ASSESSMENT OF STATISTICAL UNDERSTANDING IN A MEDIA CONTEXT

Jane M. Watson, University of Tasmania, Australia

This paper will report on the use of a three-tiered model to assess the statistical understanding of school students in the context of a media report. Data from 229 Australian students will be presented to show levels of performance in relation to the model reflecting increasing complexity of response. Four groups of students were tested at two-year intervals with one of these groups also tested after four years. The media item used in this report contained a misleading claim about a population based on an unrepresentative sample. The assessment was designed to determine if students could question the claim. Results indicate that educators have a long way to go in creating a statistically literate citizenry.

INTRODUCTION

The assessment of statistical understanding has received increased attention (Gal and Garfield, 1997) as statistics has attained a higher profile in the school mathematics curriculum and as methods of assessment have become more varied to meet the goals of statistics education. The importance of statistical literacy as students leave school to take an active part in society is acknowledged in many countries. In Australia, for example, an Australian Bureau of Statistics survey of adult literacy (McLennan, 1997), contained three subscales: prose literacy, document literacy and quantitative literacy. The last two parts were heavily dependent on basic data interpretation from graphs and tables, adapted from newspaper formats. The importance of applying statistical understanding in social contexts such as found in the media must be acknowledged and assessed in school where preparation for participation in society occurs.

This report will focus on the assessment of one aspect of statistical literacy - the relationship between sample and population. The development of school students' understanding over time will be assessed using a hierarchical model based on written responses to two survey items, one of which was based on a newspaper article.

The hierarchical model to be used in this report (Watson, 1997) has three tiers: (1) a basic understanding of statistical terminology; (2) an understanding of statistical language and concepts when they are embedded in the context of wider social discussion; and (3) a questioning attitude which can apply more sophisticated concepts to contradict claims made without proper statistical foundation. These skills represent increasingly sophisticated thinking and reflect concerns of other research documenting students' cognitive development more generally (e.g., Biggs and Collis, 1982). One aspect of the work of Biggs and Collis (1982) is particularly relevant here. Within the more general stages of development or modes of functioning they suggest three increasing levels of

thinking: (i) unistructural (U), (ii) multistructural (M), and (iii) relational (R). At the U level students focus on single elements required to solve a task presented and do not recognise contradictions should they arise. At the M level, students process several aspects, usually in sequence and although they may recognise conflict if it arises in solving a task, they are unable to resolve it completely in the context set. At the R level, students process all elements required to solve a task in a coordinated fashion and are able to resolve conflict in the context set. These levels of response are observed in association with the tiers of the statistical hierarchy.

ITEMS

For this report responses from two survey items presented in Figure 1 will be used to document tiers of the hierarchy for statistical literacy presented above. Students answered Item 1 as part of a 20-item Chance and Data Survey and Item 2 as part of a 6- or 11-item Media Survey. Many other questions on other aspects of probability and statistics were answered between these two questions and there is no evidence that responses to Item 1 in any way influenced responses to Item 2. In some instances the two surveys were not completed in the same session. Item 1 assesses the first tier of the hierarchy for the term "sample." This sets a base line for the second two tiers as it would not be expected that students could give reasonable responses to Item 2 without understanding what a sample is. Item 2 sets the context to assess the second and third tiers of the hierarchy depending on the response given. The two parts of the question for Item 2 are intended to assess the highest level of understanding without a hint, followed by a geographical clue to assist if required to resolve conflict.

Item 1   If you were given a "sample", what would you have?

Item 2

> ABOUT six in 10 United States high school students say they could get a handgun if they wanted one, a third of them within an hour, a survey shows. The poll of 2508 junior and senior high school students in Chicago also found 15 per cent had actually carried a handgun within the past 30 days, with 4 per cent taking one to school.

Q1 Would you make any criticisms of the claims in this article?

Q2 If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona? Why or why not?

Figure 1. Items used to assess understanding of sampling concept and representativeness.

SAMPLE

Grade 6 and grade 9 students, selected to represent a cross section of school types across the Australian state of Tasmania, first answered the questions in 1993 ($n = 54$ for grade 6 and $n = 81$ for grade 9) or in 1995 ($n = 61$ for grade 6 and $n = 33$ for grade 9). They again answered the same questions two years later, as did the 1993 grade 6 students in 1997. There was no indication from informal discussion with students that they recalled the specific questions or their responses to them from previous surveys.

METHOD

Responses to items were categorised using a language analysis software package, with the categories assigned levels according to the U-M-R structure described earlier. Exemplary responses will be given for each tier of the hierarchy for statistical literacy along with the percentages in each group who responded at each level. For Tier 1 the building of understanding of representative sample will be seen in terms of a U-M-R sequence. For Tiers 2 and 3 another similar sequence within the context provided by the newspaper article will be seen to operate. By assigning numerical values from 0 to 3, or 0 to 4, to observed levels of performance, longitudinal changes will be documented with paired t-tests for the two items. The maximum level achieved by each student for the two items will then be summarised and again longitudinal changes reported.

RESULTS

First tier. Assessment of achievement in the first tier of the hierarchy is determined from responses to Item 1 in Figure 1. As this is an open-ended question, students may not give the highest level response of which they are capable. However, while some students, 1% of the entire sample, do not respond with any idea related to the term, most students can at least give an example of a sample. The following are U responses which give an example or a single idea associated with the term sample.

U: A sample of blood.   U: A little bit.   U: Something to test.

More complex Tier 1 responses combine more than one aspect of being a sample without a complete description including the aspect of representation. They usually make reference to the part-whole relationship or the testing aspect, and are classified as M.

M: A little of something, not the whole thing but a little piece of it.

M: I would have a sample of dirt, something they have done tests on.

The most sophisticated responses in Tier 1 put together the part-whole nature of sample-population with the idea of representation, perhaps in the form of testing, and are classified as R.

R: You would have a small amount of something, like a new washing up liquid, to test in your home.

R: A thing taken out of a larger group ... to view as a representative of that group.

These responses demonstrate the concept of sample as "a small part representing a whole" which appears necessary for interpreting the term in other contexts.

The top part of Table 1 shows the levels of Tier 1 understanding of the students in this study. For all groups in all years the modal level of response is M. While there is an improvement for the grade 6 students over time (p < .02 for three comparisons), the grade 8 and 9 students show no significant change over two years.

Second tier. Assessment of achievement in Tier 2 of the hierarchy is determined from responses to Item 2 in Figure 1. Some responses to Item 2 do not at any point recognise that sampling is an issue in the article. Examples include the following.

Students shouldn't be carrying guns. (Q1)

Yes because one or more of them could get mad and just shoot anyone. (Q2)

No because I would like to try and help these students, not neglect them. (Q2)
Responses of this type were judged not to be in Tier 2 of the hierarchy.

Some answers to both questions in Item 2 illustrate an appreciation of some aspect of the term sample without realising the non representative nature of the Chicago sample. Those who, even with a hint, do not recognise the misleading claim but who make a comment with some statistical relevance are deemed to be in Tier 2; that is, they recognise the context of the query. At the U level are responses which address single issues associated with sampling but which are not the focus of the questioning.

U: How do you know they are not lying?

U: The students might be just saying that because they think it's tough.

More complex Tier 2 responses to the first question recognise the part-whole relationship of Chicago and the United States but then go on to demand that the population be used rather than suggesting a more representative sampling procedure. These responses are classified as M.

M: If they wanted to get the facts right they should survey every school in America...
For the second question, M responses tend to recognise the part-whole nature, that all schools are not the same, without seeing Chicago specifically as non representative.

M: No because every place is different.

M: No because the whole of the US wasn't surveyed, so we don't know that the hand gun situation is the same throughout the US.

Third tier. Responses which do not recognise the misleading claim in the first question of Item 2, but comment sensibly when given the hint in the second question, are

deemed able to relate together the information to illustrate an understanding which satisfies Tier 3 of the hierarchy of statistical literacy. These responses are R in recognising with assistance that "the part representing the whole" is not satisfied.

R: No because this poll is in Chicago. Results may be different in Colorado (Q2)

Responses to the first question which require no help to recognise immediately the difficulty are judged to be at a higher level again, possibly unistructural in another cycle or mode of functioning where understanding of the concept of sample is fully internalised and used as a single representative construct.

U*: The poll was only Chicago not substantiating the words 6 in 10 US high school students. (Q1)

U*: Which school in Chicago? In what area? Were the people asked representative of the entire high school population in USA? (Q1)

The bottom part of Table 1 shows the levels of response to Item 2 in Tiers 2 and 3 of the hierarchy. The increasing structural complexity of responses in applying the concept of sample in context is shown in the U and M responses in Tier 2 and R and U* responses in Tier 3. In every group, however, the modal category is for responses which do not reach Tier 2. Very small percentages of students reach the U* level before grade 10. The grade 6 and 8 groups show improvement *by* grade 10 (p < .0001 for both comparisons), as does the 1993 grade 9 group by grade 11 (p = .001), but other group comparisons, including for the two grade 6 to 8 pairs, are not significant at the .05-level.

Table 1: Percentage of Responses at Levels of Tiers for Understanding of "Sample"

	Grade								
	6	8	10	6	8	9	11	9	11
	1993	1995	1997	1995	1997	1993	1995	1995	1997
	|-------- (n = 54) --------|			|-- (n = 61) --|		|-- (n = 81) --|		|-- (n = 33) --|	
Out of context	2	2	2	3	-	1	2	3	-
Item U	28	9	6	31	16	12	10	6	15
One M	59	69	63	54	67	53	57	48	45
R	11	20	30	11	16	33	31	42	39
Out of context	80	78	48	89	72	48	32	52	48
Item U	7	7	13	-	18	20	16	6	3
Two M	9	7	19	7	5	14	16	24	9
R	4	4	7	2	3	12	26	12	15
U*	-	4	13	3	2	6	10	6	24

It would be expected that those with higher level Tier 1 understanding of the term sample would also perform at higher levels for Tier 2 and 3 questions. This is not borne out in the data. The high percentage of responses failing to reach Tier 2 in Item 2 is reflected uniformly across the U-M-R levels for Item 1. Of the 86 responses at the R or U* levels in Item 2, however, 76 of these achieved an M or R level on Item 1. Hence it appears that better achievement on Item 1 may be necessary but not sufficient for high achievement on Item 2.

In order to compare performance over time for the complete hierarchy, students were allocated the maximum U-M-R level achieved on the two items. These data are presented in Table 2. A comparison with Table 1 shows that while large percentages of most groups are not operating in Tier 2 or 3 for the context of Item 2, almost all have at least a basic understanding of "sample" with some able to challenge successfully an erroneous claim. For all longitudinal comparisons, except grade 9 (1995) to grade 11 (1997), differences are significant at the .05-level, with the difference for grade 6 (1993) to grade 10 (1997) more striking ($t = 5.21$, $p < .0001$).

Table 2: Percentage of Maximum Response Levels for Understanding of "Sample"

Grade	6	8	10	6	8	9	11	9	11
	1993	1995	1997	1995	1997	1993	1995	1995	1997
	|-------- (n = 54) --------|			|-- (n = 61) --|		|-- (n = 81) --|		|-- (n = 33) --|	
Out of context	2	2	2	3	-	-	-	3	-
U	28	9	6	28	16	10	6	3	9
M	56	63	46	56	64	44	36	36	33
R	15	22	33	10	18	40	48	52	33
U*	-	4	13	3	2	6	10	6	24

DISCUSSION

The assessment of statistical literacy using the model suggested here offers promise in achieving the detailed documentation of students' understanding which is necessary to help them achieve the objectives relating to statistically literate decision-making in society. The three-tiered hierarchy acknowledges the ultimate aim of being able to question unwarranted claims made in society, in particular in the media, while at the same time starting with basic understanding of terminology. The embedding of the U, M and R levels in association with observed outcomes for basic understanding (Tier 1) and for application in context (Tiers 2 and 3) adds detail which can assist in structuring activities to raise students' levels of response.

The results indicate that while most students have a Tier 1 understanding of sample, many are unable to apply it in a social context. This may reflect a curriculum which does not include application or a difficulty of recognising significant statistical aspects in a potentially distracting context. It may also reflect the developmental levels of students, as improvement in performance over time occurs for grade 6 on Item 1 and for two of the three older grades for Item 2. While change as observed here may also be related to out-of-school experiences, curriculum planners may need to consider extra measures, including a more purposeful linking of the three tiers of the hierarchy, if there is a desire to achieve higher levels of statistical literacy before students leave school.

REFERENCES

Biggs, J. B., and Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Gal, I, and Garfield, J. B. (Eds.). (1997). *The assessment challenge in statistics education.* Amsterdam: IOS Press and The International Statistical Institute.

McLennan, W. (1997). *Aspects of literacy: Assessed skill levels Australia 1996.* Canberra: Australian Bureau of Statistics.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal and J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.