables and the sampling and data collection methods used; using relevant contextual knowledge, exploratory data analysis, and statistical inference.

Make inferences from surveys and experiments: making informal predictions, interpolations, and extrapolations; using sample statistics to make point estimates of population parameters; recognising the effect of sample size on the variability of an estimate (p. 59).

Level 8: Carry out investigations of phenomena, using the statistical enquiry cycle: conducting experiments using experimental design principles, conducting surveys, and using existing data sets; finding, using, and assessing appropriate models (including linear regression for bivariate data and additive models for time-series data), seeking explanations, and making predictions; using informed contextual knowledge, exploratory data analysis, and statistical inference; communicating findings and evaluating all stages of the cycle.

Make inferences from surveys and experiments: determining estimates and confidence intervals for means, proportions, and differences, recognising the relevance of the central limit theorem; using methods such as resampling or randomisation to assess the strength of evidence (p. 61).

## Appendix B – Pre- and Post-Tests

### B.1 Teaching experiment 1 – Pre-test 2007

#### STATISTICAL INVESTIGATIONS PRE-TEST

Name: \_\_\_\_\_

The data below was randomly selected from the 2005 CensusAtSchool database. The variables are:

Gender: boy or girl

Age: in years

Region: they live in New Zealand

Innz: Years they have lived in New Zealand

Ethnic: ethnicity

Height: in cm

RF: right foot length in cm

Timetravel: time it takes to travel to school in minutes

Actlunch: the main activity they did at lunchtime

Year: year level

| gender | age | region               | innz | ethnic | height | RF | timetravel | actlunch | year |
|--------|-----|----------------------|------|--------|--------|----|------------|----------|------|
| boy    | 14  | Wellington Region    | 14   | Asian  | 154    | 27 | 10to20     | walk     | 9    |
| girl   | 10  | Bay of Plenty Region | 10   | NZEuro | 144    | 21 | less10     | ran      | 6    |
| boy    | 11  | Wellington Region    | 11   | NZEuro | 144    | 24 | 10to20     | ran      | 6    |
| boy    | 12  | Auckland Region      | 2    | Asian  | 171    | 24 | 10to20     | ran      | 8    |
| girl   | 10  | Manawatu-Wanganui    | 10   | NZEuro | 133    | 19 | 10to20     | ran      | 6    |
| girl   | 13  | Wellington Region    | 13   | NZEuro | 159    | 21 | 20to30     | sat      | 9    |
| girl   | 12  | Otago Region         | 12   | NZEuro | 157    | 22 | less10     | ran      | 7    |
| girl   | 15  | Auckland Region      | 15   | Maori  | 163    | 23 | 20to30     | ran      | 10   |
| girl   | 10  | Canterbury Region    | 10   | NZEuro | 134    | 21 | less10     | ran      | 6    |
| boy    | 11  | Waikato Region       | 11   | NZEuro | 135    | 21 | 10to20     | ran      | 7    |
| girl   | 13  | Manawatu-Wanganui    | 14   | Maori  | 171    | 24 | 20to30     | walk     | 9    |
| boy    | 10  | Auckland Region      | 10   | Asian  | 138    | 22 | less10     | ran      | 6    |
| girl   | 14  | Manawatu-Wanganui    | 14   | Other  | 160    | 21 | 30plus     | sat      | 9    |
| girl   | 11  | Auckland Region      | 8    | NZEuro | 160    | 23 | 20to30     | walk     | 7    |

You will be exploring this data as part of your pre-test.

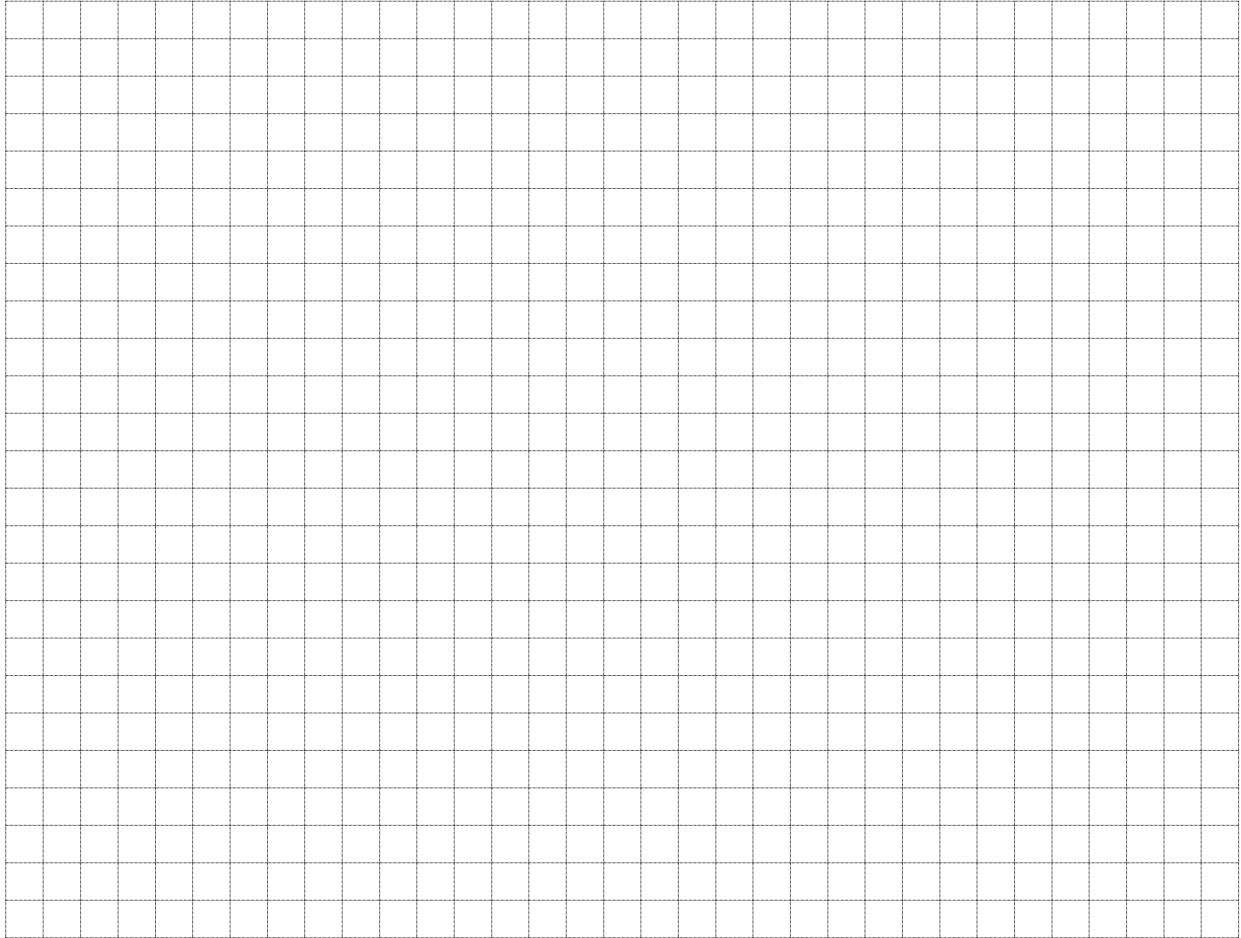
You need to think about what you could investigate using this data.

Write down one question you think you could ask or explore using this data.

1.

## Appendices

To help you answer your question, draw a graph to display the data from the table.



Make three statements about what your graph shows.

- 1.
- 2.
- 3.

Write down two more questions you think you could ask or explore using this data.

- 1.
- 2.

If you have time, draw a graph to display the data to answer one of the new questions and write two statements about the graph.

## B.2 Teaching experiment 1 – Post-test 2007

### STATISTICAL INVESTIGATIONS ASSESSMENT Name: \_\_\_\_\_

The data below was randomly selected from the 2005 Census at School database, all students are in Year 10.

The variables are:

Gender: boy or girl

Ethnic: ethnicity

RF: right foot length in cm

Wrist: wrist circumference in cm

Travel: how they travelled to school that day

Reaction: reaction time in secs

Age: in years

Height: in cm

arms: armspan in cm

lookup: the person they look up to/ role model

Timetravel: time it takes to travel to school in minutes

sleep: hours sleep last night

| gender | age | ethnic         | height | RF | arms | wrist | lookup       | travel | timetravel | reaction | sleep |
|--------|-----|----------------|--------|----|------|-------|--------------|--------|------------|----------|-------|
| boy    | 14  | Indian         | 170    | 26 | 178  | 17    | celeb        | train  | 30plus     | 0.46     | 5     |
| girl   | 15  | NZ European    | 174    | 26 | 165  | 17    | family       | bus    | 30plus     | 0.34     | 6     |
| boy    | 15  | NZ European    | 166    | 25 | 167  | 16    | family       | motor  | less10     | 0.24     | 7     |
| girl   | 14  | Asian          | 167    | 25 | 163  | 14.5  | family       | motor  | 10to20     | 0.36     | 8     |
| girl   | 15  | Other          | 175    | 25 | 174  | 15.5  | family       | bus    | 30plus     | 1.04     | 8     |
| boy    | 14  | NZ European    | 159    | 25 | 158  | 17    | family       | walk   | 10to20     | 0.74     | 10    |
| girl   | 14  | NZ European    | 169    | 25 | 150  | 15    | celeb        | bus    | 10to20     | 0.4      | 3     |
| girl   | 14  | NZ European    | 175    | 24 | 175  | 19    | sportsperson | bus    | 10to20     | 0.32     | 7     |
| girl   | 14  | Maori          | 165    | 20 | 165  | 16.4  | family       | bus    | 20to30     | 0.4      | 8.5   |
| girl   | 14  | NZ European    | 151    | 23 | 144  | 15.5  | family       | other  | less10     | 0.42     | 10    |
| boy    | 14  | Other          | 175    | 32 | 165  | 20    | celeb        | motor  | 20to30     | 0.34     | 7     |
| girl   | 14  | NZ European    | 167    | 22 | 162  | 14    | family       | motor  | less10     | 0.4      | 8.5   |
| girl   | 14  | NZ European    | 153    | 23 | 170  | 17    | sportsperson | motor  | 10to20     | 0.32     | 10    |
| girl   | 15  | Pacific Island | 163    | 23 | 158  | 16    | other        | motor  | less10     | 0.94     | 6.5   |
| girl   | 14  | NZ European    | 160    | 25 | 152  | 15    | sportsperson | walk   | less10     | 0.38     | 7.5   |
| boy    | 14  | NZ European    | 160    | 30 | 180  | 18    | family       | bike   | less10     | 0.46     | 2     |
| girl   | 14  | Pacific Island | 169    | 24 | 173  | 15.5  | family       | walk   | 10to20     | 0.4      | 10    |
| girl   | 14  | NZ European    | 164    | 24 | 170  | 14.5  | other        | bus    | 10to20     | 0.32     | 8.5   |
| girl   | 15  | Other          | 165    | 25 | 166  | 16.5  | family       | motor  | 10to20     | 2.58     | 8     |
| girl   | 14  | Other          | 164    | 25 | 161  | 15    | celeb        | bus    | less10     | 0.74     | 9     |
| girl   | 14  | NZ European    | 163    | 19 | 160  | 15    | family       | bike   | 20to30     | 0.7      | 10    |
| boy    | 15  | Indian         | 166    | 23 | 155  | 15    | business     | motor  | 20to30     | 0.76     | 2     |
| boy    | 14  | NZ European    | 145    | 25 | 151  | 14    | family       | walk   | less10     | 0.34     | 9     |
| boy    | 15  | NZ European    | 180    | 26 | 177  | 17    | family       | bus    | 30plus     | 0.4      | 8     |
| boy    | 15  | Other          | 175    | 27 | 182  | 17    | other        | bus    | 20to30     | 0.32     | 8     |
| boy    | 14  | NZ European    | 181    | 30 | 178  | 16.5  | celeb        | bus    | 20to30     | 0.6      | 6     |
| girl   | 14  | Asian          | 165    | 22 | 166  | 15.5  | family       | bus    | less10     | 0.48     | 8     |
| boy    | 14  | NZ European    | 178    | 31 | 165  | 14    | family       | motor  | 10to20     | 0.24     | 8.5   |
| girl   | 15  | NZ European    | 171    | 25 | 152  | 12    | celeb        | motor  | 10to20     | 0.36     | 9     |
| boy    | 14  | NZ European    | 168    | 25 | 164  | 18    | family       | bike   | 10to20     | 0.54     | 9     |

You will be exploring this data as part of your assessment.

You need to think about what you could investigate about this data.

Write down three summary questions (I wonder statements) you think you could ask or explore using this data.

Write down three comparison questions (I wonder statements) you think you could ask or explore using this data.

#### SUMMARY QUESTIONS

I wonder

I wonder

I wonder

#### COMPARISON QUESTIONS

I wonder

I wonder

I wonder

## Appendices

**STATISTICAL INVESTIGATIONS ASSESSMENT Name:** \_\_\_\_\_

Select one of your comparison questions to investigate for this assessment.

### **PROBLEM**

Write the question here: I wonder...

### **PLAN/DATA**

Which variables from the data sheet are you going to use to do this investigation?

Why have you selected these variables?

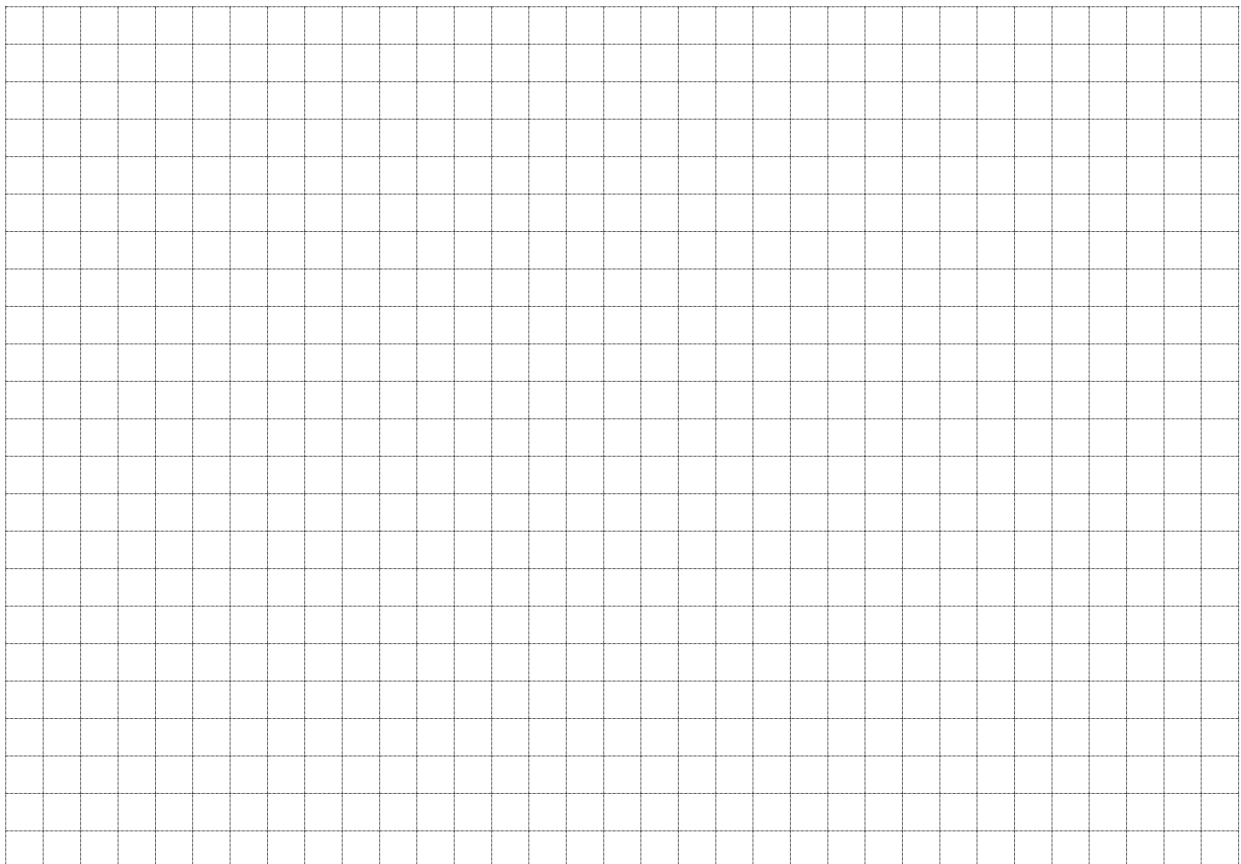
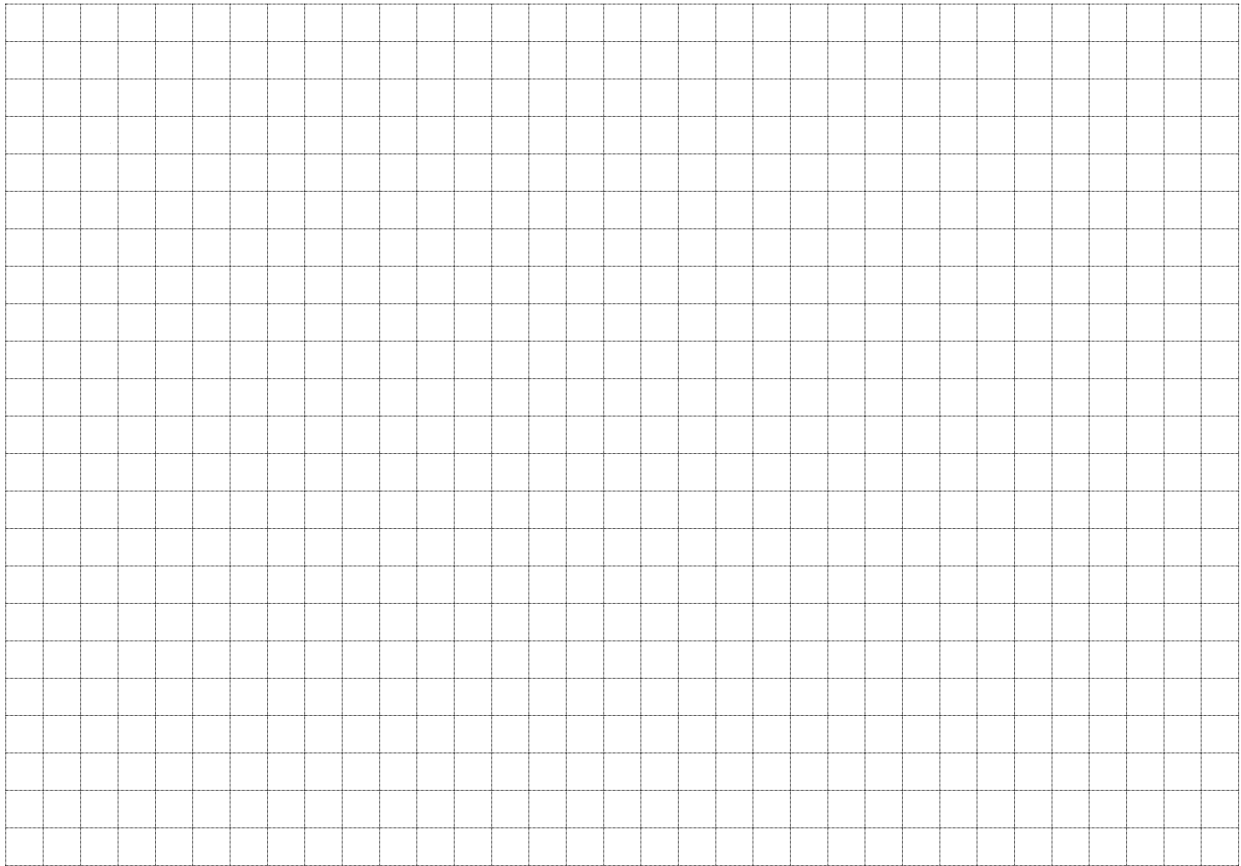
### **ANALYSIS**

Calculate your summary statistics below.

|                              |  |  |
|------------------------------|--|--|
|                              |  |  |
| Minimum                      |  |  |
| Maximum                      |  |  |
| Median                       |  |  |
| Upper Quartile               |  |  |
| Lower Quartile               |  |  |
| Range or Interquartile range |  |  |
| Mean                         |  |  |

## Appendices

You are expected to draw a dot plot and/or box and whisker plot below.



## Appendices

From your tables and/or plots make four I notice statements about the data you have displayed. You should compare the summary statistics, shape of the data, spread of the data and the middle 50% of the data.

I notice...

I notice...

I notice...

I notice...

## CONCLUSION

Your question you started with (from PROBLEM)...

Answer to your question...

Support for this answer.

What other questions has this investigation generated?

I wonder...

I wonder...



### B.3 Teaching experiment 2 – Pre-test 2008

#### STATISTICAL INVESTIGATIONS PRE-TEST DATA SHEET

The table below shows part of the data that was randomly selected from the 2007 NZ Census at School database. Altogether in the sample there are 254 Year 11 students. When the sample was selected only the “questions about you” variables were included. The variables and survey questions asked are:

| Variable            | Survey Question                                                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl | 1. Are you: male/female                                                                                                                                                |
| age: in years       | 2. How old are you?                                                                                                                                                    |
| handed:             | 3. Are you right-handed, left-handed or ambidextrous? (An ambidextrous person is able to use their right and left hands equally well.)                                 |
| languages:          | 4. In how many languages can you hold a conversation about a lot of everyday things?                                                                                   |
| height: in cm       | 5. How tall are you without your shoes on? Answer to the nearest centimetre.                                                                                           |
| right foot: in cm   | 6. What is the length of your right foot, without a shoe? Answer to the nearest centimetre.                                                                            |
| armspan: in cm      | 7. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.) |
| hairlength: in cm   | 8. What is the length of your hair? Answer to the nearest centimetre. (Pull one hair out from the back of your head and measure it.)                                   |

yr11aboutme.csv

|     | gender | age | handed | languages | height | rightfoot | armspan | hairlength |
|-----|--------|-----|--------|-----------|--------|-----------|---------|------------|
| 78  | girl   | 15  | right  | 1         | 164    | 25        | 182     | 43         |
| 79  | boy    | 15  | ambi   | 3         | 150    | 28        | 150     | 4          |
| 80  | girl   | 15  | right  | 1         | 168    | 22        | 171     | 23         |
| 81  | girl   | 16  | right  | 1         | 176    | 22        | 175     | 15         |
| 82  | girl   | 15  | right  | 4         | 153    | 25        | 20      | 30         |
| 83  | girl   | 15  | right  | 1         | 162    | 24        | 160     | 40         |
| 84  | girl   | 15  | right  | 1         | 157    | 27        | 155     | 31         |
| 85  | girl   | 15  | right  | 2         | 157    | 28        | 35      | 128        |
| 86  | girl   | 16  | right  | 2         | 159    | 25        | 169     | 44         |
| 87  | boy    | 16  | right  | 2         | 176    | 24        | 174     | 29         |
| 88  | boy    | 16  | right  | 2         | 186    | 34        | 200     | 4          |
| 89  | girl   | 15  | right  | 2         | 165    | 24        | 60      | 40         |
| 90  | girl   | 15  | right  | 2         | 162    | 26        | 174     | 44         |
| 91  | boy    | 16  | right  | 2         | 193    | 25        | 193     | 2          |
| 92  | girl   | 16  | left   | 1         | 165    | 25        | 165     | 29         |
| 93  | girl   | 15  | right  | 2         | 169    | 30        | 169     | 30         |
| 94  | boy    | 15  | right  | 1         | 164    | 27        | 75      | 7          |
| 95  | girl   | 15  | right  | 2         | 169    | 28        | 169     | 30         |
| 96  | girl   | 16  | right  | 1         | 175    | 24        | 168     | 20         |
| 97  | boy    | 15  | right  | 3         | 175    | 28        | 173     | 19         |
| 98  | boy    | 16  | right  | 1         | 170    | 20        | 150     | 6          |
| 99  | girl   | 15  | right  | 1         | 160    | 25        | 60      | 25         |
| 100 | girl   | 15  | right  | 2         | 168    | 25        | 165     | 46         |
| 101 | girl   | 15  | ambi   | 2         | 155    | 22        | 157     | 8          |

TABLE 1: Part of the data table (total number of students in the sample is 254)

## Appendices

### STATISTICAL INVESTIGATIONS PRE-TEST

Name: \_\_\_\_\_

#### TASK A

Your task is to pose questions about the 2007 NZ Census At School data using the variables given in the data sheet. You need to pose three summary type questions, three comparison type questions, and three relationship type questions.

##### Summary questions

*I wonder...*

*I wonder...*

*I wonder...*

##### Comparison questions

*I wonder...*

*I wonder...*

*I wonder...*

##### Relationship questions

*I wonder...*

*I wonder...*

*I wonder...*

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS PRE-TEST**

**Name:** \_\_\_\_\_

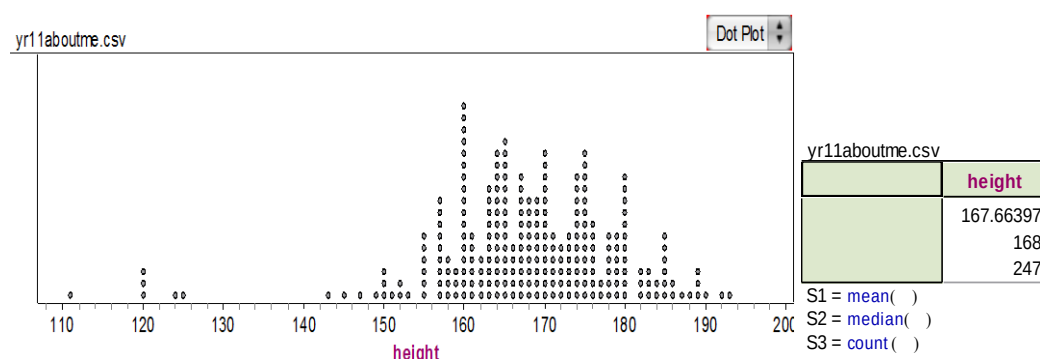
**TASK B**

Problem: I wonder what is the typical height of NZ Year 11 students?

Plan: These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on height. The question asked to get these data was: How tall are you without your shoes on? Answer to the nearest centimetre.

Data: Managed through NZ Census At School survey team.

Analysis:



GRAPH 1: Heights of Yr 11 students from NZ Census At School sample. TABLE 2: Statistics from the sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Shape:** I notice...

**Middle group:** I notice...

**Spread:** I notice...

**Anything unusual:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**

STATISTICAL INVESTIGATIONS PRE-TEST

Name: \_\_\_\_\_

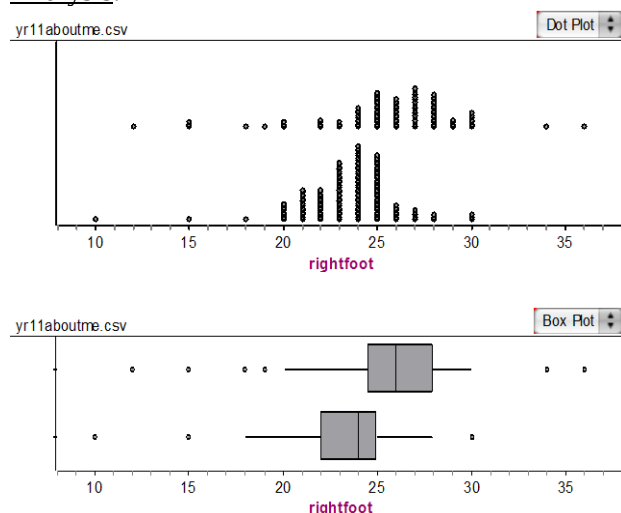
TASK C

**Problem:** I wonder if Year 11 NZ boys tend to have longer right foot lengths than Year 11 NZ girls?

**Plan:** These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on right foot lengths and involves gender. The questions asked to get these data were: Are you: male/female? What is the length of your right foot, without a shoe? Answer to the nearest centimetre.

**Data:** Managed through NZ Census At School survey team.

**Analysis:**



| yr11aboutme.csv |  | gender |           | Row<br>Summary |
|-----------------|--|--------|-----------|----------------|
|                 |  | boy    | girl      |                |
| rightfoot       |  | 25.67  | 23.524138 | 24.4           |
|                 |  | 100    | 145       | 245            |
|                 |  | 12     | 10        | 10             |
|                 |  | 24.5   | 22        | 23             |
|                 |  | 26     | 24        | 24             |
|                 |  | 28     | 25        | 26             |
|                 |  | 36     | 30        | 36             |
| S1 = mean ( )   |  |        |           |                |
| S2 = count ( )  |  |        |           |                |
| S3 = min ( )    |  |        |           |                |
| S4 = Q1 ( )     |  |        |           |                |
| S5 = median ( ) |  |        |           |                |
| S6 = Q3 ( )     |  |        |           |                |
| S7 = max ( )    |  |        |           |                |

GRAPHS 2&3: Right foot length of Year 11 students from NZ Census at school sample.

TABLE 3: Statistics from the sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Summary statistics:** I notice...

**Shape:** I notice...

**Spread:** I notice...

**Middle 50%:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**

**STATISTICAL INVESTIGATIONS PRE-TEST Name:** \_\_\_\_\_

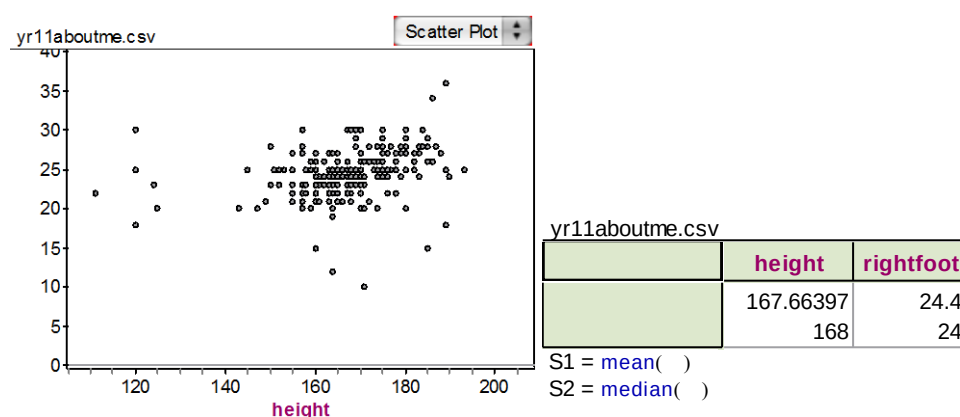
**TASK D**

Problem: I wonder if there is a relationship between the height of NZ Year 11 students and their right foot length?

Plan: These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on height and right foot lengths. The questions asked to get these data were: How tall are you without your shoes on? Answer to the nearest centimetre. What is the length of your right foot, without a shoe? Answer to the nearest centimetre.

Data: Managed through NZ Census At School survey team.

Analysis:



GRAPH 4: Heights and right foot length of Year 11 students from NZ Census at School sample.

TABLE 4: Statistics from the sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Trend (shape):** I notice...

**Trend (middle group):** I notice...

**Spread:** I notice...

**Anything unusual:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**

## B.4 Teaching experiment 2 – Post-test 2008

### STATISTICAL INVESTIGATIONS POST TEST DATA SHEET

The table below shows part of the data that was randomly selected from the 2007 NZ Census at School database. Altogether in the sample there are 254 Year 11 students. When the sample was selected only the “questions about you” variables were included. The variables and survey questions asked are:

| Variable            | Survey Question                                                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl | 1. Are you: male/female                                                                                                                                                |
| age: in years       | 2. How old are you?                                                                                                                                                    |
| handed:             | 3. Are you right-handed, left-handed or ambidextrous? (An ambidextrous person is able to use their right and left hands equally well.)                                 |
| languages:          | 4. In how many languages can you hold a conversation about a lot of everyday things?                                                                                   |
| height: in cm       | 5. How tall are you without your shoes on? Answer to the nearest centimetre.                                                                                           |
| right foot: in cm   | 6. What is the length of your right foot, without a shoe? Answer to the nearest centimetre.                                                                            |
| armspan: in cm      | 7. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.) |
| hairlength: in cm   | 8. What is the length of your hair? Answer to the nearest centimetre. (Pull one hair out from the back of your head and measure it.)                                   |

yr11aboutme.csv

|     | gender | age | handed | languages | height | rightfoot | armspan | hairlength |
|-----|--------|-----|--------|-----------|--------|-----------|---------|------------|
| 78  | girl   | 15  | right  | 1         | 164    | 25        | 182     | 43         |
| 79  | boy    | 15  | ambi   | 3         | 150    | 28        | 150     | 4          |
| 80  | girl   | 15  | right  | 1         | 168    | 22        | 171     | 23         |
| 81  | girl   | 16  | right  | 1         | 176    | 22        | 175     | 15         |
| 82  | girl   | 15  | right  | 4         | 153    | 25        | 20      | 30         |
| 83  | girl   | 15  | right  | 1         | 162    | 24        | 160     | 40         |
| 84  | girl   | 15  | right  | 1         | 157    | 27        | 155     | 31         |
| 85  | girl   | 15  | right  | 2         | 157    | 28        | 35      | 128        |
| 86  | girl   | 16  | right  | 2         | 159    | 25        | 169     | 44         |
| 87  | boy    | 16  | right  | 2         | 176    | 24        | 174     | 29         |
| 88  | boy    | 16  | right  | 2         | 186    | 34        | 200     | 4          |
| 89  | girl   | 15  | right  | 2         | 165    | 24        | 60      | 40         |
| 90  | girl   | 15  | right  | 2         | 162    | 26        | 174     | 44         |
| 91  | boy    | 16  | right  | 2         | 193    | 25        | 193     | 2          |
| 92  | girl   | 16  | left   | 1         | 165    | 25        | 165     | 29         |
| 93  | girl   | 15  | right  | 2         | 169    | 30        | 169     | 30         |
| 94  | boy    | 15  | right  | 1         | 164    | 27        | 75      | 7          |
| 95  | girl   | 15  | right  | 2         | 169    | 28        | 169     | 30         |
| 96  | girl   | 16  | right  | 1         | 175    | 24        | 168     | 20         |
| 97  | boy    | 15  | right  | 3         | 175    | 28        | 173     | 19         |
| 98  | boy    | 16  | right  | 1         | 170    | 20        | 150     | 6          |
| 99  | girl   | 15  | right  | 1         | 160    | 25        | 60      | 25         |
| 100 | girl   | 15  | right  | 2         | 168    | 25        | 165     | 46         |
| 101 | girl   | 15  | ambi   | 2         | 155    | 22        | 157     | 8          |

TABLE 1: Part of the data table (total number of students in the sample is 254)

## Appendices

### STATISTICAL INVESTIGATIONS POST TEST      Name: \_\_\_\_\_

#### TASK A

Your task is to pose questions about the 2007 NZ Census At School data using the variables given in the data sheet. You need to pose three summary type questions, three comparison type questions, and three relationship type questions.

##### Summary questions

*I wonder...*

*I wonder...*

*I wonder...*

##### Comparison questions

*I wonder...*

*I wonder...*

*I wonder...*

##### Relationship questions

*I wonder...*

*I wonder...*

*I wonder...*

**Hand this sheet in when finished and get the next task.**

## Appendices

### STATISTICAL INVESTIGATIONS POST TEST      Name: \_\_\_\_\_

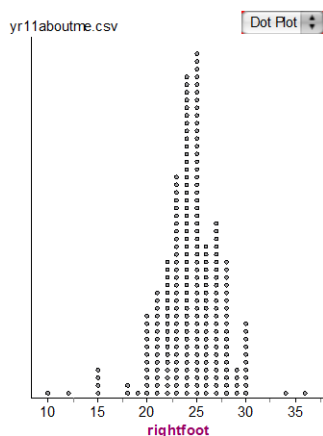
#### TASK B

Problem: I wonder what is the typical right foot length of NZ Year 11 students?

Plan: These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on right foot length. The question asked to get these data was: What is the length of your right foot, without a shoe? Answer to the nearest centimetre.

Data: Managed through NZ Census At School survey team.

Analysis:



GRAPH 1: Right foot length of Year 11 students from NZ.  
Census at School sample

| yr11aboutme.csv |           |
|-----------------|-----------|
|                 | rightfoot |
|                 | 24.4      |
|                 | 24        |
|                 | 245       |

S1 = mean( )

S2 = median( )

S3 = count( )

TABLE 2: Statistics from the sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Shape:** I notice...

**Middle group:** I notice...

**Spread:** I notice...

**Anything unusual:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**



**STATISTICAL INVESTIGATIONS POST-TEST Name:** \_\_\_\_\_

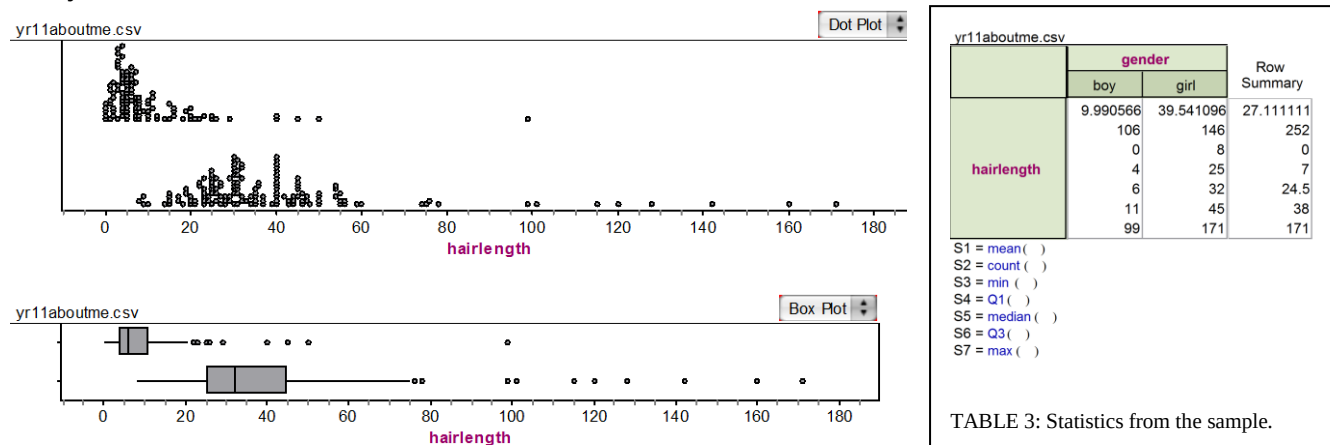
**TASK C**

Problem: I wonder if Year 11 NZ boys tend to have shorter hair than Year 11 NZ girls?

Plan: These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on hair length and involves gender. The questions asked to get these data were: Are you: male/female? What is the length of your hair? Answer to the nearest centimetre. (Pull one hair out from the back of your head and measure it.)

Data: Managed through NZ Census At School survey team.

Analysis:



GRAPHS 1 & 2: Hair length of Year 11 students from NZ Census at school sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Summary statistics:** I notice...

**Shape:** I notice...

**Spread:** I notice...

**Middle 50%:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**

**STATISTICAL INVESTIGATIONS POST TEST Name:** \_\_\_\_\_

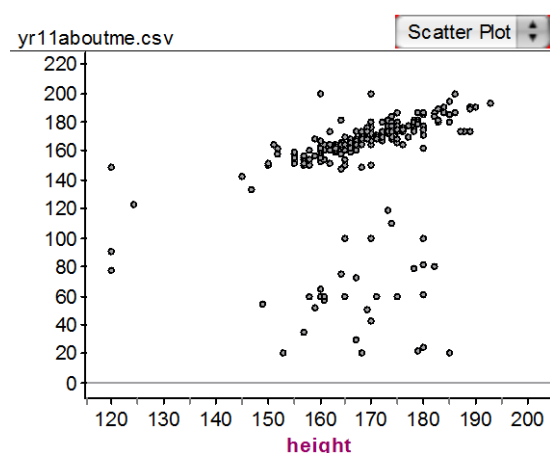
**TASK D**

Problem: I wonder if there is a relationship between the height of NZ Year 11 students and their armspan?

Plan: These data are from the 2007 NZ Census At School database. 254 Year 11 students were selected using the random sampler. As the sample is random it would be expected to be representative of all the Year 11 students in the database. The students are a mixture of boys and girls from Year 11. The problem focuses on height and right foot lengths. The questions asked to get these data were: How tall are you without your shoes on? Answer to the nearest centimetre. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.)

Data: Managed through NZ Census At School survey team.

Analysis:



| yr11aboutme.csv |           |           |
|-----------------|-----------|-----------|
|                 | height    | arm span  |
|                 | 167.66397 | 152.30453 |
|                 | 168       | 165       |

S1 = mean( )  
S2 = median( )

GRAPH 4: Heights and armspan of Year 11 students from NZ Census at school sample.

TABLE 4: Statistics from the sample.

**Make statements about what you notice about the graph/statistics. You should make statements under the headings given.**

**Shape:** I notice...

**Pattern:** I notice...

**Spread:** I notice...

**Anything unusual:** I notice...

**Conclusion:** Write a conclusion using the headings below.

**Answer to the problem:**

**Support for this answer:**

**Generalisation of the sample findings to the population:**

## B.5 Teaching experiment 3 – Pre-test 2009

### STATISTICAL INVESTIGATIONS PRETEST DATA SHEET

Table 1 shows part of a data set from the 2007 NZ Census at School. The data shown were collected from Year 11 students and refer to the following survey questions:

| Variable            | Survey Question                                                                                                                                                        |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl | 1. Are you: male/female                                                                                                                                                |
| age: in years       | 2. How old are you?                                                                                                                                                    |
| handed:             | 3. Are you right-handed, left-handed or ambidextrous? (An ambidextrous person is able to use their right and left hands equally well.)                                 |
| languages:          | 4. In how many languages can you hold a conversation about a lot of everyday things?                                                                                   |
| height: in cm       | 5. How tall are you without your shoes on? Answer to the nearest centimetre.                                                                                           |
| right foot: in cm   | 6. What is the length of your right foot, without a shoe? Answer to the nearest centimetre.                                                                            |
| armspan: in cm      | 7. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.) |
| hairlength: in cm   | 8. What is the length of your hair? Answer to the nearest centimetre. (Pull one hair out from the back of your head and measure it.)                                   |

|     | gender | age | handed | languages | height | rightfoot | armspan | hairlength |
|-----|--------|-----|--------|-----------|--------|-----------|---------|------------|
| 78  | girl   | 15  | right  | 1         | 164    | 25        | 182     | 43         |
| 79  | boy    | 15  | ambi   | 3         | 150    | 28        | 150     | 4          |
| 80  | girl   | 15  | right  | 1         | 168    | 22        | 171     | 23         |
| 81  | girl   | 16  | right  | 1         | 176    | 22        | 175     | 15         |
| 82  | girl   | 15  | right  | 4         | 153    | 25        | 20      | 30         |
| 83  | girl   | 15  | right  | 1         | 162    | 24        | 160     | 40         |
| 84  | girl   | 15  | right  | 1         | 157    | 27        | 155     | 31         |
| 85  | girl   | 15  | right  | 2         | 157    | 28        | 35      | 128        |
| 86  | girl   | 16  | right  | 2         | 159    | 25        | 169     | 44         |
| 87  | boy    | 16  | right  | 2         | 176    | 24        | 174     | 29         |
| 88  | boy    | 16  | right  | 2         | 186    | 34        | 200     | 4          |
| 89  | girl   | 15  | right  | 2         | 165    | 24        | 60      | 40         |
| 90  | girl   | 15  | right  | 2         | 162    | 26        | 174     | 44         |
| 91  | boy    | 16  | right  | 2         | 193    | 25        | 193     | 2          |
| 92  | girl   | 16  | left   | 1         | 165    | 25        | 165     | 29         |
| 93  | girl   | 15  | right  | 2         | 169    | 30        | 169     | 30         |
| 94  | boy    | 15  | right  | 1         | 164    | 27        | 75      | 7          |
| 95  | girl   | 15  | right  | 2         | 169    | 28        | 169     | 30         |
| 96  | girl   | 16  | right  | 1         | 175    | 24        | 168     | 20         |
| 97  | boy    | 15  | right  | 3         | 175    | 28        | 173     | 19         |
| 98  | boy    | 16  | right  | 1         | 170    | 20        | 150     | 6          |
| 99  | girl   | 15  | right  | 1         | 160    | 25        | 60      | 25         |
| 100 | girl   | 15  | right  | 2         | 168    | 25        | 165     | 46         |
| 101 | girl   | 15  | ambi   | 2         | 155    | 22        | 157     | 8          |

TABLE 1: Part of the data set for Year 11 students

## Appendices

### STATISTICAL INVESTIGATIONS PRETEST

Name: \_\_\_\_\_

#### TASK A

1. Your task is to pose questions about the 2007 NZ Census At School data using the variables given in the data sheet. You need to pose at least six questions.

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

2. Beside each of your questions write whether the question you posed is a **summary** or a **comparison** or a **relationship** question.

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS PRETEST**

**Name:** \_\_\_\_\_

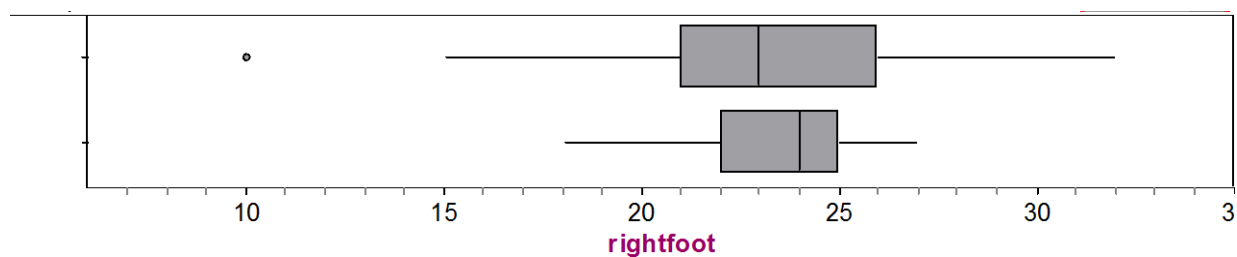
**TASK B**

Emma takes a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls. She can find out their right foot lengths without shoes on in cm.

1. Emma knows that her *random sample* of 30 Year 8 NZ boys can be used to find out ...

Emma is interested in comparing the right foot lengths (in cm) of Year 8 NZ boys and girls. She plots her graph correctly.

**Emma's graph**



Emma looks at her graph and claims that the right foot lengths of Year 8 NZ girls tend to be bigger than the right foot lengths of Year 8 NZ boys.

2. Would you make the same claim as Emma? Why?

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS PRETEST**

**Name:** \_\_\_\_\_

**TASK C**

**Problem:** Do the heights of Year 11 NZ boys tend to be bigger than the heights of Year 11 NZ girls?

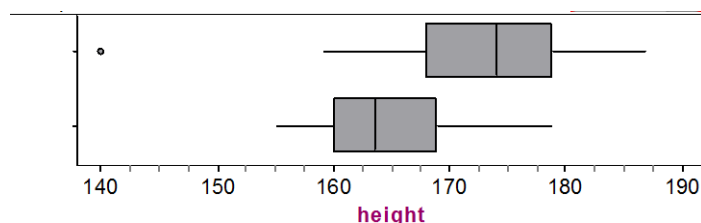
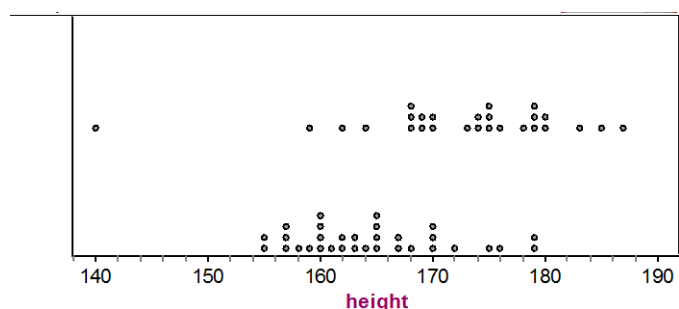
**Plan:** Assume that the 2007 NZ Census At School database is representative of the NZ population. A random sample of 35 Year 11 boys and a random sample of 35 Year 11 girls are taken from the database.

The survey questions asked to get these data were:

- Are you: male/female?
- How tall are you without your shoes on? Answer to the nearest centimetre.

**Data:** It was found that some students did not record their heights. The data from 27 boys and 32 girls are used in this analysis.

**Analysis:**



|                | Gender |       |
|----------------|--------|-------|
| Height         | boy    | girl  |
| Sample size    | 27     | 32    |
| Mean           | 172.6  | 164.6 |
| Minimum        | 140    | 155   |
| Lower Quartile | 168    | 160   |
| Median         | 174    | 163.5 |
| Upper Quartile | 179    | 169   |
| Maximum        | 187    | 179   |

**Plots comparing the heights in cm of Year 11 boys and girls      Statistics for the heights (cm) of Year 11 boys and girls**

**Make statements about what you notice and think about as you look at the graphs/statistics.**

**You should make statements under the headings given.**

**1. Middle 50%:**

a. I notice...

b. I notice ...

c. Back in the two populations I wonder ...

## Appendices

### 2. Anything unusual:

- a. I notice...
  
- b. I worry or think that ...

### 3. Shape:

- a. I notice...
  
- b. Back in the two populations I wonder ...

### 4. Spread:

- a. I notice...
  
- b. Back in the two populations I wonder ...

### **5. Conclusion: Write a conclusion using the headings below.**

- a. Answer the problem:  
“Do the heights of Year 11 NZ boys tend to be bigger than the heights of Year 11 NZ girls?”
  
- b. Statistical support for this conclusion:
  
  
  
  
- c. Does this conclusion make sense with what you personally know about the heights of Year 11 students? Why?
  
  
  
  
- d. Is there anything else you should investigate? Why?

**Hand this sheet in when finished. Thank you for completing these tasks.**

## B.6 Teaching experiment 3 – Post-test 2009

### STATISTICAL INVESTIGATIONS POST-TEST DATA SHEET

Table 1 shows part of a data set from the 2009 NZ Census at School. The data shown were collected from secondary students (Years 9–13) and refer to the following survey questions:

| Variable                 | Survey Question                                                                                                                                                                                                     |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl      | 1. Are you: male/female                                                                                                                                                                                             |
| age: in years            | 2. How old are you?                                                                                                                                                                                                 |
| languages:               | 6. In how many languages can you hold a conversation about a lot of everyday things?                                                                                                                                |
| armspan: in cm           | 9. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.)                                              |
| wrist: in cm             | 10. What is the circumference of your wrist? Answer to the nearest centimetre.                                                                                                                                      |
| neck: in cm              | 11. What is the circumference of your neck? Answer to the nearest centimetre.                                                                                                                                       |
| popliteal: in cm         | 12. What is your popliteal length? Answer to the nearest centimetre. (The popliteal length is the measurement from the underside of the leg right behind the knee when seated, to the floor. Taken with shoes off.) |
| indexfinger: in mm       | 13. What is the length of your index finger to the nearest millimetre?                                                                                                                                              |
| ringfinger: in mm        | 14. What is the length of your ring finger to the nearest millimetre?                                                                                                                                               |
| fitlevel:                | 27. How physically fit do you think you are? Unfit A little bit fit Quite fit Very fit                                                                                                                              |
| pulserest: beats per min | 28. What is your resting pulse rate? (Measure for 15 seconds and multiply by 4, measure after sitting for at least 10 minutes.)                                                                                     |
| year:                    | By default – year at school                                                                                                                                                                                         |
| region:                  | By default – region the school is in                                                                                                                                                                                |

|    | gender | age | languages | armspan | wrist | neck | popliteal | indexfinger | ringfinger | fitlevel  | pulserest | year | region            |
|----|--------|-----|-----------|---------|-------|------|-----------|-------------|------------|-----------|-----------|------|-------------------|
| 8  | girl   | 15  | 1         | 158     | 15    | 30   | 40        | 80          | 70         | littlefit | 84        | 10   | Southland Region  |
| 9  | boy    | 14  | 2         | 178     | 18    | 35   | 40        | 90          | 95         | littlefit | 104       | 10   | Auckland Region   |
| 10 | girl   | 15  | 2         | 171     | 16    | 35   | 41        | 80          | 84         | quitefit  | 76        | 11   | Auckland Region   |
| 11 | girl   | 13  | 1         | 167     | 15    | 29   | 58        | 85          | 70         | quitefit  | 68        | 9    | Otago Region      |
| 12 | girl   | 15  | 1         | 155     | 17    | 33   | 41        | 70          | 70         | littlefit | 110       | 11   | Auckland Region   |
| 13 | boy    | 14  | 1         | 177     | 19    | 37   | 74        | 65          | 70         | veryfit   | 71        | 10   | Waikato Region    |
| 14 | boy    | 12  | 1         | 156     | 16    | 101  | 41        | 80          | 70         | veryfit   | -1        | 9    | Wellington Region |
| 15 | girl   | 15  | 1         | 168     | 11    | 30   | 40        | 90          | 75         | quitefit  | 80        | 11   | Auckland Region   |
| 16 | girl   | 14  | 2         | 166     | 14    | 30   | 40        | 70          | 65         | quitefit  | 80        | 10   | Auckland Region   |
| 17 | boy    | 15  | 2         | 180     | 18    | 40   | 51        | 95          | 75         | littlefit | 80        | 11   | Auckland Region   |
| 18 | girl   | 15  | 2         | 138     | 14    | 28   | 39        | 65          | 70         | littlefit | 80        | 11   | Auckland Region   |
| 19 | boy    | 14  | 1         | 170     | 18    | 34   | 49        | 90          | 89         | quitefit  | 84        | 10   | Southland Region  |
| 20 | girl   | 14  | 1         | 148     | 15    | 31   | 35        | 60          | 70         | quitefit  | 72        | 10   | Bay of Plenty ReÉ |
| 21 | girl   | 13  | 2         | 160     | 15    | 30   | 41        | 93          | 96         | quitefit  | 72        | 9    | Nelson Region     |
| 22 | boy    | 13  | 2         | 166     | 18    | 43   | 66        | 70          | 78         | veryfit   | 48        | 9    | Auckland Region   |
| 23 | boy    | 14  | 2         | 170     | 20    | 35   | 62        | 50          | 53         | quitefit  | 72        | 10   | Wellington Region |
| 24 | boy    | 17  | 2         | 200     | 18    | 38   | 43        | 90          | 100        | quitefit  | 60        | 13   | Wellington Region |
| 25 | girl   | 12  | 0         | 22      | 15    | 3    | 43        | 8           | 6          | veryfit   | 100       | 9    | Manaw atu-WanÉ    |
| 26 | girl   | 12  | 1         | 154     | 13    | 28   | 39        | 89          | 95         | littlefit | 84        | 9    | Wellington Region |
| 27 | girl   | 13  | 1         | 184     | 14    | 33   | 48        | 77          | 66         | quitefit  | 70        | 10   | Waikato Region    |
| 28 | girl   | 15  | 1         | 162     | 16    | 31   | 46        | 83          | 70         | littlefit | 76        | 10   | Canterbury Region |
| 29 | girl   | 17  | 1         | 123     | 15    | 32   | 42        | 83          | 60         | quitefit  | 64        | 13   | Auckland Region   |
| 30 | girl   | 16  | 2         | 162     | 6     | 33   | 39        | 73          | 76         | littlefit | 65        | 12   | Auckland Region   |
| 31 | girl   | 15  | 1         | 164     | 17    | 35   | 46        | 71          | 75         | littlefit | 64        | 11   | Auckland Region   |
| 32 | girl   | 13  | 1         | 148     | 6     | 13   | 16        | 80          | 50         | littlefit | 80        | 9    | Canterbury Region |
| 33 | girl   | 12  | 2         | 155     | 20    | 27   | 48        | 70          | 65         | unfit     | 1         | 9    | Auckland Region   |

TABLE 1: Part of the data set for secondary students



## Appendices

### STATISTICAL INVESTIGATIONS POST-TEST    Name: \_\_\_\_\_

#### TASK A

1. Your task is to pose questions about the 2009 NZ Census At School data using the variables given in the data sheet. You need to pose at least six questions.

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

*I wonder...*

2. Beside each of your questions write whether the question you posed is a **summary** or a **comparison** or a **relationship** question.

**Hand this sheet in when finished and get the next task.**

## STATISTICAL INVESTIGATIONS POST-TEST      Name: \_\_\_\_\_

### TASK B

The following questions were posed about the data set given in Table 1. For each question comment on whether you think the question is a good question or not. Give reasons for why or why not. If the question is not a good question, change it to make it a better question.

1. I wonder if girls have longer popliteal lengths than boys?
2. I wonder who has the biggest armspan?
3. I wonder what are typical neck circumferences for these students?
4. I wonder if the armspan of NZ secondary school boys tends to be longer than the armspan of NZ secondary school girls?
5. I wonder what is the most popular sport played?

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS POST-TEST**      **Name:** \_\_\_\_\_

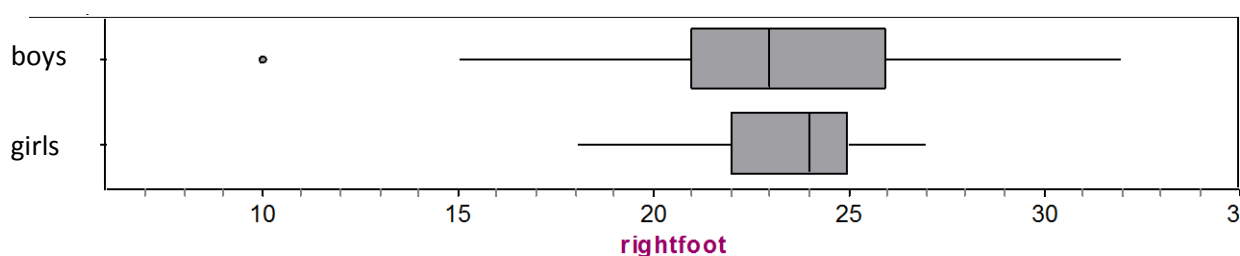
**TASK C**

Emma takes a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls. She can find out their right foot lengths without shoes on in cm.

1. Emma knows that her *random sample* of 30 Year 8 NZ boys can be used to find out ...

Emma is interested in comparing the right foot lengths (in cm) of Year 8 NZ boys and girls. She plots her graph correctly.

**Emma's graph**



Emma looks at her graph and claims that the right foot lengths of Year 8 NZ girls tend to be bigger than the right foot lengths of Year 8 NZ boys.

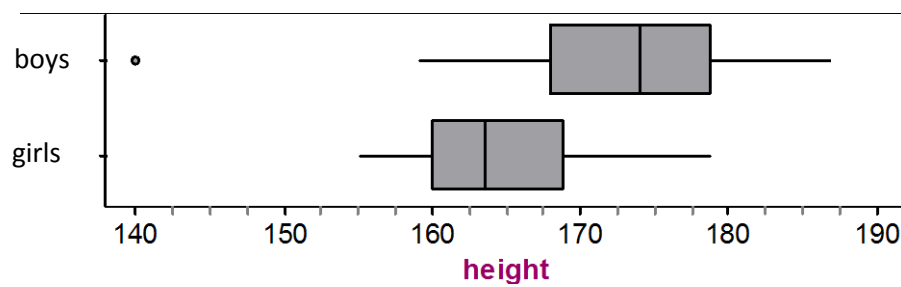
2. Would you make the same claim as Emma? Why?

## Appendices

Emma then takes a *random sample* of 30 Year 11 NZ boys and a *random sample* of 30 Year 11 NZ girls. She can find out their height without shoes on in cm.

Emma is interested in comparing the heights (in cm) of Year 11 NZ boys and girls. She plots her graph correctly.

### Emma's graph



Emma looks at her graph and claims that the heights of Year 11 NZ boys tend to be greater than the heights of Year 11 NZ girls.

3. Would you make the same claim as Emma? Why?

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS POST-TEST**      **Name:** \_\_\_\_\_

**TASK D**

**Background:** It was reported in the newspaper that NZ school students were not getting enough sleep because they were sending text messages all night. We will check this claim in this investigation.

**Problem:** Do the hours of sleep per night for NZ Year 5 to Year 10 students who own cell phones tend to be less than the hours of sleep per night for NZ Year 5 to Year 10 students who do not own cell phones?

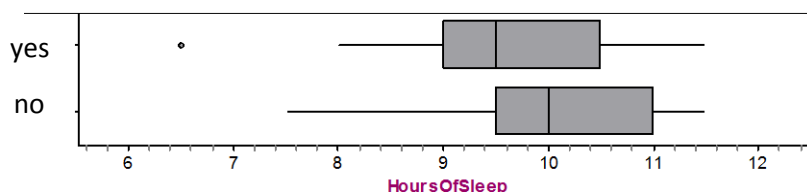
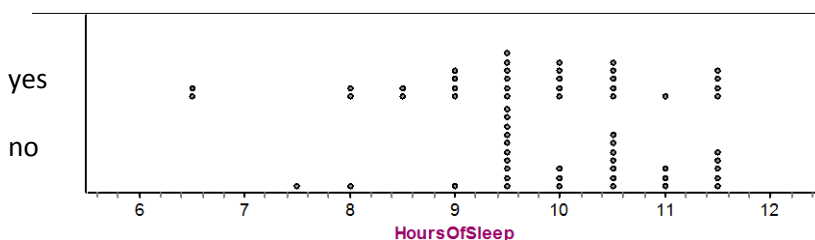
**Plan:** Assume that the 2005 NZ Census At School database is representative of the NZ Year 5 to Year 10 population. A random sample of 31 cell phone owners and a random sample of 31 non-cell phone owners are taken from the database.

The survey questions asked to get these data were:

- Do you own a cell phone: yes/no
- How many hours did you sleep last night? Answer to the nearest half hour.

**Data:** The data from 31 cell phone owners and 31 non-cell phone owners are used in this analysis.

**Analysis:**



|                | Cell phone |      |
|----------------|------------|------|
| Hours of Sleep | No         | Yes  |
| Sample size    | 31         | 31   |
| Mean           | 10.1       | 9.6  |
| Minimum        | 7.5        | 6.5  |
| Lower Quartile | 9.5        | 9    |
| Median         | 10         | 9.5  |
| Upper Quartile | 11         | 10.5 |
| Maximum        | 11.5       | 11.5 |

**Plots comparing the hours of sleep per night and cell phone ownership for Year 5 to Year 10 students**

**Statistics for hours of sleep per night & cell phone ownership for Year 5 to 10 students**

**Make statements about what you notice and think about as you look at the graphs/statistics. You should make statements under the headings given.**

**1. Middle 50%:**

- a. From the samples I notice...
- b. From the samples I notice ...
- c. I think that back in the two populations ...

## Appendices

### 2. Anything unusual:

- a. From the samples I notice...
- b. I worry or think that ...

### 3. Shape:

- a. From the samples I notice...
- b. I think that back in the two populations ...

### 4. Spread:

- a. From the samples I notice...
- b. I think that back in the two populations ...

### **5. Conclusion: Write a conclusion using the headings below.**

- a. Answer the problem:  
“Do the hours of sleep per night for NZ Year 5 to Year 10 students who own cell phones tend to be less than the hours of sleep per night for NZ Year 5 to Year 10 students who do not own cell phones?”
- b. Statistical support for this conclusion:
- c. Does this conclusion make sense with what you personally know about the hours of sleep per night for Year 5 to 10 students?
- d. Is there anything else you should investigate? Why?

**Hand this sheet in when finished and get the next task.**

**1. Describe three main ideas that you have learnt in the statistics unit.**

- 2. Describe the part in the statistics unit that you found interesting. Explain why you found this part interesting.**

284

## B.7 Teaching experiment 4 – Pre-test 2011

### STATISTICAL INVESTIGATIONS PRE-TEST DATA SHEET

Table 1 shows part of a sample of data from the 2009 NZ Census at School. The data shown were collected from secondary students (Years 9–13) and refer to the following survey questions:

| Variable                 | Survey Question                                                                                                                                                                                                     |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl      | 1. Are you: male/female                                                                                                                                                                                             |
| age: in years            | 2. How old are you?                                                                                                                                                                                                 |
| languages:               | 6. In how many languages can you hold a conversation about a lot of everyday things?                                                                                                                                |
| armspan: in cm           | 9. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.)                                              |
| wrist: in cm             | 10. What is the circumference of your wrist? Answer to the nearest centimetre.                                                                                                                                      |
| neck: in cm              | 11. What is the circumference of your neck? Answer to the nearest centimetre.                                                                                                                                       |
| popliteal: in cm         | 12. What is your popliteal length? Answer to the nearest centimetre. (The popliteal length is the measurement from the underside of the leg right behind the knee when seated, to the floor. Taken with shoes off.) |
| indexfinger: in mm       | 13. What is the length of your index finger to the nearest millimetre?                                                                                                                                              |
| ringfinger: in mm        | 14. What is the length of your ring finger to the nearest millimetre?                                                                                                                                               |
| fitlevel:                | 27. How physically fit do you think you are? Unfit A little bit fit Quite fit Very fit                                                                                                                              |
| pulserest: beats per min | 28. What is your resting pulse rate? (Measure for 15 seconds and multiply by 4, measure after sitting for at least 10 minutes.)                                                                                     |
| year:                    | By default – year at school                                                                                                                                                                                         |
| region:                  | By default – region the school is in                                                                                                                                                                                |

|    | gender | age | languages | armspan | wrist | neck | popliteal | indexfinger | ringfinger | fitlevel  | pulserest | year | region              |
|----|--------|-----|-----------|---------|-------|------|-----------|-------------|------------|-----------|-----------|------|---------------------|
| 8  | girl   | 15  | 1         | 158     | 15    | 30   | 40        | 80          | 70         | littlefit | 84        | 10   | Southland Region    |
| 9  | boy    | 14  | 2         | 178     | 18    | 35   | 40        | 90          | 95         | littlefit | 104       | 10   | Auckland Region     |
| 10 | girl   | 15  | 2         | 171     | 16    | 35   | 41        | 80          | 84         | quitefit  | 76        | 11   | Auckland Region     |
| 11 | girl   | 13  | 1         | 167     | 15    | 29   | 58        | 85          | 70         | quitefit  | 68        | 9    | Otago Region        |
| 12 | girl   | 15  | 1         | 155     | 17    | 33   | 41        | 70          | 70         | littlefit | 110       | 11   | Auckland Region     |
| 13 | boy    | 14  | 1         | 177     | 19    | 37   | 74        | 65          | 70         | veryfit   | 71        | 10   | Waikato Region      |
| 14 | boy    | 12  | 1         | 156     | 16    | 101  | 41        | 80          | 70         | veryfit   | -1        | 9    | Wellington Region   |
| 15 | girl   | 15  | 1         | 168     | 11    | 30   | 40        | 90          | 75         | quitefit  | 80        | 11   | Auckland Region     |
| 16 | girl   | 14  | 2         | 166     | 14    | 30   | 40        | 70          | 65         | quitefit  | 80        | 10   | Auckland Region     |
| 17 | boy    | 15  | 2         | 180     | 18    | 40   | 51        | 95          | 75         | littlefit | 80        | 11   | Auckland Region     |
| 18 | girl   | 15  | 2         | 138     | 14    | 28   | 39        | 65          | 70         | littlefit | 80        | 11   | Auckland Region     |
| 19 | boy    | 14  | 1         | 170     | 18    | 34   | 49        | 90          | 89         | quitefit  | 84        | 10   | Southland Region    |
| 20 | girl   | 14  | 1         | 148     | 15    | 31   | 35        | 60          | 70         | quitefit  | 72        | 10   | Bay of Plenty Re... |
| 21 | girl   | 13  | 2         | 160     | 15    | 30   | 41        | 93          | 96         | quitefit  | 72        | 9    | Nelson Region       |
| 22 | boy    | 13  | 2         | 166     | 18    | 43   | 66        | 70          | 78         | veryfit   | 48        | 9    | Auckland Region     |
| 23 | boy    | 14  | 2         | 170     | 20    | 35   | 62        | 50          | 53         | quitefit  | 72        | 10   | Wellington Region   |
| 24 | boy    | 17  | 2         | 200     | 18    | 38   | 43        | 90          | 100        | quitefit  | 60        | 13   | Wellington Region   |
| 25 | girl   | 12  | 0         | 22      | 15    | 3    | 43        | 8           | 6          | veryfit   | 100       | 9    | Manawatu-Wan...     |
| 26 | girl   | 12  | 1         | 154     | 13    | 28   | 39        | 89          | 95         | littlefit | 84        | 9    | Wellington Region   |
| 27 | girl   | 13  | 1         | 184     | 14    | 33   | 48        | 77          | 66         | quitefit  | 70        | 10   | Waikato Region      |
| 28 | girl   | 15  | 1         | 162     | 16    | 31   | 46        | 83          | 70         | littlefit | 76        | 10   | Canterbury Region   |
| 29 | girl   | 17  | 1         | 123     | 15    | 32   | 42        | 83          | 60         | quitefit  | 64        | 13   | Auckland Region     |
| 30 | girl   | 16  | 2         | 162     | 6     | 33   | 39        | 73          | 76         | littlefit | 65        | 12   | Auckland Region     |
| 31 | girl   | 15  | 1         | 164     | 17    | 35   | 46        | 71          | 75         | littlefit | 64        | 11   | Auckland Region     |
| 32 | girl   | 13  | 1         | 148     | 6     | 13   | 16        | 80          | 50         | littlefit | 80        | 9    | Canterbury Region   |
| 33 | girl   | 12  | 2         | 155     | 20    | 37   | 48        | 70          | 65         | unfit     | 4         | 9    | Auckland Region     |

TABLE 1: Part of the sample of data from 2009 NZ Census at School



**STATISTICAL INVESTIGATIONS PRE-TEST**

**Name:** \_\_\_\_\_

**TASK A**

Your task is to pose investigative questions that can be answered using the sample of data from the 2009 NZ Census At School survey as outlined in the data sheet. The population that the sample comes from and the variables available are detailed in the data sheet. You need to pose three summary type investigative questions and three comparison type investigative questions.

**Summary questions**

*1. I wonder...*

*2. I wonder...*

*3. I wonder...*

**Comparison questions**

*4. I wonder...*

*5. I wonder...*

*6. I wonder...*

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS PRE-TEST**      **Name:** \_\_\_\_\_

**TASK B**

The following investigative questions were posed about the data set given in Table 1. For each investigative question comment on whether you think the investigative question is a good investigative question or not. Give reasons for why or why not. If the investigative question is not a good investigative question, change it to make it a better investigative question.

1. Do girls have longer popliteal lengths than boys?
  
  
  
  
  
  
  
  
  
  
2. Who has the biggest armspan?
  
  
  
  
  
  
  
  
  
  
3. What are typical neck circumferences for these students?
  
  
  
  
  
  
  
  
  
  
4. Do the armspans of NZ secondary school boys tend to be longer than the armspans of NZ secondary school girls?
  
  
  
  
  
  
  
  
  
  
5. What is the most popular sport played?

**Hand this sheet in when finished and get the next task.**

## Appendices

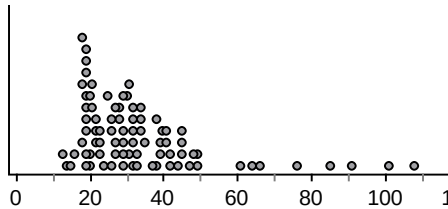
### STATISTICAL INVESTIGATIONS PRE-TEST

Name: \_\_\_\_\_

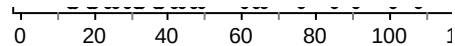
#### TASK C

For each of the three situations given, sketch the shape of the distribution of the variable and write two statements about the distribution of the variable.

**All Blacks scores in Test Matches 2005–2010** 1. Sketch of shape of distribution of ABs scores

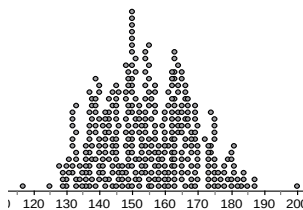


All Blacks' scores

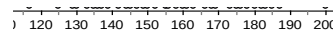


2. Describe the distribution of ABs scores

**Heights of NZ Yr 5–10 students** 3. Sketch of shape of distribution of NZ Yr 5-10 student heights



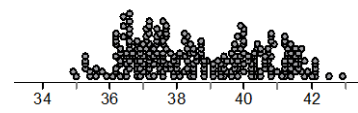
height (cm)



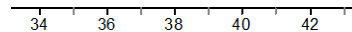
4. Describe the distribution of NZ Yr5–10 student heights

**Heights of Tokoeka Kiwis**

5. Sketch of shape of distribution of Tokoeka kiwi heights



height (cm)



6. Describe the distribution of Tokoeka kiwi heights

**Hand this sheet in when finished and get the next task.**

## Appendices

### STATISTICAL INVESTIGATIONS PRE-TEST

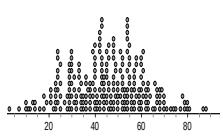
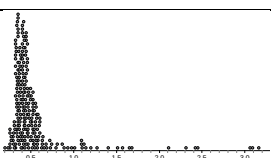
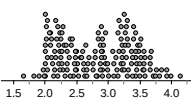
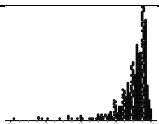
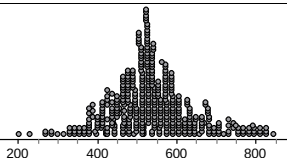
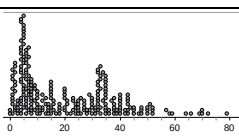
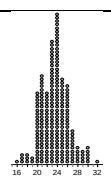
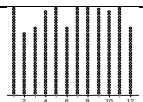
Name: \_\_\_\_\_

#### TASK D

Complete the table by selecting one or more shape descriptors for each graph.

#### SHAPE DESCRIPTORS:

**bimodal    left skew    right skew    symmetric    uniform    unimodal**  
**none of these**

| Graph                                                                               | shape descriptor(s) |
|-------------------------------------------------------------------------------------|---------------------|
|    | 1.                  |
|    | 2.                  |
|  | 3.                  |
|  | 4.                  |
|  | 5.                  |
|  | 6.                  |
|  | 7.                  |
|  | 8.                  |

Hand this sheet in when finished and get the next task.

**STATISTICAL INVESTIGATIONS PRE-TEST**

**Name:** \_\_\_\_\_

**TASK E**

You are required to sketch what you suspect the distribution of data for the variable given (age), and for each of the three populations (Year 10, C@S Yr 5–13, NZAMT participants). **Give some idea of the values that each of the variables will have.**

|                                                                                                                                                       | age (in years)                                                     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| <p>Year 10</p> <p>Data collected from a year 10 mathematics class.</p> <p>28 students in the class.</p>                                               | <p>Sketch 1</p>                                                    |
| <p>Reason</p>                                                                                                                                         | <p>1. I think the distribution will look like this because ...</p> |
| <p>C@S (Year 5-13)</p> <p>Data collected on Census At School for all students from year 5 through to year 13.</p> <p>22,000 students participated</p> | <p>Sketch 2</p>                                                    |
| <p>Reason</p>                                                                                                                                         | <p>2. I think the distribution will look like this because ...</p> |
| <p>Maths teachers</p> <p>Data collected from NZ Maths teachers at a conference.</p> <p>147 participants.</p>                                          | <p>Sketch 3</p>                                                    |
| <p>Reason</p>                                                                                                                                         | <p>3. I think the distribution will look like this because ...</p> |

4. If your three age distributions are different explain why.

**Hand this sheet in when finished and get the next task.**

**STATISTICAL INVESTIGATIONS PRE-TEST**

**Name:** \_\_\_\_\_

**TASK F**

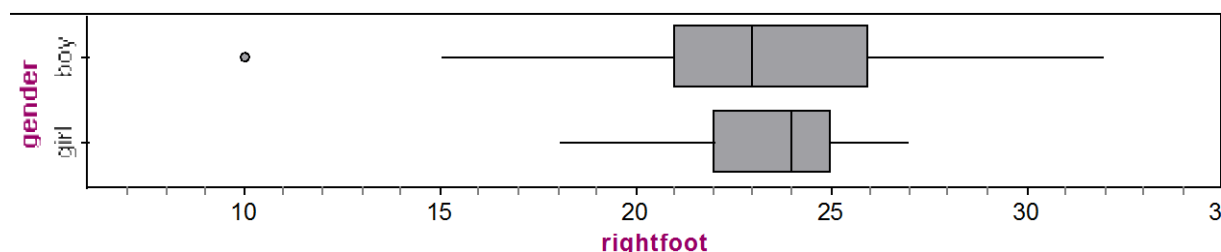
Emma is interested in comparing the right foot lengths (in cm) of Year 8 NZ boys and girls. She takes a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls. She can find out their right foot lengths without shoes on in cm.

She plots the right foot lengths from her samples correctly.

1. Emma knows that her *random sample* of 30 Year 8 NZ boys can be used to find out ...

Emma is interested in comparing the right foot lengths (in cm) of Year 8 NZ boys and girls. She plots her graph correctly.

**Emma's graph**



Emma looks at her graph and claims that the right foot lengths of Year 8 NZ girls tend to be bigger than the right foot lengths of Year 8 NZ boys.

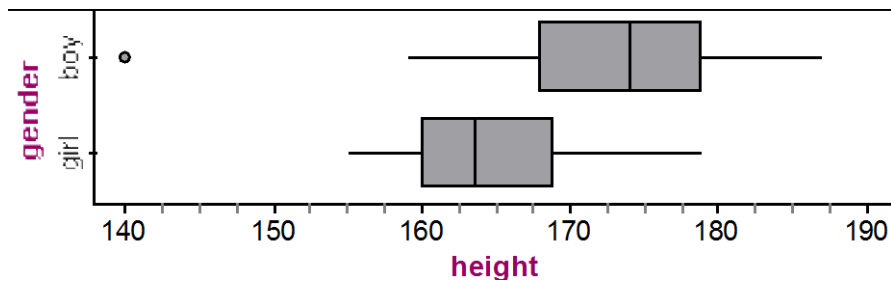
2. Would you make the same claim as Emma? Why?

## Appendices

Emma is interested in comparing the heights (in cm) of Year 11 NZ boys and girls. She takes a *random sample* of 30 Year 11 NZ boys and a *random sample* of 30 Year 11 NZ girls. She can find out their height without shoes on in cm.

She plots the heights from her samples correctly.

### Emma's graph



Emma looks at her graph and claims that the heights of Year 11 NZ boys tend to be greater than the heights of Year 11 NZ girls.

3. Would you make the same claim as Emma? Why?

**Hand this sheet in when finished. Thank you for completing the pre-test.**

## B.8 Teaching experiment 4 – Post-test 2011

### STATISTICAL INVESTIGATIONS POST-TEST DATA SHEET

Table 1 shows part of a sample of data from the 2009 NZ Census at School. The data shown were collected from secondary students (Years 9–13) and refer to the following survey questions:

| Variable                 | Survey Question                                                                                                                                                                                                     |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| gender: boy or girl      | 1. Are you: male/female                                                                                                                                                                                             |
| age: in years            | 2. How old are you?                                                                                                                                                                                                 |
| languages:               | 6. In how many languages can you hold a conversation about a lot of everyday things?                                                                                                                                |
| armspan: in cm           | 9. What is your arm span? Answer to the nearest centimetre. (Open arms wide, measure distance from tip of right hand middle finger to tip of left hand middle finger.)                                              |
| wrist: in cm             | 10. What is the circumference of your wrist? Answer to the nearest centimetre.                                                                                                                                      |
| neck: in cm              | 11. What is the circumference of your neck? Answer to the nearest centimetre.                                                                                                                                       |
| popliteal: in cm         | 12. What is your popliteal length? Answer to the nearest centimetre. (The popliteal length is the measurement from the underside of the leg right behind the knee when seated, to the floor. Taken with shoes off.) |
| indexfinger: in mm       | 13. What is the length of your index finger to the nearest millimetre?                                                                                                                                              |
| ringfinger: in mm        | 14. What is the length of your ring finger to the nearest millimetre?                                                                                                                                               |
| fitlevel:                | 27. How physically fit do you think you are? Unfit A little bit fit Quite fit Very fit                                                                                                                              |
| pulserest: beats per min | 28. What is your resting pulse rate? (Measure for 15 seconds and multiply by 4, measure after sitting for at least 10 minutes.)                                                                                     |
| year:                    | By default – year at school                                                                                                                                                                                         |
| region:                  | By default – region the school is in                                                                                                                                                                                |

|    | gender | age | languages | armspan | wrist | neck | popliteal | indexfinger | ringfinger | fitlevel  | pulserest | year | region              |
|----|--------|-----|-----------|---------|-------|------|-----------|-------------|------------|-----------|-----------|------|---------------------|
| 8  | girl   | 15  | 1         | 158     | 15    | 30   | 40        | 80          | 70         | littlefit | 84        | 10   | Southland Region    |
| 9  | boy    | 14  | 2         | 178     | 18    | 35   | 40        | 90          | 95         | littlefit | 104       | 10   | Auckland Region     |
| 10 | girl   | 15  | 2         | 171     | 16    | 35   | 41        | 80          | 84         | quitefit  | 76        | 11   | Auckland Region     |
| 11 | girl   | 13  | 1         | 167     | 15    | 29   | 58        | 85          | 70         | quitefit  | 68        | 9    | Otago Region        |
| 12 | girl   | 15  | 1         | 155     | 17    | 33   | 41        | 70          | 70         | littlefit | 110       | 11   | Auckland Region     |
| 13 | boy    | 14  | 1         | 177     | 19    | 37   | 74        | 65          | 70         | veryfit   | 71        | 10   | Waikato Region      |
| 14 | boy    | 12  | 1         | 156     | 16    | 101  | 41        | 80          | 70         | veryfit   | -1        | 9    | Wellington Region   |
| 15 | girl   | 15  | 1         | 168     | 11    | 30   | 40        | 90          | 75         | quitefit  | 80        | 11   | Auckland Region     |
| 16 | girl   | 14  | 2         | 166     | 14    | 30   | 40        | 70          | 65         | quitefit  | 80        | 10   | Auckland Region     |
| 17 | boy    | 15  | 2         | 180     | 18    | 40   | 51        | 95          | 75         | littlefit | 80        | 11   | Auckland Region     |
| 18 | girl   | 15  | 2         | 138     | 14    | 28   | 39        | 65          | 70         | littlefit | 80        | 11   | Auckland Region     |
| 19 | boy    | 14  | 1         | 170     | 18    | 34   | 49        | 90          | 89         | quitefit  | 84        | 10   | Southland Region    |
| 20 | girl   | 14  | 1         | 148     | 15    | 31   | 35        | 60          | 70         | quitefit  | 72        | 10   | Bay of Plenty Re... |
| 21 | girl   | 13  | 2         | 160     | 15    | 30   | 41        | 93          | 96         | quitefit  | 72        | 9    | Nelson Region       |
| 22 | boy    | 13  | 2         | 166     | 18    | 43   | 66        | 70          | 78         | veryfit   | 48        | 9    | Auckland Region     |
| 23 | boy    | 14  | 2         | 170     | 20    | 35   | 62        | 50          | 53         | quitefit  | 72        | 10   | Wellington Region   |
| 24 | boy    | 17  | 2         | 200     | 18    | 38   | 43        | 90          | 100        | quitefit  | 60        | 13   | Wellington Region   |
| 25 | girl   | 12  | 0         | 22      | 15    | 3    | 43        | 8           | 6          | veryfit   | 100       | 9    | Manawatu-Wan...     |
| 26 | girl   | 12  | 1         | 154     | 13    | 28   | 39        | 89          | 95         | littlefit | 84        | 9    | Wellington Region   |
| 27 | girl   | 13  | 1         | 184     | 14    | 33   | 48        | 77          | 66         | quitefit  | 70        | 10   | Waikato Region      |
| 28 | girl   | 15  | 1         | 162     | 16    | 31   | 46        | 83          | 70         | littlefit | 76        | 10   | Canterbury Region   |
| 29 | girl   | 17  | 1         | 123     | 15    | 32   | 42        | 83          | 60         | quitefit  | 64        | 13   | Auckland Region     |
| 30 | girl   | 16  | 2         | 162     | 6     | 33   | 39        | 73          | 76         | littlefit | 65        | 12   | Auckland Region     |
| 31 | girl   | 15  | 1         | 164     | 17    | 35   | 46        | 71          | 75         | littlefit | 64        | 11   | Auckland Region     |
| 32 | girl   | 13  | 1         | 148     | 6     | 13   | 16        | 80          | 50         | littlefit | 80        | 9    | Canterbury Region   |
| 33 | girl   | 12  | 2         | 155     | 20    | 27   | 48        | 70          | 65         | unfit     | 4         | 9    | Auckland Region     |

TABLE 1: Part of the sample of data from 2009 NZ Census at School



**STATISTICAL INVESTIGATIONS POST-TEST**    **Name:** \_\_\_\_\_

**TASK A**

Your task is to pose investigative questions that can be answered using the sample of data from the 2009 NZ Census At School survey as outlined in the data sheet. The population that the sample comes from and the variables available are detailed in the data sheet. You need to pose three summary type investigative questions and three comparison type investigative questions.

**Summary questions**

*1. I wonder...*

*2. I wonder...*

*3. I wonder...*

**Comparison questions**

*4. I wonder...*

*5. I wonder...*

*6. I wonder...*

**Hand this sheet in when finished and get the next task.**

The following investigative questions were posed about the data set given in Table 1. For each investigative question comment on whether you think the investigative question is a good investigative question or not. Give reasons for why or why not. If the investigative question is not a good investigative question, change it to make it a better investigative question.

- Hand this sheet in when finished and get the next task.**

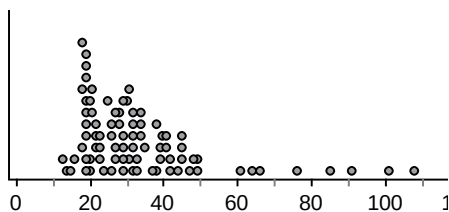
## Appendices

### STATISTICAL INVESTIGATIONS POST-TEST Name: \_\_\_\_\_

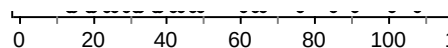
#### TASK C

For each of the three situations given, sketch the shape of the distribution of the variable and write two statements about the distribution of the variable.

**All Blacks scores in Test Matches 2005–2010** 1. Sketch of shape of distribution of ABs scores

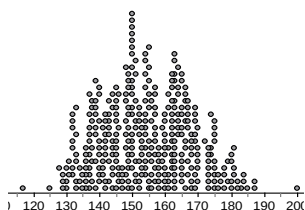


All Blacks' scores

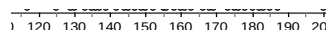


2. Describe the distribution of ABs scores

**Heights of NZ Yr 5-10 students** 3. Sketch of shape of distribution of NZ Yr 5–10 student heights



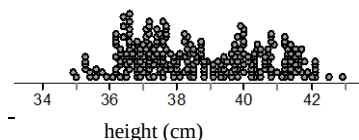
height (cm)



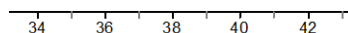
4. Describe the distribution of NZ Yr 5–10 student heights

**Heights of Tokoeka Kiwis**

5. Sketch of shape of distribution of Tokoeka kiwi heights



height (cm)



6. Describe the distribution of Tokoeka kiwi heights

**Hand this sheet in when finished and get the next task.**

## Appendices

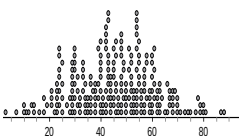
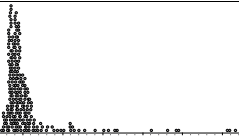
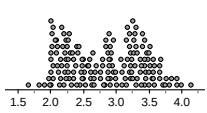
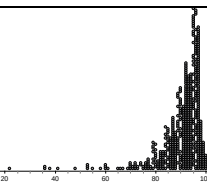
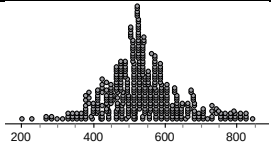
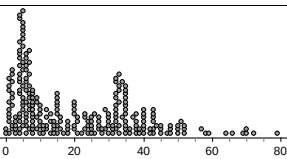
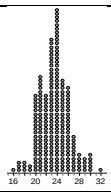
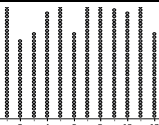
### STATISTICAL INVESTIGATIONS POST-TEST Name: \_\_\_\_\_

#### TASK D

Complete the table by selecting one or more shape descriptors for each graph.

#### SHAPE DESCRIPTORS:

**bimodal    left skew    right skew    symmetric    uniform    unimodal**  
**none of these**

| Graph                                                                               | Shape descriptor(s) |
|-------------------------------------------------------------------------------------|---------------------|
|    | 1.                  |
|    | 2.                  |
|   | 3.                  |
|  | 4.                  |
|  | 5.                  |
|  | 6.                  |
|  | 7.                  |
|  | 8.                  |

Hand this sheet in when finished and get the next task.

**STATISTICAL INVESTIGATIONS POST-TEST**    **Name:** \_\_\_\_\_

**TASK E**

You are required to sketch what you suspect the distribution of data for the variable given (age), and for each of the three populations (Year 10, C@S Yr 5–13, NZ Maths teachers at a conference). **Give some idea of the values that each of the variables will have.**

|                                                                                                                                                        | Age (in years)                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|
| <p>Year 10</p> <p>Data collected from a year 10 mathematics class.</p> <p>28 students in the class.</p>                                                | <p>Sketch 1</p>                                                    |
| <p>Reason</p>                                                                                                                                          | <p>1. I think the distribution will look like this because ...</p> |
| <p>C@S (Year 5-13)</p> <p>Data collected on Census At School for all students from year 5 through to year 13.</p> <p>22,000 students participated.</p> | <p>Sketch 2</p>                                                    |
| <p>Reason</p>                                                                                                                                          | <p>2. I think the distribution will look like this because ...</p> |
| <p>Maths teachers</p> <p>Data collected from NZ Maths teachers at a conference.</p> <p>147 participants.</p>                                           | <p>Sketch 3</p>                                                    |
| <p>Reason</p>                                                                                                                                          | <p>3. I think the distribution will look like this because ...</p> |

4. If your three age distributions are different explain why.

**Hand this sheet in when finished and get the next task.**

Matt is interested in comparing the schoolbag weights (in g) of Year 11 NZ boys and girls. He takes a *random sample* of 30 Year 11 NZ boys and a *random sample* of 30 Year 11 NZ girls. He can find out their school bag weights to the nearest 100g.

1. Matt knows that his *random sample* of 30 Year 11 NZ boys can be used to find out ...

A box plot showing the distribution of bagweight for boys and girls. The y-axis is labeled 'gender' with categories 'boy' and 'girl'. The x-axis is labeled 'bagweight' with values from 0 to 10000. The 'boy' box has a median around 4000, a Q1 around 2500, a Q3 around 4500, and a whisker extending to 7500. The 'girl' box has a median around 3000, a Q1 around 2000, a Q3 around 5200, and a whisker extending to 9000. There is an outlier for boys at approximately 8000.

2. Would you make the same claim as Matt? Why?

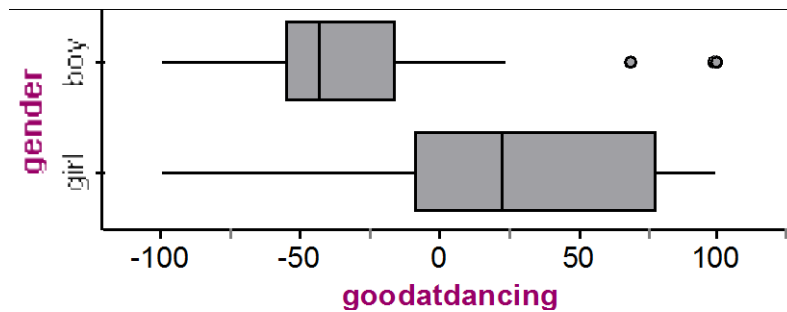
299

## Appendices

Matt is interested in comparing how Year 9 NZ boys and girls rate themselves at dancing. He takes a *random sample* of 30 Year 9 NZ boys and a *random sample* of 30 Year 9 NZ girls. He can find out how good they rate themselves at dancing on a scale of  $-100$  (no good) to  $+100$  (very good).

He plots the results from his samples correctly.

### Matt's graph



Matt looks at his graph and claims that Year 9 NZ girls tend to rate themselves better at dancing than Year 9 NZ boys.

4. Would you make the same claim as Matt? Why?
5. Does Matt's claim make sense with what you personally know about how Year 9 boys and girls would rate themselves at dancing? Why?

**Hand this sheet in when finished and get the next task.**

1. Describe three main ideas that you have learnt in the statistics unit.

2. Describe the part in the statistics unit that you found interesting. Explain why you found this part interesting.

301



## Appendix C – Marking criteria for Teaching Experiment 1 pre- and post-tests

Question marking criteria is in Table C–1, data displays in Table C–2, descriptive statements in Table C–3, and conclusion in Table C–4.

**Table C–1. Question marking criteria**

| Mark | Description                                                                                                                             | Examples                                               |
|------|-----------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| 0    | A nonsense question or a question that was unable to be answered with the given variables.                                              | Is the Asians circled above related?                   |
| 1    | A question that is related to the variables, but not answerable with the given variables.                                               | What region is the most preferred to live in?          |
| 2    | A question that asks how many there are of a particular category for the given variables.                                               | How many children are NZ European?                     |
| 3    | A question that asks for the most popular category for the given variables.                                                             | What is the most ethnicity population in this data?    |
| 4    | A summary or comparison question that can be answered with the given variables.                                                         | What gender has the longest average right foot length? |
| 5    | A summary or comparison question that can be answered with the given variables and considers the population from which the sample came. | No examples.                                           |

**Table C–2. Data displays marking criteria**

| Mark | Description                                                                                       | Examples                                                        |
|------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| 0    | Incorrect or no graph for the data given.                                                         | A line graph.                                                   |
| 1    | An appropriate graph for the data given, but not related to the question posed.                   | Graphed age when their question was about years in New Zealand. |
| 2    | Graph/statistics almost correct or complete.                                                      | Bar graph without gaps, inconsistent scale.                     |
| 3    | Graph(s) constructed correctly [but scales might not allow comparison (for comparison question)]. | Box plots with different scales.                                |
| 4    | Included appropriate statistics for their question.                                               | Five number summary.                                            |
| 5    | Used multiple displays.                                                                           | A dot and a box plot for comparison.                            |

**Table C–3. Descriptive statements marking criteria**

| Mark | Description                                                                             | Examples                                                                                                                                                                                                                                    |
|------|-----------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 0    | No analysis or irrelevant statements.                                                   | No descriptions.                                                                                                                                                                                                                            |
| 1    | Irrelevant statement(s).                                                                | No examples.                                                                                                                                                                                                                                |
| 2    | At least one statement that reflected the question asked and reflected the sample data. | Boys have a larger right foot of 32.                                                                                                                                                                                                        |
| 3    | Relevant statements for the question asked and reflected the sample data.               | Some of the girl students were taller than some of the boy students in the census data. The minimum height of the girls was taller than the minimum height of the boys. The boys' maximum height was taller than the girls' maximum height. |
| 4    | As for 3., and reflects the context, including variables, values and units.             | The girls' smallest height is bigger than the boys' smallest height by 6 cm. The boys' biggest height is taller than the girls' biggest height by 6 cm also. The boys' dot plot was a lot more spread out than the girls'.                  |
| 5    | As for 4., but reflects a broader population rather than just the given sample data.    | No examples.                                                                                                                                                                                                                                |

**Table C–4. Conclusion marking criteria**

| Mark | Description                                                                                                                               | Examples                                                                                                                |
|------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
| 0    | No conclusion or irrelevant statements.                                                                                                   | No conclusion.                                                                                                          |
| 1    | Wrote a comment related to the question but not directly answering the question.                                                          | Most students got 8 hours sleep which is the average, so the answer is no.                                              |
| 2    | Answered the question that they posed but irrelevant support for their answer.                                                            | NZ Euro get the most sleep. Pacific Islanders only have two people surveyed, NZ have 16 people surveyed.                |
| 3    | Answered the question that they posed but support either not relevant or uses just a single point of comparison, e.g. maximum or minimum. | Girls have a smaller right foot, boys' smallest right foot is 23, girls' smallest right foot is 19.                     |
| 4    | Answered the question that they posed with support limited to comparing at least two individual summary points.                           | Yes, boys are taller than girls. Although girls have a higher minimum than boys, boys have a higher maximum and median. |
| 5    | Answered the question that they posed with support including comparing the aggregate picture.                                             | No examples.                                                                                                            |

## **Appendix D – Interview schedules for teaching experiments 3 and 4 – related interview questions to making the call**

### ***2009 Pre-interview schedule – Task B and C***

#### **Questions about their basis for making a claim (task B)**

11. Ask about response to Question 1 [in task B].
12. Why was a sample taken? What is a sample? What will a sample tell you?
13. What does random sample mean?
14. What “group” of people might you make statements about using this data set?
15. What can the samples tell you about all Year 8 NZ girls’ and boys’ right foot lengths?
16. How confident are you that a random sample can tell you something about all Year 8 NZ girls’ and boys’ right foot lengths? [population concept] How close will the average length be to the “real one”?
17. How big does the sample have to be to tell you something about all Year 8 NZ girls’ and boys’ right foot lengths?
18. What is the population for this data set?
19. Ask about response to Question 2 [in task B] and probe reasoning of student.
20. If another person took a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls, would their graphs be the same? Why?
21. Sketch what their graph might look like. [Student to do on pre-test sheet.]
22. Is it possible that they could claim that the right foot lengths of Year 8 NZ boys tend to be bigger than the right foot lengths of Year 8 NZ girls? Why?

#### **Questions about their analysis (task C)**

30. Ask about responses to Question 5 [in task C] and probe reasoning. Why did you use the mean/median for this task and not the one in task B?
31. How did you decide how to make that claim? How confident are you in making that claim? Why? Probe reasoning for claim for task C compared with task B.
32. Do you think samples give reliable information about what is happening back in the populations? Why or why not?

## Appendices

33. For those students who believe the sample sizes should be the same: If I took another Year 11 boy at random from the database and his height was recorded, what height would you expect him to be? Probe reasoning further.

### **2009 Post-interview schedule – Tasks C and D**

#### **Questions about their basis for making a claim (task C)**

17. Ask about response to Question 1 [in task C].
18. Why was a sample taken? What is a sample? What will a sample tell you?
19. What does random sample mean?
20. What “group” of people might you make statements about using this data set?
21. What can the samples tell you about all Year 8 NZ girls’ and boys’ right foot lengths?
22. How confident are you that a random sample can tell you something about all Year 8 NZ girls’ and boys’ right foot lengths? [population concept] How close will the average length be to the “real one”?
23. How big does the sample have to be to tell you something about all Year 8 NZ girls’ and boys’ right foot lengths?
24. What is the population for this data set?
25. Ask about response to Question 2 [in task C] and probe reasoning of student.
26. If another person took a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls, would their graphs be the same? Why?
27. Sketch what their graph might look like. [Student to do on post-test sheet.]
28. Is it possible that they could claim that the right foot lengths of Year 8 NZ boys tend to be bigger than the right foot lengths of Year 8 NZ girls? Why?
29. Which situation is this? Situation 1 or Situation 2?
30. Show with your hands what you think could happen with repeated sampling from the population to the graphs?
31. Will girls’ median right foot length always be higher?
32. Would you expect the boxes to overlap in another sample?
33. Ask about response to question 3 and probe reasoning of student.
34. If another person took a *random sample* of 30 Year 11 NZ boys and a *random sample* of 30 Year 11 NZ girls, would their graphs be the same? Why?
35. Sketch what their graph might look like. [Student to do on post-test sheet.]

## Appendices

36. Is it possible that they could claim that heights of Year 11 NZ boys tend to be greater than heights of Year 11 NZ girls? Why?
37. Which situation is this? Situation 1 or Situation 2?
38. Show with your hands what you think could happen with repeated sampling from the population to the graphs?
39. Will boys' median height always be higher?
40. Would you expect the boxes to overlap in another sample?

### **Questions about their analysis (task D)**

48. Ask about responses to Question 5 [in task D] and probe reasoning. Why did you use the mean/median for this task and not the one in task C?
49. How did you decide how to make that claim? How confident are you in making that claim? Why? Probe reasoning for claim for task D compared with task C.
50. Do you think samples give reliable information about what is happening back in the populations? Why or why not?
51. For those students who believe the sample sizes should be the same: If I took another Year 11 boy at random from the database and his height was recorded, what height would you expect him to be? Probe reasoning further.

### **Related web application**

52. Here is a web application that you looked at in class. Tell me what the visual images are showing you. (show Karekaremovie and Karekaremovie2).
53. How do these images help you “make a call”?

**2011 Pre-interview schedule – Task F**

**Questions about their basis for making a claim (task F)**

11. What does random sample mean?
12. Why was a sample taken? What is a sample? What will a sample tell you?
13. What “group” of people might you make statements about using this data set?
14. What can the sample of 30 Year 8 NZ boys tell you about all Year 8 NZ boys’ right foot lengths?
15. How confident are you that a random sample can tell you something about all Year 8 NZ boys’ right foot lengths? [population concept] How close will the average length of the sample be to the “real one”?
16. How big does the sample have to be to tell you something about all Year 8 NZ boys’ right foot lengths?
17. What is the population for this data set?
18. Ask about responses to Question 2 [in task F]. Particularly ask: Why did you use the median/UQ/range to make your claim? Why did you use the UQ and not the medians? Can you suggest on what basis Emma made her claim?
19. If another person took a *random sample* of 30 Year 8 NZ boys and a *random sample* of 30 Year 8 NZ girls, would their graphs be the same? Why?
20. Sketch what their graph might look like. [Student to do on pre-test sheet]
21. Is it possible that they could claim that the right foot lengths of Year 8 NZ boys tend to be bigger than the right foot lengths of Year 8 NZ girls? Why?
22. Repeat Questions 19–21 several times to get an idea of the extent of sampling variability they are prepared for intuitively.
23. Probe responses to Question 3 [in task F] (see #2 above) including how confident they are in their claim and how close the average heights are to the “real ones”.
24. Comparing claims made in Questions 2 & 3 [in task F]: probe differences in responses. E.g. Why claim on UQ in Q2 and median in Q3? Which claim are they more confident about? Why?

*Extra questions for students who mention samples and populations*

25. Is Emma’s claim about the samples (the data collected) or the populations (the group from which the data sampled)? Explain.

## Appendices

26. In the database what do you think the population distributions would show? Sketch.  
What do you think your sample would show? Sketch.

### **Definitions – ask about student definitions**

27. List of words
- Population
  - Sample

### ***2011 Post-interview schedule***

#### **Questions about their basis for making a claim (task F)**

13. What does random sample mean?
14. Why was a sample taken? What is a sample? What will a sample tell you?
15. What “group” of people might you make statements about using this data set?
16. What can the sample of 30 Year 11 NZ boys tell you about all Year 11 NZ boys’ bag weights?
17. How confident are you that a random sample can tell you something about all Year 11 NZ boys’ bag weights? [population concept] How close will the average weight of the sample be to the “real one”?
18. How big does the sample have to be to tell you something about all Year 11 NZ boys’ bag weights?
19. What is the population for this data set?
20. Ask about responses to Question 2 [in task F] Particularly ask: Why did you use the median/UQ/range to make your claim? Why did you use the UQ and not the medians? Can you suggest on what basis Matt made his claim?
21. If another person took a *random sample* of 30 Year 11 NZ boys and a *random sample* of 30 Year 11 NZ girls, would their graphs be the same? Why?
22. Sketch what their graph might look like. [Student to do on post-test sheet.]
23. Is it possible that they could claim that the bag weights of Year 11 NZ boys tend to be bigger than the bag weights of Year 11 NZ girls? Why?
24. Repeat Questions 21–23 several times to get an idea of the extent of sampling variability they are prepared for intuitively.
25. How confident are you in making/not making the claim?

## Appendices

26. Question 3 [in task F]: looking at bag weights, what other things could Matt investigate? What factors might make a difference to bag weights other than gender?
27. Probe responses to Question 4 [in task F]. In particular, ask: Why did you use the median/UQ/range to make your claim? Why did you use the UQ and not the medians? Can you suggest on what basis Matt made his claim?
28. If another person took a *random sample* of 30 Year 9 NZ boys and a *random sample* of 30 Year 9 NZ girls, would their graphs be the same? Why?
29. Sketch what their graph might look like. [Student to do on post-test sheet.]
30. Is it possible that they could claim that the dancing ratings of Year 9 NZ girls tend to be higher than the dancing ratings of Year 9 NZ boys? Why?
31. Repeat Questions 28–31 several times to get an idea of the extent of sampling variability they are prepared for intuitively ...
32. ... and how confident they are in their claim and how close the average ratings are to the “real ones”.
33. Comparing claims made in Questions 2 and 4 [in task F]: probe differences in responses. E.g. Why claim on UQ in Q2 and median in Q4? Which claim are they more confident about? Why?
34. Ask about Question 5 [in task F] responses.

### *Extra questions for students who mention samples and populations*

35. Is Matt’s claim about the samples (the data collected) or the populations (the group from which the data sampled)? Explain
36. In the database what do you think the population distributions would show? Sketch. What do you think your sample would show? Sketch.

### **Definitions – ask about student definitions**

37. List of words
  - a. Population
  - b. Sample



## **Appendix E – Investigative questions in assessment and curriculum materials**

### ***Assessment material***

Posing an *investigative* question is explicitly stated in:

- AS91264 Mathematics and Statistics 2.9: Use statistical methods to make an inference (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2012/as91264.pdf>)
- AS91265 Mathematics and Statistics 2.8: Conduct an experiment to investigate a situation (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2012/as91265.pdf>)
- AS91582 Mathematics and Statistics 3.10: Use statistical methods to make a formal inference ([http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2013/as91582 .pdf](http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2013/as91582.pdf))
- AS 91583 Mathematics and Statistics 3.11: Conduct an experiment to investigate a situation using experiment design principles (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2013/as91583.pdf>).

Posing a question is the requirement, *investigative* question is implied in:

- AS91035 Mathematics and Statistics 1.10: Investigate a multivariate data set using the statistical enquiry cycle (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2011/as91035.pdf>)
- AS91036 Mathematics and Statistics 1.11: Investigate bivariate numerical data using the statistical enquiry cycle (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2011/as91036.pdf>)
- AS91581 Mathematics and Statistics 3.9: Investigate bivariate measurement data (<http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2013/as91581.pdf>).

### ***Curriculum material***

Three achievement objectives at curriculum levels 6–8 (ages 15–18) make reference to posing investigative questions with a link at level 6 to material on posing investigative questions that is directly from this thesis. The achievement objectives are:

## Appendices

- S6-1: Plan and conduct investigations using the statistical enquiry cycle  
(<http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/Achievement-objectives-by-level/AO-S6-1>)
- S7-1: Carry out investigations of phenomena, using the statistical enquiry cycle  
(<http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/Achievement-objectives-by-level/AO-S7-1>)
- S8-1: Carry out investigations of phenomena, using the statistical enquiry cycle  
(<http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/Achievement-objectives-by-level/AO-S8-1>).

## Appendix F – Presentations on posing investigative questions

- Arnold, P. (2008) *Statistics education early 21st-century New Zealand*. Multiple key note presentations to secondary and primary teachers. November 2008–September 2009. Dunedin, Christchurch, Wellington (primary and secondary), Palmerston North, New Plymouth, Havelock North, Whakatane (BOP), Hamilton, Auckland (primary and secondary). <http://www.censusatschool.org.nz/2008/pip-statsed/>.
- Arnold, P. (2009, 11 June) *Posing investigative questions*. Workshop at HOD day, Auckland.
- Arnold, P. (2009, 16 June) *Statistics session for SNP teachers*. Included posing investigative questions. <http://aucksnp.wikispaces.com/2008+PCK+statistics+2009>.
- Arnold, P. (2009, 30 September) *Posing investigative questions*. Workshop at NZAMT conference, Palmerston North, September 2009. <http://nzstatsedn.wikispaces.com/NZAMT+11#L2>.
- Arnold, P. (2009, 13 October) Part of workshop for teachers at Farm Cove Intermediate, Auckland. <http://nzstatsedn.wikispaces.com/Farm+Cove+Intermediate>.
- Arnold, P. (2009) Lesson plan for Alfriston College <http://nzstatsedn.wikispaces.com/lesson+plan+1> (see posing investigative questions). This draft lesson plan became an activity on CensusAtSchool <http://www.censusatschool.org.nz/classroom-activities/posing-questions/>.
- Arnold, P. (2011, 4 March) *Statistics level 5*. Whole-day workshop for Gisborne teachers. <http://nzstatsedn.wikispaces.com/Gisborne> (see section on sorting questions).
- Arnold, P. (2011, 10 & 17 May). *Using Fathom in statistics at levels 5 and 6*. (Included posing investigative questions). In-service training for secondary teachers at St Cuthbert's College, Auckland.
- Arnold, P. (2011, 7 July). *Statistics level 5 and 6*. (Included posing investigative questions.) Afternoon workshop in Christchurch for maths teachers..
- Arnold, P. (2011, 21 July). *Posing questions*. NZAMT 12 workshop. <http://nzstatsedn.wikispaces.com/NZAMT+12#L2>.
- Arnold, P. (2011, 22 July). *Statistics level 5 and 6*. (Included posing investigative questions.) Afternoon workshop for Otago maths teachers.

## Appendices

- Arnold, P. (2012, 28 September). *What is the story? Being a data detective*. (Included posing investigative questions.) Key note presentation at Napier Regional Mathematics Symposia (primary and intermediate teachers).
- Arnold, P. (2013, 19, 22 & 23 April). *Statistical investigations level 4*. Workshop presentations at Palmerston North, Napier and Wellington primary teachers symposia. (Included posing investigative questions.)

## **Appendix G – Teaching and Learning Research Initiative (TLRI) papers and presentations**

Included are all papers and presentations by members of the TLRI team.

### ***Journal articles***

- Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education, 19*(2), 1–32.
- Pfannkuch, M., Regan, M., Wild, C. J., & Horton, N. (2010). Telling data stories: essential dialogues for comparative reasoning. *Journal of Statistics Education, 18*(1), 1–38.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society), 174*(2), 247–295. doi: 10.1111/j.1467-985X.2010.00678.x

### ***Conference papers***

- Arnold, P., & Pfannkuch, M. (2010, July). Enhancing students' inferential reasoning: from hands on to "movie snapshots". In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Pfannkuch, M. (2010, July). Inferential reasoning: learning to "make a call" in practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2010, July). Inferential reasoning: learning to "make a call" in theory. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.

### ***Keynote presentations***

#### **2009 Keynote presentations**

- Wild, C., & Pfannkuch, M. (2009, February). *Building inferential reasoning in statistics*. Keynote presentation at the National Numeracy Conference, Auckland.  
Their recorded talks, slides and handouts are at:  
<http://www.censusatschool.org.nz/2009/informal-inference/>

## 2010 Keynote presentations

Wild, C., Pfannkuch, M., Regan, M., & Horton, N. (2010, October). *Towards more accessible conceptions of statistical inference*. Paper (read by Chris Wild) presented at World Statistics Day, Royal Statistical Society launch of the UK 10-year statistical literacy campaign, Royal Statistical Society London Headquarters, England. The paper will be published along with discussions in Series A of the *Journal of the Royal Statistical Society*.

Pfannkuch, M., Regan, M., & Wild, C. (2010, October). *Generalising from data*. Keynote presentation at the CensusAtSchool Day, Royal Statistical Society, London, England.

Wild, C. (2010, October). *Laying foundations for inference*. Keynote presentation at the CensusAtSchool Day, Royal Statistical Society, London, England.

Pfannkuch, M., Regan, M., & Wild, C. (2010, October). *Generalising from data*. Keynote presentation at the getStats Week, Plymouth University, Plymouth, England.

Wild, C. (2010, October). *Laying foundations for inference*. Keynote presentation at the getStats Week, Plymouth University, Plymouth, England.

## 2011 Keynote presentations

Arnold, P. (2011, July). *What is the story?* Keynote presentation at the 12th New Zealand Association of Mathematics Teachers Conference, Dunedin, New Zealand

Arnold, P. (2011, September). *What is the story?* Keynote presentation to the Auckland Mathematics Association, Auckland, New Zealand

Pfannkuch, M. (2011, July). *Developing secondary students' inferential reasoning in statistics*. Keynote presentation at the 51st Hungarian Annual Teachers Conference (51. László Rátz), Komarom, Hungary.

Pfannkuch, M. (2011, July). *Engaging with shape*. Keynote presentation at the 51st Hungarian Annual Teachers Conference (51. László Rátz), Komarom, Hungary.

Pfannkuch, M. (2011, July). *Developing students' inferential reasoning in statistics: From hands-on to "movies"*. Keynote presentation to Ludwigsburg University of Education, Germany.

***Statistics teachers' days***

**2009 Statistics teachers' day**

**Auckland Mathematical Association Statistics Teachers Day**, 24 November 2009, Auckland.

The purpose of the day was to disseminate the research findings of the project to all teachers. Attended by more than 150 teachers. See: <http://www.censusatschool.org.nz/2009/informal-inference/>

**Two Plenary Talks:**

Pfannkuch, M., Arnold, P., & Regan, M. *NCEA 2011: Laying foundations for inference in 2010*

Pfannkuch, M., Arnold, P. *NCEA 2011: Laying foundations for inference in 2010 Wrap Up*

**Two Workshops:**

- Workshop 1: *Building conceptions of populations and samples and the connections between*
- Workshop 2: *Introduction to making a call using box plots*

These one-hour workshops were run by members of the TLRI project team in five parallel sessions and were attended by all teachers. Members of the team who were involved in the delivery of the workshops were M. Pfannkuch, P. Arnold, J. Florence, J. Ellwood, L. Smith, S. Wright, A. Martin, M. McFarland and P. Doyle.

**Wellington Mathematics Association Teacher Day**, 20 November 2009

Arnold, P. *Building conceptions of populations and samples and the connections between, and Introduction to making a call using box plots*. Ninety-minute workshop in two parts (repeated).

**Bay of Plenty Mathematics Association Teacher Day**, 3 December 2009

**Plenary:**

Arnold, P. *NCEA 2011: Laying foundations for inference in 2010*.

**Two Workshops:**

- Workshop 1: *Building conceptions of populations and samples and the connections between.*
- Workshop 2: *Introduction to making a call using box plots.*

## Appendices

These one-hour workshops were run by members of the TLRI project team in three parallel sessions and were attended by all teachers. Members of the team who were involved in the delivery of the workshops were P. Arnold, J. Ellwood and P. Doyle.

### **2010 Statistics teachers' day**

#### **Auckland Mathematical Association Statistics Teachers Day, 2 December 2010**

The purpose of the day was to disseminate the research findings of the project to all teachers. Attended by more than 200 teachers. Interest was very high and registrations had to close at 200 because of facility limits.

See: [www.censusatschool.org.nz/2009/informal-inference/](http://www.censusatschool.org.nz/2009/informal-inference/) for plenary hand-out and workshop resource material.

#### **Plenary (90 minutes):**

Pfannkuch, M., Arnold, P., Wright, S., Florence, J., Ellwood, J., Smith, L., & McFarland, M.  
*Developing inferential reasoning: Stories from Statistics classes.*

#### **Seven Year 11 Workshops and One Year 10 Workshop (75 minutes):**

These workshops were run by members of the TLRI project team in eight parallel sessions and were attended by all teachers. Members of the team who were involved in the delivery of the workshops were M. Pfannkuch, P. Arnold, J. Florence, J. Ellwood, L. Smith, S. Wright, M. McFarland, J. Saunders, M. Regan and P. Doyle.

### ***Workshop presentations***

#### **2009 Workshop presentations**

Pfannkuch, M. (2009, August). *Building students' inferential reasoning in statistics: Research in progress*. Presentation to the Mathematics Education Unit Seminar, University of Auckland, Auckland.

#### **2010 Workshops presentations**

Arnold, P. (2010, March). *Year 11 making the call*. Workshop conducted with the mathematics department of the trial school.

Arnold, P. (2010, March). Auckland Mathematics Association, Auckland.

- Workshop 1: *Building conceptions of populations and samples and the connections between*
- Workshop 2: *Introduction to making a call using box plots*



## Appendices

Arnold, P. (2010, June). *Making the call*. Workshop conducted with the Secondary Numeracy In-school facilitators training, Auckland.

Arnold, P. (2010, September). *Making the call*. Workshop conducted with the Manurewa High School mathematics department, Auckland.

Pfannkuch, M. (2010, October). *Laying foundations for inference*. Presentation to the Mathematics Education Unit Seminar, University of Auckland, Auckland.

Pfannkuch, M., & Regan, M. (2010, October). Two workshops presented at the CensusAtSchool Day, Royal Statistical Society, London, England.

- Workshop 1: *Engaging in Shape*
- Workshop 2: *Making a claim*

Pfannkuch, M. & Regan, M. (2010, October). Two workshops presented at the getStats Week, Plymouth, England.

- Workshop 1: *Engaging in Shape*
- Workshop 2: *Making a claim*

Arnold, P. (2010, December). *Enhancing students' inferential reasoning: From hands-on to "movie snapshots"*. Presentation to the Cognition Education Seminar, Auckland.

### **2011 Workshop presentations**

P. Arnold gave the following workshop presentations in 2011:

- *Year 10 making the call*. To Gisborne mathematics teachers, 4 March.
- *Year 11 Inference*. To Auckland Mathematics Association, 12 March.
- *Year 11 inference with Fathom*. To St Cuthbert's Mathematics Department, 10 & 17 May.
- *Year 10 making the call*. To Christchurch mathematics teachers, 7 July.
- *Year 10 making the call*. To Otago mathematics teachers, 22 July.

## Appendix H – Lessons on shape

Figure H-1 is split into two columns. The first column outlines the lesson detail that was developed for the teachers to use in class (all the year 10 mathematics teachers at the school used the same lesson plans as the teacher being observed for the statistics topic). The second column describes the background to or purpose of the particular aspect of the activity. Also noted is the aspect of distribution that was being attended to: (1) the notion of distribution, (2) shape of distributions, (3) predicting distributions, and (4) contextual knowledge.

| Lesson number | Activity detail                                                                                                                                                                                                                                  | Background/purpose to the particular aspect of the activity                                                                                                                                                                                                                                                                                     |
|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lesson 2      | 1. Students make 15 “squares” of paper from an A5 sheet. Get students to label their 15 “squares” of paper with the numbers 1–15 in the top left hand corner. This is to help with identification later on.                                      | Organisation/preparation.                                                                                                                                                                                                                                                                                                                       |
|               | 2. Using the prepared PowerPoint presentation, show each of the 15 graphs for a very short time, 1-2 seconds and get students to sketch the shape they see in the quick glimpse. See Figure H-2 for the 15 graphs.                               | Looking at the gross shape of the data rather than specific detail. Use bigger sample sizes so that the shape was clearer. Large population/sample to help make the shape more “obvious”.<br><br><b>(2) shape of distributions</b>                                                                                                              |
|               | 3. When all 15 graphs are drawn students should check with their neighbour and compare what they have sketched for each graph. At this stage the teacher can put up the 15 “teacher” sketched shapes and they can compare against these as well. | Students to compare their sketches with the view that they might decide that one was better than the other, and also compare this with the teacher graph and what the teacher graph might offer that theirs doesn’t. A more generalised shape. See Figure H-3 for the teacher graph (sketched shapes).<br><br><b>(2) shape of distributions</b> |
|               | 4. Get pairs of students to sort one set of graphs into similar shapes. Collate responses from the class and arrive at a consensus as to which shapers are similar. Use the teacher shapes (Figure H-3) on the board.                            | Grouping the shapes was about trying to see the patterns that are there. Generally statistical graphs fall into a limited group of patterns, there is not an infinite number of patterns.<br><br><b>(2) shape of distributions</b><br><br><i>Notes: Symmetric LS RS uniform</i>                                                                 |
|               | 5. For each group get students to describe the shape they see using words that they are comfortable with. Note these words under each of the groups of graphs.                                                                                   | Starting with student language for the shapes that they see to give a foundation for the development of statistical language.<br><br><b>(2) shape of distributions</b>                                                                                                                                                                          |

## Appendices

| Lesson number | Activity detail                                                                                                                                                                                                                                                                                                                                             | Background/purpose to the particular aspect of the activity                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | 6. Introduce the statistical words used for describing graphs – teacher prepared resources. Have a good discussion with the students about what they think the different words mean both in everyday and statistical sense. Get the students to suggest which words might best go with which group of graphs.                                               | Gave students the language to see what they do with the language, see how much of it is intuitive.<br><br><b>(2) shape of distributions</b><br><br><i>Then have the conversation with at the end about what the words mean.</i>                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Lesson 3      | 7. Hand out strips of graphs (Figure H-2). Get students to cut and paste the graph and their sketch into their book under each of the description words. Allow room for the variable, justification, other examples and the description. Suggested layout below. Need about six pages in double spread. This will become a reference resource for students. | <br>Organisation of graphs, but also to start a “library” of contexts that are similar shapes. Building their contextual knowledge library.<br><br><b>(4) contextual knowledge</b>                                                                                                                                                                                                                                                                                                                                                                             |
|               | 8. Put up the list of variables that made the graphs. Before students match them with the graphs get them to predict what shape they think the graph of the variables will be and why. Discuss as a class. Collect ideas on the board.                                                                                                                      | Consideration of what the data might look like; want students to think about data before they sketch it. That is to get students to think about what might be sensible values for a particular variable. The prediction is also about thinking about the shape of the data and using contextual knowledge to decide on what the shape might be.<br><br>Understanding when data is incorrect, cleaning data. Getting the students to start to think about the context a bit more, building their contextual knowledge.<br><br><b>(3) predicting distributions</b><br><i>Note: E.g. right foot length, reaction time, attendance, birth months</i> |
|               | 9. Get students to match the context with the graph – get them to use the mix and match labels initially and record the final context in their book with their justification. Add the variable and the unit to the graph.                                                                                                                                   | Organisation, but also using their contextual knowledge and information from previous activity to see what makes sense.<br><br><b>(4) contextual knowledge</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|               | 10. Once this is finished get students to look back at their graphs from the previous lesson and decide which “shape” they are. Add these contexts in the appropriate space.                                                                                                                                                                                | Organisation, but also building their contextual knowledge library for different variables with the same shape.<br><br><b>(2) shape of distributions</b><br><b>(4) contextual knowledge</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

| Lesson number | Activity detail                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Background/purpose to the particular aspect of the activity                                                                                                                                                                                                                                                                                                     |
|---------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Lesson 4      | <p>A. Review activity: Mix and match – statistical graphs and shape descriptors</p> <p>Resource: mix and match activity – shape descriptors</p> <ul style="list-style-type: none"> <li>Students place the statistical graphs under one of the headings. There may be different numbers of graphs under each of the headings.</li> <li>Add the contexts (and paste the graph) to the other examples in the work done previously.</li> </ul>                                                                                                                                                                                                                                                                                          | <p>Further work on deciding on shapes. Adds to the “library”. Opportunity to check use of language, especially with skewed graphs.</p> <p><b>(2) shape of distributions</b><br/><b>(4) contextual knowledge</b></p>                                                                                                                                             |
|               | <p>B. Describing distributions</p> <p>Discuss with students what key features of a graph to describe are.</p> <ul style="list-style-type: none"> <li>Put the challenge out if they had to draw the graph from the description only what info would they need. Collect ideas from the class. <ul style="list-style-type: none"> <li>Suggest may be: shape, description of range, median/centre, middle group, and peak(s) – there may be other features, discuss as a department first.</li> </ul> </li> </ul>                                                                                                                                                                                                                       | <p>About developing what makes a good description.</p> <p><b>(1) notion of distribution</b><br/><b>(2) shape of distributions</b><br/><b>(4) contextual knowledge</b></p>                                                                                                                                                                                       |
|               | <ul style="list-style-type: none"> <li>Model for #9 and #4. Model this process for the students. <ul style="list-style-type: none"> <li>Talk out loud your thinking and get them to contribute.</li> <li>Eg. What shape is the graph? Write the first sentence explaining the use of approximately and the use of the variable and the group we are talking about.</li> <li>What values do the heights range from and to?</li> <li>Write the next sentence and so on. The questions should be around the features you decided on with the class.</li> <li>Remember to include the CONTEXT. <b>Variable, values and units.</b></li> <li>Use active reflection, i.e. making descriptions correct and complete.</li> </ul> </li> </ul> | <p>Modelling the thinking process for students when writing a description, also modelling the language to use, including the intertwining of the context throughout the description including especially the variable, values and units.</p> <p><b>(1) notion of distribution</b><br/><b>(2) shape of distributions</b><br/><b>(4) contextual knowledge</b></p> |

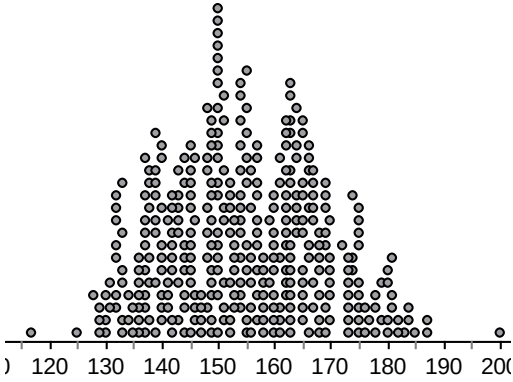
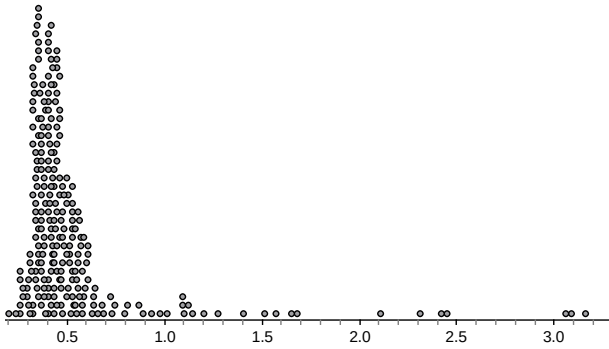
| Lesson number | Activity detail                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Background/purpose to the particular aspect of the activity                                                                                                                                                                                                                                                                      |
|---------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|               | <p>Examples:</p> <p>#9 Graph is: heights in cm of Yr 5–10 students</p>  <p>The distribution of heights for these year 5–10 students is approximately symmetrical and unimodal. The heights range from 116 cm to 200 cm. The median height is about 155 cm and the middle group of heights is between 142 cm and 167 cm.</p> <p>#4 Graph is: reaction times in secs of yr 4–13 students</p>  <p>The distribution of reaction times for these yr 4–13 students is right skewed. Nearly all of the reaction times are tightly bunched between 0.2 and 0.6 secs. There are some reaction times slower than 0.6 secs and they spread out to 3.15 secs. The graph of reaction times peaks at about 0.4 secs and is approximately symmetrical between 0.2 and 0.6 secs.</p> |                                                                                                                                                                                                                                                                                                                                  |
|               | <ul style="list-style-type: none"> <li>Students to do rest of the descriptions for homework, one per night over the next few weeks. Review these at the beginning of the following lesson, remembering to model good practice (see above).</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    | <p>To continue to develop their descriptive skills over the whole unit of work, to keep the focus in this area, and to provide plenty of practice at writing descriptions, a new skill to be developed.</p> <p><b>(1) notion of distribution</b><br/> <b>(2) shape of distributions</b><br/> <b>(4) contextual knowledge</b></p> |

Figure H-1. Detailed lesson planning for teaching experiment four

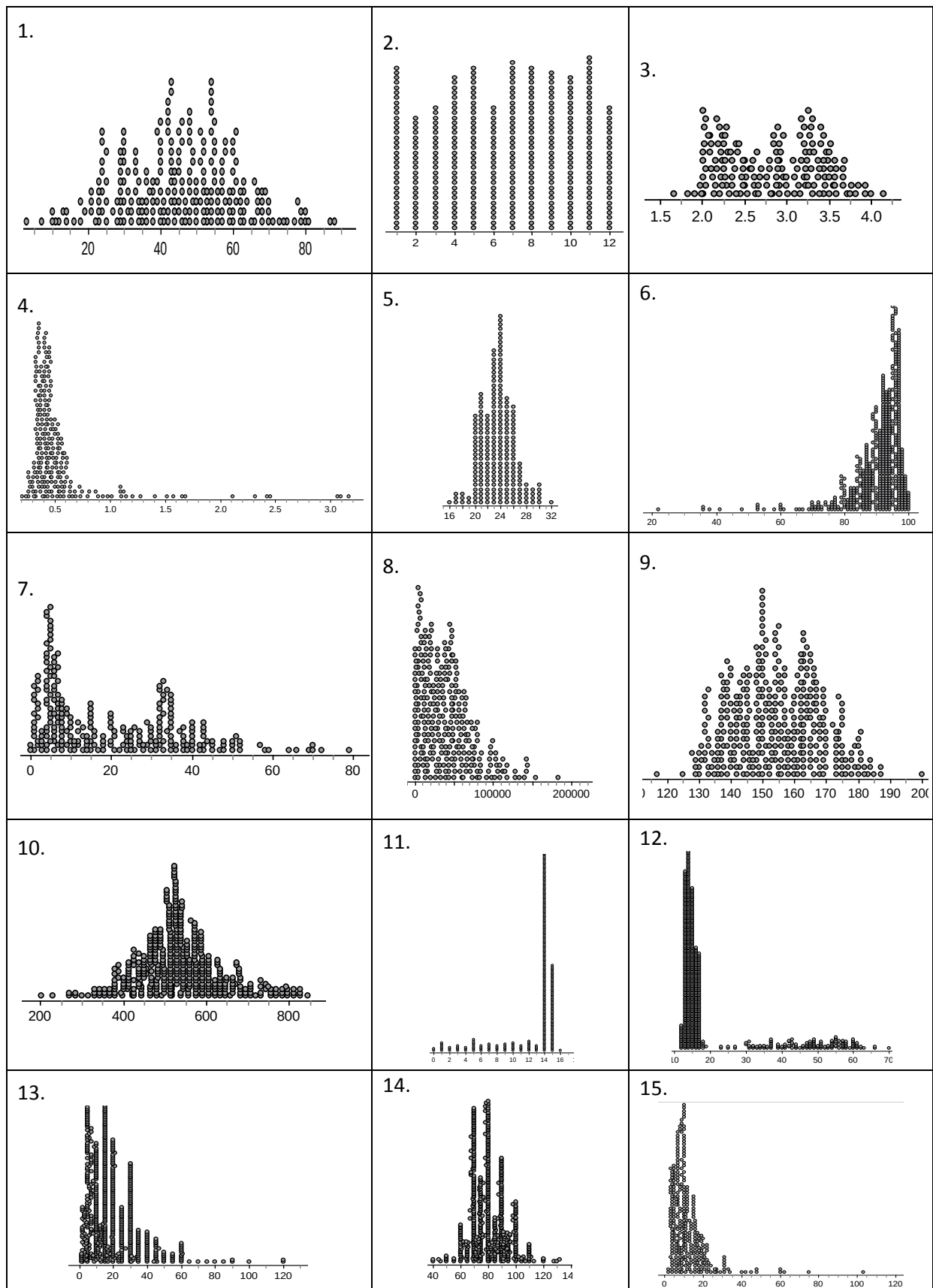


Figure H-2. Actual graphs used in shape lessons

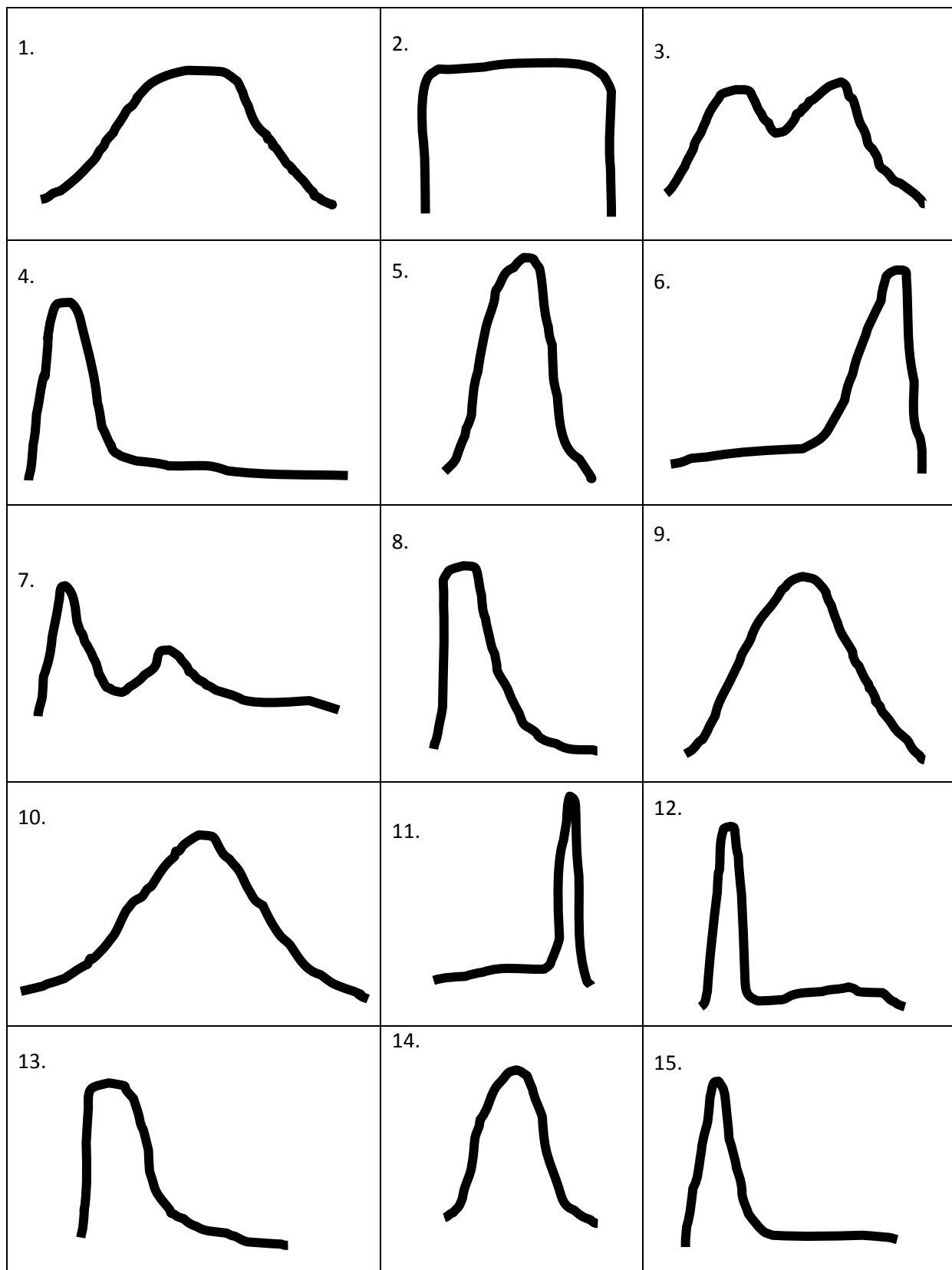


Figure H-3. Shape sketches used in shape lessons

## References

- Afflack, R. (1982). *Beyond equals: To encourage the participation of women in mathematics*. Berkeley, CA: Regents, University of California.
- Ainley, J., Nardi, E., & Pratt, D. (2000). The construction of meanings for trend in active graphing. *International Journal of Computers for Mathematical Learning*, 5, 85–114. doi: 10.1023/A:1009854103737
- Arnold, P. (2008a). Developing new statistical content knowledge with secondary school mathematics teachers. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference (July, 2008), Monterrey, Mexico* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Arnold, P. (2008b, July). *What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation*. Paper presented at the 11th International Congress on Mathematical Education (ICME-11), Monterrey, Mexico. <http://tsg.icme11.org/document/get/481>
- Arnold, P. (2009). Context and its role in posing investigative questions. In K. Makar (Ed.), *The Role of Context and Evidence in Informal Inferential Reasoning. Proceedings of the Sixth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-6, July, 2009), Brisbane, Australia* (pp. 101–115). Brisbane, Australia: University of Queensland.
- Arnold, P., & Pfannkuch, M. (2010). Enhancing students' inferential reasoning: from hands on to "movie snapshots". In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8, July, 2010), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Arnold, P., & Pfannkuch, M. (2012). The language of shape. *Proceedings of the 12th International Congress on Mathematical Education (ICME-12, July, 2012), Seoul, South Korea* (pp. 2446–2455). Seoul, South Korea: ICME-12.
- Arnold, P., Pfannkuch, M., Wild, C. J., Regan, M., & Budgett, S. (2011). Enhancing students' inferential reasoning: From hands-on to "movies". *Journal of Statistics Education*, 19(2), 1–32.
- Aspinwall, L., Haciomeroglu, E., & Presmeg, N. (2008). Students' verbal descriptions that support visual and analytic thinking in calculus. In O. Figueras, J. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), *Proceedings of the Joint Meeting of PME 32 and PME-NA 30 (July, 2008) (Vol. 2, pp. 97–104)*. Morelia, Mexico: Cinvestav-UMSNH.
- Bakker, A. (2004a). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: Freudenthal Institute.
- Bakker, A. (2004b). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64–83.



## References

- Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147–168). Dordrecht, The Netherlands: Kluwer.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407. doi: 10.1177/0022487108324554
- Banister, P., Burman, E., Parker, I., Taylor, M., & Tindall, C. (1994). *Qualitative methods in psychology : a research guide*. Buckingham, England: Open University Press.
- Barton, D. (2006). *Sigma mathematics: NCEA level 3 statistics and modelling*. Auckland, New Zealand: Pearson Education
- Ben-Zvi, D., & Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–22). Brisbane, Australia: University of Queensland.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35–65. doi: 10.1023/A:1013809201228
- Ben-Zvi, D., Gil, E., & Apel, N. (2007). What is hidden beyond the data? Helping young students to reason and argue about some wider universe. In D. Pratt & J. Ainley (Eds.), *Reasoning about statistical inference: Innovative ways of connecting chance and data. Proceedings of the Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5, August, 2007), University of Warwick, England* (pp. 1–26). Warwick, England: University of Warwick.
- Biehler, R. (1997). Students' difficulties in practicing computer-supported data analysis: Some hypothetical generalizations from results of two exploratory studies. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics. Proceedings of the International Association for Statistical Education Round Table Conference (July, 1996), Granada, Spain* (pp. 169–190). Voorburg, The Netherlands: International Statistical Institute.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York, NY: Academic Press.
- Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in the biological sciences. In C. Batanero (Ed.), *Training researchers in the use of statistics* (pp. 215–226). Granada, Spain: International Association for Statistical Education and International Statistical Institute.
- Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, 14, 325–341. doi: 10.1016/j.learninstruc.2004.06.006
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, 2(2), 141–178. doi: 10.1207/s15327809jls0202\_2
- Bruner, J. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

## References

- Burgess, T. (2007). *Investigating the nature of teacher knowledge needed and used in teaching statistics* (Doctoral thesis.) Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/07.Burgess.Dissertation.pdf>
- CensusAtSchool. (2008). Data detective poster. Retrieved from <http://new.censusatschool.org.nz/resource/data-detective-poster/>
- Clark, J., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210. doi: 10.1007/BF01320076
- Cobb, P. (2000a). Conducting teaching experiments in collaboration with teachers. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 307–333). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cobb, P. (2000b). The importance of a situated view of learning to the design of research and instruction. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 45–82). Westport, CT: Ablex.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32, 9–13. doi: 10.3102/0013189X032001009
- Cobb, P., & Gravemeijer, K. (2008). Experimenting to support and understand learning processes. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 68–95). New York, NY: Routledge.
- Cohen, L., & Manion, L. (1994). *Research methods in education* (4th ed.). London, England: Routledge.
- Cohen, L., Morrison, K., & Manion, L. (2007). Research methods in education [electronic resource]. Retrieved from <http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=193801>
- Confrey, J., & Lachance, A. (2000). Transformative teaching experiments through conjecture-driven research design. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 230–265). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cozby, P. C. (2009). *Methods in behavioral research*. Boston, MA: McGraw-Hill.
- Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Davies, J. (2010). Preparation and process of qualitative interviews and focus groups. In L. Dahlberg & C. McCaig (Eds.), *Practical Research and Evaluation: A Start-to-finish Guide for Practitioners* (pp. 126–144). London, England: Sage.
- delMas, R. (1997). A framework for the evaluation of software for teaching statistical concepts. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics. Proceedings of the International Association for Statistical Education Round Table Conference (July, 1996), Granada, Spain* (pp. 75–90). Voorburg, The Netherlands: International Statistical Institute.
- delMas, R. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 79–95). Dordrecht, The Netherlands: Kluwer.

## References

- delMas, R., Garfield, J., & Ooms, A. (2005). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–17). Brisbane, Australia: University of Queensland.
- Department of Education. (1986). *Mathematics form 7 teachers guide: Mathematics with statistics*. Wellington, New Zealand: Longman Paul Limited.
- Department of Education. (1987). *Mathematics: Forms 1 to 4*. Wellington, New Zealand: V. R. Ward, Government Printer.
- Downie, D., Slesnick, T., & Stenmark, J. K. (1981). *Math for girls and other problem solvers*. Berkeley, CA: Regents, University of California.
- Dunkels, A. (1991). Interplay of number concept and statistics using attitudes and techniques of EDA. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics (ICOTS-3, July, 1991), Dunedin, New Zealand* (Vol. 1, pp. 129–139). Voorburg, The Netherlands: International Statistical Institute.
- Ellerton, N. F., & Clements, M. A. (1997). Pencil-and-paper mathematics tests under the microscope. In F. Biddulph & K. Carr (Eds.), *People in mathematics education. Proceedings of the Twentieth Annual Conference of the Mathematics Research Group of Australasia (MERGA-20, July, 1990), Rotorua, New Zealand* (Vol. 1, pp. 155–162). Rotorua, New Zealand: Mathematics Education Research Group of Australasia.
- EQUALS, & California Mathematics Council. (1989). *Assessment alternatives in mathematics*. Berkeley, CA: Regents, University of California.
- Ernest, P. (1998). A postmodern perspective on research in mathematics education. In A. Sierpiska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity: An ICMI study* (Vol. 1, pp. 71–85). Dordrecht, The Netherlands: Kluwer Academic.
- Fernie, S., & Smith, K. (2010). Action research. In L. Dahlberg & C. McCaig (Eds.), *Practical Research and Evaluation: A Start-to-finish Guide for Practitioners*. London, England: Sage.
- Finzer, W. (2007). *Fathom dynamic data software version 2 [computer software]*. Emeryville, CA: Key Curriculum Press.
- Fontana, A., & Frey, J. (1994). Interviewing. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 361–376). Thousand Oaks, CA: Sage.
- Fontana, A., & Frey, J. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 695–727). Thousand Oaks, CA: Sage.
- Fowler, F. J., & Cosenza, C. (2009). Design and evaluation of survey questions. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (2nd ed., pp. 375–412). Thousand Oaks, CA: Sage.
- Franke, M., Webb, N., Chan, A., Ing, M., Freund, D., & Battey, D. (2009). Teaching questioning to elicit students' mathematical thinking in elementary school classrooms. *Journal of Teacher Education*, 60(4), 380–392. doi: 10.1177/0022487109339906

## References

- Franklin, C., & Garfield, J. (2006). The GAISE Project. Developing statistics education guidelines for grades pre-K–12 and college courses. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 345–375). Reston, VA: National Council of Teachers of Mathematics.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Schaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association.
- Friel, S. N., & Bright, G. W. (1998). Teach-Stat: A model for professional development in data analysis and statistics for teachers K–6. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 89–117). Mahwah, NJ: Lawrence Erlbaum Associates.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158.
- Gardiner, W., & Parata, H. (2007). He aratohu: Integrating Kaupapa Māori into Mainstream Secondary School Teaching and Learning Programmes. Retrieved from <http://tetereauraki.tki.org.nz/content/download/651/3185/file/He%20Aratohu.pdf>
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396. doi: 10.1111/j.1751-5823.2007.00029.x
- Gay, L., Mills, G., & Airasian, P. (2012). *Educational research: Competencies for analysis and applications* (10th ed.). Boston, MA: Pearson.
- Glencross, M., & Mji, A. (2001). The role of a research resource centre in the training of social science researchers. In C. Batanero (Ed.), *Training researchers in the use of statistics*. Granada, Spain: International Association for Statistical Education, and International Statistical Institute.
- Graham, A. (2006). *Developing Thinking in Statistics*. London, England: Paul Chapman.
- Gravemeijer, K. (1998). Developmental research as a research method. In A. Sierpiska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity: An ICMI study* (pp. 277–295). Dordrecht, The Netherlands: Kluwer Academic.
- Greer, B. (2009). Representational flexibility and mathematical expertise. *ZDM – The International Journal on Mathematics Education*, 41(5), 697–702. doi: 10.1007/s11858-009-0211-7
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Thousand Oaks, CA: Sage.
- Haciomeroglu, E., Aspinwall, L., & Presmeg, N. (2010). Contrasting cases of calculus students' understanding of derivative graphs. *Mathematical Thinking and Learning*, 12, 152–176. doi: 10.1080/10986060903480300
- Hancock, C., Kaput, J., & Goldsmith, L. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 2(3), 337–364. doi: 10.1207/s15326985ep2703\_5
- Hjalmarson, M., & Lesh, R. A. (2008). Engineering and design research. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education:*

## References

- Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 96–110). New York, NY: Routledge.
- Hook, P., & Mills, J. (2011). *SOLO taxonomy: A guide for schools*. Laughton, England: Essential Resources.
- Huberman, M., & Miles, M. (1994). Data management and analysis methods. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428–444). Thousand Oaks, CA: Sage.
- Kelly, A. E., & Lesh, R. A. (Eds.). (2000). *Handbook of research design in mathematics and science education*. Mahwah, NJ: Lawrence Erlbaum.
- Kemmis, S., & McTaggart, R. (2005). Participatory action research – Communicative action and the public sphere. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3rd ed., pp. 559–603). Thousand Oaks, CA: Sage.
- Kieran, C. (1998). Models in mathematics education research: a broader view of research results. In A. Sierpiska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity: An ICMI study* (pp. 213–225). Dordrecht, The Netherlands: Kluwer Academic.
- Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Shifter, & V. Bastable (Eds.), *Developing mathematical ideas: Collecting, representing, and analyzing data* (pp. 165–201). Parsippany, NJ: Dale Seymour.
- Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Shifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics.
- Konold, C., Higgins, T., Russell, S. J., & Khalil, K. (2004). *Data seen through different lenses*. (Unpublished manuscript.) Scientific Reasoning Research Institute, University of Massachusetts, Amherst, MA, and TERC, Cambridge, MA.
- Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, 2(1), 1–37.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259–289.
- Langbein, L., & Felbinger, C. L. (2006). Public program evaluation [electronic resource]: A statistical guide Retrieved from <http://site.ebrary.com.ezproxy.auckland.ac.nz/lib/auckland/docDetail.action?docID=10178081>
- Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction*, 14(1), 69–108.
- Lesh, R. A., & Kelly, A. E. (2000). Multitiered teaching experiments. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 197–230). Mahwah, NJ: Lawrence Erlbaum Associates.
- Lowe, I. (1991). *Mathematics at work: modelling your world* (Vol. 2). Canberra, ACT, Australia: Australian Academy of Science.
- MacKay, R., & Oldford, W. (1994). *Stat 231 course notes fall 1994*. Waterloo, Canada: University of Waterloo.

## References

- Makar, K. (2004). *Developing statistical inquiry: Prospective secondary mathematics and science teachers' investigations of equity and fairness through analysis of accountability data*. (Doctor of Philosophy dissertation.) The University of Texas, Austin.
- Makar, K., & Confrey, J. (2005). Using distributions as statistical evidence in well-structured and ill-structured problems. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–18). Brisbane, Australia: University of Queensland.
- Makar, K., & Rubin, A. (2009). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 8(1), 82–105.
- Mark, M., & Reichardt, C. (2009). Quasi-experimentation. In L. Bickman & D. J. Rog (Eds.), *The SAGE handbook of applied social research methods* (2nd ed., pp. 182–213). Thousand Oaks, CA: Sage.
- Marsden, E., & Torgerson, C. J. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education*, 36(5), 583–616. doi: 10.1080/03054985.2012.731208
- Marshall, C., & Rossman, G. B. (2006). *Designing qualitative research* (4th ed.). Thousands Oaks, CA: Sage.
- Martignon, L. (2008). Instructing future teachers of statistics: the benefits of en-active analogue modelling. In C. Batenero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching statistics in school mathematics. Challenges for teaching and teacher education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference (July, 2008), Monterrey, Mexico* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Mayer, R. E. (2009). *Multimedia Learning*. New York, NY: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2002). Animation as an aid to multimedia learning. *Educational Psychology Review*, 14(1), 87–99. doi: 10.1023/A:1013184611077
- McClain, K. (2005). The evolution of teachers' understandings of statistical data analysis: A focus on distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–17). Brisbane, Australia: University of Queensland.
- McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, 45, 103–129. doi: 10.1023/A:1013874514650
- McGillivray, H. (2007). *Stage one statistics lecture notes*. University of Technology, Queensland, Australia.
- McNiff, J., & Whitehead, J. (2006). *All you need to know about action research*. London, England: Sage.
- Meletiou-Mavrotheris, M., Lee, C., & Fouladi, R. (2007). Introductory statistics, college student attitudes and knowledge – a qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology*, 38(1), 65–83. doi: 10.1080/00207390601002765

## References

- Merton, R., Fiske, M., & Kendall, P. (1956/1990). *The focused interview* (2nd ed.). New York, NY: The Free Press.
- Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2005a). *Book 1: The number framework*. Wellington, New Zealand: Ministry of Education.
- Ministry of Education. (2005b). *Book 3: Getting started*. Wellington, New Zealand: Ministry of Education.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington, New Zealand: Learning Media.
- Ministry of Education. (2009a). Ka hikitia - managing for success: The Maori education strategy. Retrieved from <http://www.minedu.govt.nz/theMinistry/PolicyAndStrategy/KaHikitia.aspx>
- Ministry of Education. (2009b). Pasifika education plan 2009–2012. Retrieved from <http://www.minedu.govt.nz/NZEducation/EducationPolicies/PasifikaEducation/PasifikaEducationPlan.aspx>
- Ministry of Education. (2010a). *Senior secondary guides – mathematics and statistics: Glossary*. Retrieved from <http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Glossary/Glossary-page-S>
- Ministry of Education. (2010b). *Statistics: What has changed*. Retrieved from <http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/Statistics-What-has-changed>
- Moore, D. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.
- New Zealand Qualifications Authority. (2010). *Level 1 Mathematics and Statistics achievement standards*. Retrieved from <http://www.nzqa.govt.nz/ncea/assessment/search.do?query=Mathematics&view=all&level=01>
- Opie, C. (2004). Research procedures. In C. Opie & P. J. Sikes (Eds.), *Doing educational research: A guide to first-time researchers* (pp. 95–129). London, England: Sage.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27–45.
- Pfannkuch, M. (2008, July). *Building sampling concepts for statistical inference: A case study*. Paper presented at the 11th International Congress on Mathematical Education (ICME-11), Monterrey, Mexico. <http://tsg.icme11.org/document/get/476>
- Pfannkuch, M. (2010, July). Inferential reasoning: learning to “make a call” in practice. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*, Ljubljana, Slovenia (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.

## References

- Pfannkuch, M., Arnold, P., & Wild, C. J. (2011). *Statistics: It's reasoning not calculating*. (Summary research report on Building students' inferential reasoning: Levels 5 and 6) Retrieved from <http://www.tlri.org.nz/tlri-research/research-completed/school-sector/building-students-inferential-reasoning-statistics>
- Pfannkuch, M., Arnold, P., & Wild, C. J. (2012). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics* (accepted).
- Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Round Table* (pp. 204–218). Voorburg, The Netherlands: International Statistical Institute.
- Pfannkuch, M., Regan, M., & Arnold, P. (2009, November). *NCEA 2011: Laying foundations for inference in 2010*. Paper presented at the Statistics Teachers Day, Auckland, New Zealand. <http://new.censusatschool.org.nz/resource/laying-the-foundations-for-inference/>
- Pfannkuch, M., Regan, M., Wild, C. J., & Horton, N. (2010). Telling data stories: essential dialogues for comparative reasoning. *Journal of Statistics Education*, 18(1), 1–38.
- Pfannkuch, M., & Watson, J. (2003). Statistics education. In B. Perry, G. Anthony, & C. Diezmann (Eds.), *Research in mathematics education in Australasia 2000 – 2003* (pp. 265–289). Flaxton, Australia: Post Pressed.
- Pfannkuch, M., Wild, C., Horton, N., & Regan, M. (2009). *A teacher's guide to informal comparative reasoning*. Retrieved from <http://www.censusatschool.org.nz/2008/informal-inference/nzteachersguide-2009-08-04.pdf>
- Pfannkuch, M., & Wild, C. J. (2012). Laying foundations for statistical inference. *Proceedings of the 12th International Congress on Mathematical Education (ICME-12, July, 2012), Seoul, South Korea* (pp. 317–329). Seoul, South Korea: ICME-12.
- Pirie, S., & Kieran, T. (1989). A recursive theory of mathematical understanding. *For the Learning of Mathematics*, 9(3), 7–11.
- Pirie, S., & Kieran, T. (1994). Beyond metaphor: Formalising in mathematical understanding within constructivist environments. *For the Learning of Mathematics*, 14(1), 39–43.
- Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 3–4.
- Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and global thinking in statistical inference. *Statistics Education Research Journal*, 7(2), 107–129.
- Presmeg, N. (2006). Research on visualization in learning and teaching mathematics: Emergence from psychology. In P. Boero & A. Gutierrez (Eds.), *Handbook of research on the psychology of mathematics education: Past, present and future* (pp. 205–235). Rotterdam, The Netherlands: Sense.
- Punch, K. (2009). *Introduction to research methods in education*. Los Angeles, CA: Sage.
- Quinlan, C., Low, B., Sawyer, E., & White, P. (1993). *A concrete approach to algebra, units 1, 2, 3, & 4*. Sydney, NSW, Australia: Mathematical Association of New South Wales.



## References

- Radford, L. (2009). Why do gestures matter? Sensuous cognition and the palpability for mathematical meanings. *Educational Studies in Mathematics*, 70, 111–126. doi: 10.1007/s10649-008-9127-3
- Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: from a variation perspective. *Statistics Education Research Journal*, 5(2), 46–68.
- Ridgway, J., McCusker, S., & Nicholson, J. (2005). Uncovering and developing student statistical competences via new interfaces. In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Round table* (pp. 75–92). Voorburg, The Netherlands: International Statistical Institute.
- Roth, W.-M. (2005). *Doing qualitative research: Praxis of method*. Rotterdam, The Netherlands: Sense.
- Rubick, A. (2000). *The statistical thinking of twelve year 7 and 8 students* (Unpublished master's thesis). The University of Auckland, Auckland, New Zealand.
- Rubin, A., Bruce, B., & Tenny, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314–319). Voorburg, The Netherlands: International Statistical Institute.
- Rubin, A., Hammerman, J., Puttick, G., & Campbell, C. (2005). The effect of distributional shape on group comparison strategies. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–22). Brisbane, Australia: University of Queensland.
- Russell, S. J. (2006). What does it mean that “5 has a lot”? From the world to data and back. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth Yearbook* (pp. 17–29). Reston, VA: National Council of Teachers of Mathematics.
- Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257–270. doi: 10.1023/A:1023692604014
- Scheaffer, R. (2000). Statistics for a new century. In M. Burke & F. R. Curcio (Eds.), *Learning mathematics for a new century* (pp. 158–174). Reston, VA: National Council of Teachers of Mathematics.
- Schoenfeld, A. (2007). Methods. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 69–107). Charlotte, NC: Information Age.
- Schram, T. H. (2003). *Conceptualizing qualitative inquiry: Mindwork for fieldwork in education and the social sciences*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Schwartz, D., Chang, J., & Martin, L. (2008). Instrumentation and innovation in design experiments. In A. E. Kelly, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education: Innovations in science, technology, engineering, and mathematics learning and teaching* (pp. 45–67). New York, NY: Routledge.
- Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education*.

## References

- Proceedings of the Twentieth Annual Conference of the Mathematics Research Group of Australasia (MERGA-20, July, 1990), Rotorua, New Zealand* (Vol. 1, pp. 6–22). Rotorua, New Zealand: Mathematics Education Research Group of Australasia.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics* (Vol. 2, pp. 957–1009). Charlotte, NC: Information Age.
- Sikes, P. J. (2004). Methodology, procedures and ethical concerns. In C. Opie & P. J. Sikes (Eds.), *Doing educational research: A guide to first-time researchers* (pp. 15–33). London, England: Sage.
- Silverman, D. (2000). *Doing qualitative research: A practical handbook*. Thousand Oaks, CA: Sage.
- Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26(2), 114–145.
- Slavin, R. E. (2007). *Educational research in an age of accountability*. Boston, MA: Pearson.
- Smith, K., & Davies, J. (2010). Qualitative data analysis. In L. Dahlberg & C. McCaig (Eds.), *Practical research and evaluation: A start-to-finish guide for practitioners* (pp. 145–158). London, England: Sage.
- Smith, M., & Bowers-Brown, T. (2010). Different kinds of qualitative data collection methods. In L. Dahlberg & C. McCaig (Eds.), *Practical research and evaluation: A start-to-finish guide for practitioners* (pp. 111–125). London, England: Sage.
- Sotos, A., Vanhoof, S., Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2, 98–113. doi: 10.1016/j.edurev.2007.04.001
- Steffe, L., & Thompson, P. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267–306). Mahwah, NJ: Lawrence Erlbaum Associates.
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Thousand Oaks, CA: Sage.
- Stuart, M. (1995). Changing the teaching of statistics. *The Statistician*, 44(1), 45–54.
- Turner, M., & Nightingale, D. (1997). *Links: A course in mathematics at curriculum level 6*. Auckland, New Zealand: New House.
- Uniservices asTTle team. (2008). *Quality questioning using the SOLO taxonomy*. Retrieved from <http://www.slideshare.net/jocelynam/solo-taxonomy>
- University of Bristol, & Creative Dimension Software Ltd. (2011). Canine Skeleton. (Photo.) Retrieved from <http://www.real3danatomy.com/bones/dog-skeleton-3d.html>
- van Nes, F., & Doorman, M. (2010). The interaction between multimedia data analysis and theory development in design research. *Mathematics Education Research Journal*, 22(1), 6–30. doi: 10.1007/BF03217557
- Walford, G. (2001). *Doing qualitative educational research: A personal guide to the research process*. London, England: Continuum.

## References

- Watson, J. (2004). Developing reasoning about samples. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277–294). Dordrecht, The Netherlands: Kluwer Academic.
- Watson, J. (2005). Developing an awareness of distribution. In K. Makar (Ed.), *Reasoning about distribution: A collection of research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-4, July, 2005), Auckland, New Zealand* (pp. 1–33). Brisbane, Australia: University of Queensland.
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Watson, J. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal*, 7(2), 59–82.
- Watson, J., & Moritz, J. (1999). The beginning of statistical inference: comparing two data sets. *Educational Studies in Mathematics*, 37, 145–168. doi: 10.1023/A:1003594832397
- Wellington, J. J. (2000). *Educational research: Contemporary issues and practical approaches*. London, England: Continuum.
- Whittin, D. (2006). Learning to talk back to a statistic. In G. Burrill & P. Elliot (Eds.), *Thinking and reasoning with data and chance: Sixty-eighth yearbook* (pp. 31–39). Reston, VA: National Council of Teachers of Mathematics.
- Wikipedia. (2012). Structure of observed learning outcome. Retrieved from [http://en.wikipedia.org/wiki/SOLO\\_Taxonomy](http://en.wikipedia.org/wiki/SOLO_Taxonomy)
- Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal*, 5(2), 10–26.
- Wild, C. J. (2007, September). *CensusAtSchool and changes in the statistics curriculum*. Paper presented at the New Zealand Association of Mathematics Teachers Conference (NZAMT-10), Auckland, New Zealand.
- Wild, C. J., Horton, N., Pfannkuch, M., & Regan, M. (2008). *Guidelines for “how to make the call”*. Retrieved from CensusAtSchool website: <http://www.censusatschool.org.nz/2010/documents/Shifts%20diagram.11.02.04.pdf>
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–265. doi: 10.1111/j.1751-5823.1999.tb00442.x
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2010, July). Inferential reasoning: learning to “make a call” in theory. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS-8), Ljubljana, Slovenia* (pp. 1–6). Voorburg, The Netherlands: International Statistical Institute.
- Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. (2011). Towards more accessible conceptions of statistical inference (with discussion). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(2), 247–295. doi: 10.1111/j.1467-985X.2010.00678.x
- Wild, C. J., & Seber, G. (2000). *Chance encounters: A first course in data analysis and inference*. New York, NY: John Wiley & Sons.

## References

- Wittman, E. (1998). Mathematics education as a “design science”. In A. Sierpinska & J. Kilpatrick (Eds.), *Mathematics education as a research domain: A search for identity: An ICMI study* (pp. 87–103). Dordrecht, The Netherlands: Kluwer Academic.
- Zayac, S. (1991). Improving open-ended problem solving: lessons learned in industry. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics (ICOTS-3, July, 1991), Dunedin, New Zealand* (Vol. 1, pp. 79–83). Voorburg, The Netherlands: International Statistical Institute.
- Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40–58.