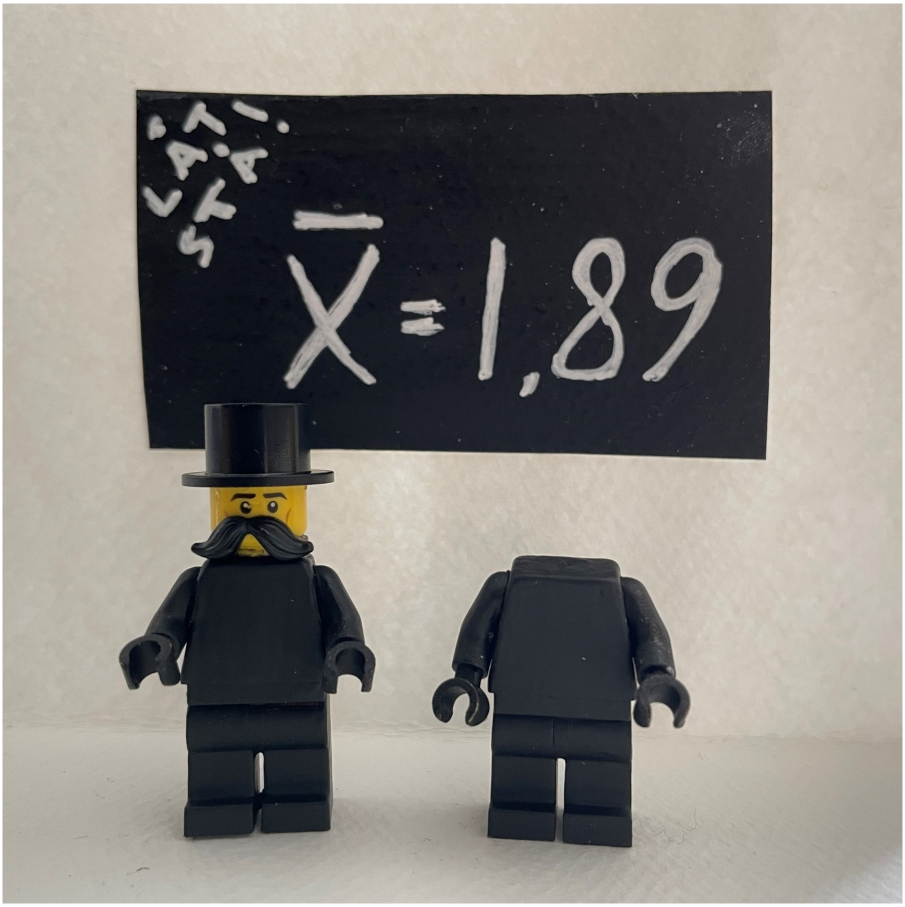
Stockholm University

Mean, median, and mode in school years 4–6

A study about aspects of statistical literacy

Karin Landtblom



Doctoral Thesis in Mathematics Education at Stockholm University, Sweden 2023

Mean, median, and mode in school years 4–6

A study about aspects of statistical literacy

Karin Landtblom

Academic dissertation for the Degree of Doctor of Philosophy in Mathematics Education at Stockholm University to be publicly defended on Friday 5 May 2023 at 10.00 in hörsal 9, hus D, Universitetsvägen 10 D and online via Zoom, public link is available at the department website.

Abstract

This thesis explores different aspects of statistical literacy such as mathematical knowledge, context knowledge, use of words, and conceptions. The focus is on measures of central tendency: mean, median, and mode, and school years 4–6 (ages 10–13). The thesis contains of five papers and the phenomenon is studied in different contexts. In the first three papers, data is generated through a questionnaire answered by prospective teachers, teachers, and students in grade 6. In papers 4–5, second hand data is generated through textbook analysis where all tasks about the measures in seven different textbook series were analysed.

Papers 1–3 showed that all respondent groups, primarily express procedural knowledge. They sometimes mix up the definitions of mean and median. Mean appears to be the most familiar measure and different contexts appropriate for mean are suggested. Median and mode appear to be less familiar, especially median which according to the students only exists in a school context. All respondent groups show several ways to express the mean using different colloquial connotations. Median and mode on the other hand do not bring any connotations, leading to difficulties to express explanations. For mode, some students used a homonym that gives a wrong meaning to the concept. This implies that the support for understanding mean, based on colloquial words, is not available for median or mode.

The results from paper 4 show a high proportion of procedural tasks dealing with above all quantitative values for mode. Only one textbook definition of mode exemplified with qualitative values. In paper 5, tasks were examined out of mathematical properties related to input object, transformation, and output object. Here, tasks about both mean, median, and mode were examined. The results show that the distribution of the tasks was skew, meaning that students have different opportunities to learn various mathematical properties of the three concepts, something that was even more complex given that in many tasks, the mathematical properties in focus were implicit.

Overall, statistical literacy according to the results generated in the five different studies, appear to be pre-dominantly about numbers and procedures. Very little attention seems to be on contextual knowledge, something that is crucial in statistical literacy.

Keywords: *Mathematics teaching, statistical literacy, mathematics teachers, prospective teachers, students, text books, school years 4–6.*

Stockholm 2023
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-215493

ISBN 978-91-8014-238-0
ISBN 978-91-8014-239-7

Stockholm University

Department of Teaching and Learning

Stockholm University, 106 91 Stockholm

MEAN, MEDIAN, AND MODE IN SCHOOL YEARS 4–6

Karin Landtblom

Stockholm University

Mean, median, and mode in school years 4–6

A study about aspects of statistical literacy

Karin Landtblom

ISBN print 978-91-8014-238-0
ISBN PDF 978-91-8014-239-7

Printed in Sweden by Universitetsservice US-AB, Stockholm 2023

Till mamma

Acknowledgements

Thanks to

My mother: for lifelong support and encouragement
Lasse: for love, support, and good food
Sören Karlsson and Gunilla Nilsson: generous and talented colleagues who have been important to me in my development as a teacher
Tor Englund and Lilian Ahlm: teaching statistics came into my life thanks to you
Lovisa Sumpter: a supervisor beyond the ordinary - respect!
Kerstin Larsson and Judy Sayers: for support and joy
Paul Andrews: for support when needed
Tim Houghton: for the best language review imaginable
Inger Ridderlind and Anna Almqvist: for being my gun bearers through many years and for relieving me when there has been a lot to do
Kattis, Matilda and John: because you are my cute children and make sure I do something other than work
Wilma, Birk, Freja and Alva: simply because you are my sweet grandchildren
And of course – everyone else who has supported me in various ways – no one mentioned and no one forgotten

Thank you!

Karin Landtblom, Gustavsberg, March 2023

Pre-introduction
Illustrated on the front page

> Dr Frankenstein var en man med övernaturliga egenskaper. Han var statistiker på Statistiska Centralbyrån och trollade utan större ansträngning fram små svarta gubbar med mystisk innebörd: ibland innehöll en sån där liten gubbe alla nykterister i Västerbottens län, ibland kunde den lilla svarta gubben föreställa alla svenskar som röstade på högerpartiet för två år sen, och bredvid stod då en annan liten gubbe och föreställde alla svenskar som röstade på högerpartiet nu senast, fast den sista lilla gubben hade inget huve. Så ni förstår att Dr Frankenstein han var allt en riktig trollkarl!

Inledningen av Berättelsen om Sven-Erik Medeltal, statistikerns idealmänniska. Hämtat ur Sagor för barn över 18 år (Danielsson, 1964).

> Dr. Frankenstein was a man with supernatural qualities. He was a statistician at Statistiska Centralbyrån and without much effort conjured up little black men with mysterious meaning: sometimes such a little man contained all the sober people in Västerbotten County, sometimes the little black man could represent all the Swedes who voted for the right-wing party two years ago, and next to him stood another little man and represented all the Swedes who voted for the right-wing party recently, although the last little man had no head. So, you see that Dr. Frankenstein he was a real wizard!

Translation of the introduction to the Story of Sven-Erik Medeltal, the statistician's ideal person. (Danielsson, 1964).

Included papers

I. Landtblom, K. K. (2018). Prospective teachers’ conceptions of the concepts mean, median and mode. In H. Palmér, & J. Skott (Eds.), *Students’ and teachers’ values, attitudes, feelings and beliefs in mathematics classrooms* (pp. 43–52). Springer.

II. Landtblom, K., & Sumpter, L. (2021). Teachers and prospective teachers’ conceptions about averages. *Journal of Adult Learning*, *Knowledge and Innovation*, *4*(1), 1–8.

III. Landtblom, K., & Sumpter, L. (submitted). Which measure is most useful? Grade 6 students’ expressed statistical literacy.

IV. Landtblom, K. (2018). Is data a quantitative thing? An analysis of the concept of the mode in textbooks for grade 4–6. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10*). International Statistical Institute.

V. Landtblom, K. (submitted). Afforded opportunities to learn mean, median, and mode in textbook tasks.

Contents

1. Introduction 1
1.1 Aim and research questions 5
1.2 The outline of this thesis 6

2. Literature review—statistical literacy 8
2.1 Historical development of statistics 8
2.2 Statistics as a school subject 9
2.3 Definitions of statistical literacy 10
2.4 Models of statistical literacy 12
2.5 Previous studies of textbook tasks on mean, median, and mode 14
2.6 Research on mean, median, and mode 15
2.7 Students' understanding and conceptions 17
2.8 Teachers and prospective teachers' understanding and conceptions 18

3 Theoretical considerations 21
3.1 Aspects of statistical literacy 21
3.1.1 Mathematical knowledge 22
3.1.2 Use of Words 26
3.1.3 Context knowledge 28
3.1.4 Conceptions 30
3.2 Affordance and opportunities to learn 31

4 Methodology and methods 33
4.1 Methodological considerations 33
4.1.1 Questionnaires 35
4.1.2 Textbook analysis 36
4.2 Sampling and methods 37
4.2.1 Questionnaires 38
4.2.2 Textbook analysis 40
4.3 Data analyses 41
4.3.1 Questionnaires 42
4.3.2 Textbook analysis 44
4.4 Methodological reflections 45
4.4.1 Validity of data-generation methods 45
4.4.2 Validity of interpretation 47
4.4.3 Reliability 50
4.4.4 Ethical considerations 51

5 Results 53
5.1 Paper 1—prospective teachers' conceptions of the concepts mean, median, and mode 53
5.2 Paper 2—teachers and prospective teachers' conceptions about averages 54
5.3 Paper 3—which measure is most useful? Grade 6 students' expressed statistical literacy. 55
5.4 Paper 4—is data a quantitative thing? An analysis of the concept of the mode in textbook tasks for grades 4–6 56

5.5 Paper 5—afforded opportunities to learn mean, median, and mode in textbook tasks 57

6. Discussion 59
6.1 Mathematical knowledge 59
6.1.1 Procedural knowledge 59
6.1.2 Mathematical properties 61
6.2 Use of words 61
6.3 Context knowledge 63
6.4 Conceptions 63
6.5 Methodology 64
6.6 Statistical literacy 67

7. Svensk sammanfattning 70
8.1 Artikel 1–3 71
8.2 Artikel 4–5 74
8.3 Sammanfattning 76

References 79

Appendix 100
Appendix I: Information to prospective teachers 100
Appendix II: Information to principals for approval of teachers' participation 101
Appendix III: Information to principals for approval of students' participation 102
Appendix IV: Consent, parents and caretakers 103

1. Introduction

> In the early days when statistical literacy was low, those who could read and write this strange new language were set off and apart from the others (Ogburn, 1940, p. 260).

Developing statistical literacy, developing reasoning about statistics in realistic situations, is an overarching goal for the teaching and learning of statistics (Ben-Zvi & Garfield, 2004; Franklin et al., 2007; Franklin & Kader, 2010; Hannigan et al., 2013; Shaughnessy, 2007). For this reason, knowledge about how to teach for statistical literacy is an important issue (Ben-Zvi & Garfield, 2004). For teachers and prospective teachers, this includes both developing conceptions of statistics, and becoming aware of students' conceptual understanding and their levels of statistical literacy (Pfannkuch & Ben-Zvi, 2011).

To start with, statistics and mathematics can be viewed in several ways. Two major differences between statistics and mathematics are context and variability (Cobb & Moore, 1997; Groth, 2007, 2013). Compared to mathematics, statistics deals with data in a specific way where data are numbers with a context (Cobb & Moore, 1997). The difference is that statistic content arises from the omnipresence of variability (Cobb & Moore, 1997), which is indicated by patterns and relationships between variables (Reading & Shaughnessy, 2004). These patterns help with making predictions based on variability and are a foundation for reasoning about statistics (Ben-Zvi, 2004). The context of data is also valuable for learners as it helps them to find meaning in the observed patterns (Pfannkuch, 2011). Statistics exists to support other fields with ideas about how to deal with data where mathematics is one tool (Cobb & Moore, 1997; Groth 2007, 2013; Wild et al., 2018). As an example, for the arithmetic mean, the mathematics is about performing the calculation, whereas the statistics is about understanding what the mean represents in terms of the dataset, starting from context and variability (Mokros & Russell, 1995). Being statistically literate then means being both able to do the mathematical transformations (e.g., Lithner, 2008) and being able to evaluate the answer (Callingham

& Watson, 2017). Variability and measures of central tendency (mean, median, and mode) are inseparable (Konold & Pollastek, 2004). The measures are concepts that tell us how values in a dataset are distributed in relation to other values (Ben-Zvi, 2004; Konold & Pollastek, 2004). The measures tell us how the values in the dataset are centred; for instance, the mean can be seen as a balance point (where the deviation of surrounding values equals zero), the median as a value having an equal number of values above and below (e.g., Konold & Pollastek, 2004). Properties like these are consequences of the mathematical procedure for calculating a measure (Konold & Pollastek, 2004; Strauss & Bichler, 1988). Another aspect of variability and measures is consideration of the variables in the context—what we measure. Sometimes there are repeated measurements of one object's characteristic and sometimes there is one measurement of several objects' characteristics (Konold & Pollastek, 2004). The nature of the measurement affects the interpretation of the dataset. A body of research has stressed the importance of understanding the mathematical principles of the different averages (e.g., Mason, 2011; Strauss & Bichler, 1988). Here, there is a difference between mathematics and statistics:

> In mathematics education, a distinction is made between computation and the understanding of the mathematical concepts underlying the computation. The arithmetic average is computed by adding the values to be averaged and dividing this sum by the number of values that were summed. Perhaps the simplicity of the technical side of its computation makes the average appear to be so straightforward and simple. (Strauss & Bichler, 1988, p.65).

The simplicity of the procedure, as described in the quote above, does not capture the complexity between the input values, the different procedures, and the output values. Strauss and Bichler (1988) identify seven mathematical properties related to averages, all important to the understanding of these measures. However, it is not clear what is related to the three different concepts, and mode is not mentioned at all. Also, these seven different mathematical properties are presented without distinguishing them as transformations, objects, or concepts (e.g., Lithner, 2008). In Strauss and Bichler (1988), the different properties fall into the categories statistical, abstract, and representative, but there is no in-depth description of how to separate between different mathematical properties. In my review of research, it appears that since Strauss and Bichler's (1988) study, little has been done to repeat a similar mathematical analysis and description of a method.

Given that the mathematical properties of the different averages are so central to the understanding, each task is different depending on which input value you are working with. For instance, one can talk about the average of 2.1 children, but no-one can have 2.1 children. Since mathematics education is most often closely connected with the textbook, both in terms of what students should learn and how the teaching should take place (Glasnovic Gracin, 2018; Jones & Pepin, 2016; Pepin & Haggarty, 2001; Remillard, 2005), it is of interest to investigate what is afforded in different textbooks. In addition, the different averages are often presented one at a time in different curricula, so there is a risk of students missing a holistic approach (Nilsson et al., 2018). Hence, it is not only about the different mathematical properties of each task, but also the specific context of each task and the design of each task: which properties are explicit and which are implicit.

Insight into the development of statistics is also valuable. What is taught and how it is taught are matters of historical, social, and political interest (Engel, 2019; Weiland, 2019a; Wild, 2017; Zapata-Cardona & Marrugo Escobar, 2019; Zapata-Cardona & Martínez-Castro, 2021). Historical development includes, among other things, technology. Technology reduces the need for basic skills, such as the ability to calculate a mean but increases the need for conceptual understanding (Wild, 2017). Other historical aspects include the development of statistics over time and implications of this for education (Bakker, 2003; Bakker & Gravemeijer, 2006). Social aspects include how statistical literacy is, and has been, essential for citizens to be able to engage and participate in democratic processes (Engel, 2019; Ridgway et al., 2018; UNESCO, 2013, 2017; Wells, 1903). This need, to understand statistics both as consumers and producers, was predicted more than 100 years ago (Wells, 1903). One example of how to develop students' critical abilities is the inclusion of current contexts in teaching, for example, global warming (e.g., Büscher, 2019; McCright, 2012; UNESCO, 2013; Zapata-Cardona & Martínez-Castro, 2021). As an example, McCright (2012) discusses the importance of developing both scientific literacy and statistical principles and processes when teaching about climate change. Büscher (2019) uses similar reasoning concerning the development of students' understanding of averages with respect to the growth and melting of Arctic sea ice. He captures this connection between statistical literacy and the context by using the theoretical concept 'situativity of knowledge'. The concept fits well with how Nilsson and colleagues (2018) describe the core of theories in statistics education: in statistics education, the context is part of—as McCright (2012) calls it—the statistical principles and processes. It is therefore crucial, when designing teaching in statistics that aims to develop students' statistical literacy, to consider the fact that the context is as important as the mathematical procedures and concepts.

Focusing on contexts provides the learner with the possibility of enhancing their understanding of the world (e.g., Stillman et al., 2013). However, critical education, where mathematical and statistical content interact, places great demands on the teacher and is often neglected in school (Engel, 2019; Weiland, 2019a). Civic statistics, a relatively new sub-discipline of statistical literacy addresses the kind of statistical information about society that involves statistics and political, social, and educational science (Engel, 2019). The summary of my personal journey, below, provides a brief multifaceted view of teaching statistics, and suggests some future directions for statistics education.

Here, there is an even stronger emphasis on the importance of society and its needs, not the individual's. While some studies show it is possible to change people's attitudes to climate change with scientific information that includes statistical literacy (e.g., Ranney & Clark, 2016), other studies show that affective factors might prevent any change in perception of climate change risk (Kahan et al., 2012). It is therefore relevant to understand how affective factors, such as beliefs, attitudes, conceptions, etc., are part of an individual's understanding of a mathematical or statistical concept (e.g., Juter, 2005; Leavy, 2010; Leavy & O'Loughlin, 2006). Researchers have also concluded that there is a lack of research focusing on affect and statistics (Estrada et al., 2011), so there is clearly a gap in the research area.

I started my teaching career in 1987 with school years 7–9 (ages 13–16). I remember, when the 1994 curriculum was implemented, how students entering year 7 only knew about the mean, and maybe a little about median. Another memory is that statistics was often the last chapter in the textbook, which meant that the topic was not always handled in depth. My personal interpretation is that statistics had lower status than other areas in the mathematics teaching. When a colleague and I wrote mathematics textbooks, we upgraded statistics to the first chapter. I remember students being positive about starting the semester with statistics. It even changed some students' attitude to mathematics. In 2000, I started to lecture in teacher education, and have since had the privilege of teaching statistics with a focus on school years 4–6 (ages 10–13). I noticed how prospective teachers from the outset showed above all procedural knowledge about measures of central tendency. Depending on when the prospective teachers themselves had been students in elementary school, they had been taught in different ways, following the development of curricula. Over the years, my interest in developing teaching about statistics in teacher education was aroused.

In this thesis, mean, median, and mode are either referred to by their names, as measures of central tendency, as measures, or as averages. The measures are applicable depending on data level. Sometimes, I only distinguish between qualitative and quantitative levels, sometimes there are reasons for explicitly talking about nominal and ordinal level. Data at a nominal level can only be grouped, therefore only mode is applicable. Data on an ordinal level can be both grouped and ranked, one example being shoe size. If data are rankable then the median can be calculated. Since ordinal data are not equidistant, the mean cannot be calculated (e.g., Kitto et al., 2019; Mayén & Diaz, 2010; Zawojewski & Shaughnessy, 2000). In this thesis, I do not distinguish between interval level and ratio level for quantitative values. The difference between these, that data at a ratio level have an absolute zero, is not important knowledge for students in grades 4–6.

1.1 Aim and research questions

This thesis consists of problem-driven studies (e.g., Schoenfeld, 1992). Based on the different gaps identified and presented in the text above concerning the need for a mathematical understanding and analysis of mathematical properties of the different measures (e.g., Strauss & Bichler, 1988), the need to understand the role of the context with respect to statistical literacy, the so called 'situativity of knowledge' (Büscher, 2019), and the role of affect when teaching and learning using statistics (e.g., Ranney & Clark, 2016), the aim is to explore and discuss different aspects of statistical literacy. To address this aim, the following two overarching research questions are posed: a) What aspects of statistical literacy do students, teachers, and prospective teachers express about mean, median, and mode? b) What do textbook tasks afford with respect to different aspects of statistical literacy regarding mean, median, and mode?

These two overarching research questions are general, and therefore in each of the studies, more specific questions have been asked. They are presented in the next section, the outline of the thesis.

1.2 The outline of this thesis

This is a compilation thesis consisting of a kappa and five papers—two published conference papers, one published article, and two submitted articles.

The first research question, about how students, teachers, and prospective teachers talk about mean, median, and mode, will be answered in papers 1–3 that deal with prospective teachers', students', and teachers' conceptions about mean, median, and mode. These three papers aim to address the identified gap in research about affect and statistics (Estrada et al., 2011). The second overarching research question, about what is afforded regarding mean, median, and mode in textbook tasks, is addressed in papers 4 and 5. These are both textbook analyses, investigating what statistical knowledge various textbook tasks afford, and what words are used in definitions of the mode. Their relevance is based on the argument that textbooks are an important feature of mathematics and statistics education (e.g., Glasnovic Gracin, 2018; Jones & Pepin, 2016; Pepin & Haggarty, 2001; Remillard, 2005).

The kappa consists of seven chapters that describe how this problem-driven study was performed. Chapter 1 is an introduction and Chapter 2 a literature review presenting relevant research about statistical literacy. Statistical literacy is a comprehensive concept that encompasses several various aspects required for developing knowledge about statistics—mathematical, statistical, and context knowledge, for instance. Chapter 3 presents the theoretical approaches taken in the five papers and Chapter 4 is about methodological considerations, sampling, methods of generating data, data analysis, and ethical issues. Chapter 5 presents a summary of the results of the papers. Chapter 6 discusses the results and presents conclusions and implications for further research. Chapter 7 is a summary of the thesis in Swedish.

The titles of the five papers and their specific research questions are presented in Figure 1:

Figure 1

The five papers included in the thesis with their research questions

Paper	Title	Research questions
1	Prospective teachers' conceptions of the concepts mean, median and mode (Landtblom, 2018a)	How do prospective teachers conceptualise the concepts of mean, median, and mode to a student in years 4–6?
2	Teachers and prospective teachers' conceptions about averages (Landtblom & Sumpter, 2021)	(1) What are the characteristics of the motivations given by prospective teachers and teachers to which of the averages is the easiest or hardest to explain? (2) What are their expressed conceptions about the usefulness of the averages? (3) How do prospective teachers and teachers differ in their responses?
3	Which measure is most useful? Grade 6 students' expressed statistical literacy (Landtblom & Sumpter, submitted)	(1) What different knowledge elements do the students use when explaining measures of central tendency, and how are these knowledge elements expressed? (2) Which measure is, according to grade 6 students, easiest or hardest to explain, and what characterises their motivations? (3) Which measure is considered, according to grade 6 students, most or least useful, and what characterises their motivations?
4	Is data a quantitative thing? An analysis of the concept of the mode in textbooks for grade 4–6 (Landtblom, 2018b)	What knowledge, procedural or conceptual, and quantitative or qualitative context, do textbooks in years 4–6 afford Swedish students on the concept of mode?
5	Afforded opportunities to learn mean, median, and mode in textbook tasks (Landtblom, submitted)	(1) What is the distribution among non-contextual and contextual tasks? (2) What opportunities to learn about a) input objects, b) transformations, and c) output objects do textbook tasks afford, and what does the distribution look like?

2. Literature review—statistical literacy

> On statistical literacy: "we acknowledge that the concept is dynamic, and likely to change in the face of major cultural upheavals associated with data science" (Ridgway et al., 2018, p.1).

The present thesis is situated within the theoretical framing known as statistical literacy. Although not theory-driven research (i.e., the results do not aim to develop this theory), the theory still must be scrutinised to identify appropriate theoretical units to be operationalised for each of the studies (Mason, 2018). In this section, I will first present some historical aspects of statistics as a subject and as a school subject. Thereafter, follows a review of statistical literacy, including models of statistical literacy. Finally, I present previous research relevant to the studies presented in Figure 1.

2.1 Historical development of statistics

The history of statistics allows us to follow the development of this field which reflects how different practical needs have presented over time in different contexts (Gal, 2019). In this way, statistics has evolved to enable us to understand what has been measured.

In a historical phenomenology of mean and median, Bakker and Gravemeijer (2006) give examples of how their findings relate to teaching activities and understanding of these concepts. The earliest reference to the mean is from approximately 400 BC where mean was used as an average value to estimate a total. Another historic example shows reasoning about fair share to be a precursor to the present use of mean. The last example is from a maritime context, where the excess between income and costs should be distributed fair among the interests – this was called the average. This was the first time average was mentioned with this meaning. In the sixteenth century, calculations for arithmetic mean were introduced with the intention of reducing errors, giving a truer value than the midrange does. One part of this conceptual change was the mathematical use of the decimal system in division. This view of the mean continued developing, and at the end of the nineteenth century, the mean could be understood as the most representative value. However, this development caused problems concerning the distinction between discrete and continuous values, something that remains problematic today.

Bakker and Gravemeijer (2006) found only a few historical studies on the median; the oldest from around 1600, where a construct that can be interpreted as median, was used to reduce errors. In the seventeenth century, median was also used in probability contexts based on half of the values on each side of the median. Examples of median in a statistical context appear in the late 1800's. Underlying reasons for the development of the median was that it was an easier calculation than mean, being insensitive to outliers, and further driven by the need for other measures within different research fields. One example of these research fields is intelligence and IQ. The IQ scale is on an ordinal level which implies that the median is a proper measure (ibid.).

There is much less written about the mode (Groth & Bergner, 2006). The word mode originates from Pearson (1895), who used it interchangeably with the term maximum ordinate (maximum value on the *y*-axis). Some statisticians question mode as a measure of central tendency since the mode may not be close to the centre of distribution (e.g., Utts, 2015). However, it has a unique role in being the only measure for data on a nominal level (Groth & Bergner, 2013).

2.2 Statistics as a school subject

Statistics became a school subject as a result of how statistical content developed over time (e.g., Wild et al., 2018). From a didactical perspective, the historical view of the mean as an estimation of a total, or the metaphor of fair share, are examples of how mean nowadays is conceptualised in school (Bakker & Gravemeijer, 2006). More recent changes in the environment, not least in the data universe, also affect how the content keeps changing (e.g., Wild et al., 2018). Over the last 15 to 20 years, statistics has evolved as a school subject in many countries (Arnold, 2008; Garfield & Ben-Zvi, 2004; Shaughnessy, 2007; Skolverket, 2022). The comprehensive changes to the curriculum are described as a focus shift from formulas and procedures to statistical literacy, reasoning, and thinking (SLRT) (e.g., Ben-Zvi & Makar, 2016; Callingham & Watson, 2017; Carmichael et al., 2010; Gal, 2019; Garfield &

Ben-Zvi, 2007; Gonzáles, 2016; Leavy & Hourigan, 2016; Petocz et al., 2018; Watson, 2013). This development is even described as a paradigm shift from descriptive statistics—a mathematical approach—to a statistical approach—becoming statistically literate (Pfannkuch, 2018). This paradigm shift also affects teacher education regarding professional development, research on teachers' knowledge, and development of models of the knowledge required to teach statistics (Arnold, 2008; Gonzáles, 2016; Pfannkuch & Ben-Zvi, 2011).

In Europe, statistics was introduced as a school subject in 1846 in Hungary and Belgium and 1868 in France (Bibby, 1986). In Sweden, statistics is taught in Mathematics and becomes a subject of its own at university level. The following summary of Swedish curricula through the years, below, is about the presence of statistics, especially measures of central tendency, in the subject Mathematics at school level. Content related to statistics appeared for the first time in 1919 as tables and graphs in Geography (UPL, 1920). It is not until the 1950s, however, that statistics becomes part of the mathematical content (the procedures), while tables are present in other subjects like History, Geography and Sports (Skolöverstyrelsen, 1951, 1955a, 1955b). Following the measures of central tendency through the curricula, we can see that the mean is present in the curricula of 1951 and 1962 (Skolöverstyrelsen, 1951, 1962). Mean, median, and mode are specified as concepts in 1969 and 1980 (Skolöverstyrelsen, 1969a, 1969b, 1980). In 1994, the measures are merged as elementary averages, with no mention of mean, median, and mode (Utbildningsdepartementet, 1994). Finally, from 2011, all three measures are mentioned explicitly (Skolverket, 2011). The present Swedish curriculum for school years 4–6 states that students shall learn about the measures mean, median, and mode, and how to use them in statistical investigations (Skolverket, 2022). This indicates both the mathematical aspect, the procedures, and the statistical part, managing variables in contexts.

2.3 Definitions of statistical literacy

One of the early definitions of the concept statistical literacy is the ability to understand quantitative language (Walker, 1951). Since then, the concept has proceeded to develop in step with the development of society, meaning that there are many angles of approach (e.g., Büscher, 2018, 2019; Callingham & Watson, 2017; Carmichael et al., 2010; Gal 1995, 2019; Leavy & Hourigan, 2016; Meletiou-Mavrotheris & Lee, 2003; Petocz et al., 2018; Watson, 2013; Watson & Moritz, 2000). Over the last twenty years, the use of the concept statistical literacy has increased (e.g., Ben-Zvi & Garfield, 2004; delMas, 2002; Rumsey, 2002; Schield, 2017; Watson & Callingham, 2003). The same is true for the umbrella concept statistical literacy, reasoning, and thinking (SLRT), a research area where slightly different definitions have emerged (Chance, 2002; Garfield, 2002; Rumsey, 2002). In light of what will be needed in the future, human understanding of data and knowledge about what data can or cannot do, are becoming the main focus of statistical literacy (Wild, 2017). A plausible conclusion, therefore, is that the research areas of statistical literacy and SLRT will continue to progress theoretically, given the development of society. Besides the various definitions of statistical literacy, there is often an overlap with the definitions of SLRT (delMas, 2002; Shaughnessy, 2007). There is, especially, much less agreement on the meaning of the concept statistical literacy (Ridgeway et al., 2011), which consists of two words, making it unclear which is dominant (Gal, 2004; Schield, 2017). Statistical skills is about calculations and content suitable for a certain school level (Watson, 2013). Literacy is about how reading and writing skills are associated with critical thinking and communication (Watson, 2013). Literacy is sometimes discussed as learning to read the word and the world, leading to the ability to write the word and the world (Gutstein, 2006). In the same way as general literacy is based on the symbols in a language (UNESCO, 2005), statistical literacy is based on the concepts and symbols used in statistics (Weiland, 2016). Though meaning is not present in the symbols and quantities (for mean, for instance), one must describe meaning related to a context.

The duality of the concept statistical literacy is probably the reason behind the many various definitions; for instance, activity-based, critical-thinking based, and content-based, to mention a few (Schield, 2017). Over time, new forms of communicating and displaying data have developed the meaning of being statistically literate (Gould, 2017; Ridgway et al., 2018; Wild, 2017). This makes statistical literacy a dynamic concept, which will probably continue to change in line with cultural changes linked to the development of data science (Ridgway et al., 2018). This is comparable to how literacy, in general, is dependent on the development of different infrastructures to support intellectual activities (diSessa, 2018). This means that representation is vital for what can be represented, and that literacy and statistical literacy are of both social and cultural importance (diSessa, 2018; Gal, 2004). In statistics, data are represented as a vital part of our lives (Gould, 2017). Gould's

(2017) conclusion is that becoming data literate enhances the possibility of becoming statistically literate.

2.4 Models of statistical literacy

The idea with models of statistical literacy is that they help to identify critical statistical skills like being well informed, making good decisions, being a critical thinker, having data awareness, and questioning and analysing data (Rumsey, 2002; Shaughnessy, 2007; Watson, 2013). Statistical literacy is also about knowing different ways of representing data (Ben-Zvi& Garfield, 2004) and knowing the statistical concepts required (Ridgeway et al., 2011).

One model considers reading values, comparing values, and analysing the dataset as necessary skills for being statistically literate (Chick & Pierce, 2013). This model is the development of a previous framework, which was about reading, reading between, and reading beyond the data (cf. Curcio, 1987). Another model distinguishes the questions needed to be asked when solving a task (delMas, 2002). In this model, it is about the nature of the tasks— what we ask the students to do with the content. Basic literacy in this model is described as being able to identify, describe, rephrase, translate, interpret, and read the data. A third model finds the inconsistence between the definitions of statistical literacy so broad that the concept is divided into statistical competence and statistical citizenship (Rumsey, 2002).

One frequently used framework in the field is Pedagogical Content Knowledge (PCK) (e.g., Arnold, 2008; Callingham & Watson, 2011; Callingham et al., 2016; Chick & Pierce, 2008a; Gonzáles, 2016; Groth, 2007, 2013; Groth & Bergner, 2005; Heaton & Mickelson, 2002), upon which Statistical Knowledge for Teaching (SKT) has been developed (e.g., Groth, 2007; Callingham & Watson, 2011; Chick & Pierce, 2008 a, b). SKT separates mathematical and nonmathematical knowledge. Mathematical knowledge could be about reading the data, while nonmathematical knowledge could be about how to pose questions and understanding the context from which the data were generated (Groth, 2007). A further development encompasses teachers' conceptions about teaching statistics (Gonzáles, 2016). As we can see in different models, statistical literacy is multifaceted and involves both noncognitive as well as cognitive factors (e.g., Bond et al., 2012). In different ways and to varying extents the abilities to interpret, evaluate, and communicate statistical information are the focus. One model by Gal (2004), reveals the complexity of statistical literacy (see Figure 2, below).

Figure 2

A model of statistical literacy. (Original table see Gal, 2004)

Knowledge elements	Dispositional elements
Mathematical knowledge Statistical knowledge Literacy skills Context knowledge Critical Questions	Beliefs and Attitude Critical stance
Statistical literacy	

In Figure 2, we can see how statistical literacy is divided into two different groupings—knowledge elements and dispositional elements. This theoretical construction, taking into account these different elements, is accepted and used in much research on statistical literacy (e.g., Bond et al., 2012; Callingham & Watson, 2017; Gal, 2002, 2004, 2019; Sharma, 2017; Shaughnessy, 2007). It is a mix of the knowledge elements that build the knowledge base, which enable the literate behaviour of understanding and be able to interpret statistical information (Gal, 2004, 2019). When a person understands and can interpret statistical information, it is possible to take a critical stance, something that is dependent on the dispositional elements (Gal, 2004). Altogether, the model in Figure 2 highlights that statistical literacy is not only a subset of formal statistics, but a complex competency where all elements are involved (Gal, 2019).

Another model incorporates ‘task’ as part of the model (e.g., Watson, 2013). Similar to Figure 2, mathematical, statistical, and literacy skills, and context are included. Specific knowledge of average, variation, and data collection as well as task format and task motivation are also included. The model is founded on a three-level hierarchy, namely, (a) a basic understanding of statistical terminology, (b) embedding of language and concepts in a wider context, and (c) questioning of claims. The model’s intention is to facilitate encounters with different aspects of statistical literacy and enable the understanding and critical evaluating of results from our daily lives (Watson, 1997). This model has developed to encompass six stages of statistical literacy characteristics, where each stage is described from the perspective of context: sampling, representation, average, chance, inference, and variation (Callingham & Watson, 2017; Watson, 1997).

2.5 Previous studies of textbook tasks on mean, median, and mode

Textbooks are important tools in school since they promote much of what is afforded for learning in the classroom (e.g., Weiland, 2019b). In statistics education, many types of analysis have been done on textbook tasks (e.g., Büscher, 2022a, 2022b; Glasnovic Gracin, 2018; Huey & Jackson, 2015; Jones & Jacobbe, 2014; Jones et al., 2015; Pickle, 2012; Weiland, 2016, 2019b). For instance, several studies show that textbook tasks to a high degree focus on procedures, such as how to construct diagrams and calculate measures, and to a lesser degree on conceptual content (e.g., Büscher, 2022a; Glasnovic Gracin, 2018; Pickle, 2012). Others focus on the distribution and the location of the topic within the textbooks (Jones et al., 2015; Pickle, 2012). Some of the categories related to problem-solving process are, Formulate Questions, Collect Data, Analyse Data, and Interpret Results (Jones et al., 2015). Then, there was an overwhelmingly large proportion of tasks in the data analysis phase with a significant emphasis on procedures. This means that students involved in these tasks will have a different opportunity to learn statistics with the help of problem solving. The same phases were the focus of a textbook analysis used in a course for elementary teachers (Jones & Jacobbe, 2014). Initially, besides the phases, the teachers were expected to analyse all tasks as either procedural or conceptual. They found this difficult since tasks were found to be procedural or conceptual to varying degrees. The starting point was that procedural tasks above all encouraged the use of calculations. However, there were procedural tasks that did not involve any calculation, which is why other codes were used. Procedural tasks involved either constructing or reading a display, or performing a mathematical computation. Conceptual tasks required statistical reasoning beyond calculation and calculation alone was not sufficient for solving the task. Their results showed that in two of the textbooks most tasks involved conceptual knowledge while three textbooks paid more attention to procedures. Pickle (2012) compared different textbooks focusing on the proportion of pages devoted to various statistical content, of which measures of central tendency was one. The distribution varied, with one textbook having six times as many pages about the mean than the mode, while another had about the same number of pages for each measure. This study used an analytical framework with four levels of cognitive demand. The majority of tasks were on a low cognitive demand level, with hardly any at the highest level.

Some studies focused on the context of the task. One study identified the role of context presented a framework for identifying inferential reasoning tasks (Huey & Jackson, 2015). For instance, if a task can be solved while ignoring the context it is analysed as low. In addition, tasks that only required a numerical answer were coded as low. To be coded as high, the context needed to be incorporated in the task. Weiland (2016, 2019b) focused on types of data provided in the contexts. He found that many tasks provide meaningless contexts at a superficial level (Weiland, 2016). More explicitly, contexts in tasks were found to be fictional and neutral (Weiland, 2019b) and did not meet the criteria for preparing students to become critical citizens, for which they need to encounter real data in order to make sense of statistical messages in real life (Weiland, 2019b). It is not only context, but also design of a task that is important (Büscher, 2022b). To teach for statistical literacy, tasks need to train students not only to read statistical information but also to imagine underlying data.

In sum, the studies in the above review suggest that similar needs are expressed in different ways. Some of the gaps identified include the need for tasks involving reasoning beyond procedural knowledge (e.g., Jones et al., 2015), the need to create opportunities for reasoning and making sense of data (e.g., Huey & Jackson, 2015), and the need for teachers to adapt textbooks to afford students opportunities to develop statistical literacy (e.g., Büscher, 2022a).

2.6 Research on mean, median, and mode

The decision in this thesis is to focus on the measures of central tendency: mean, median, and mode. When looking at previous research about these concepts, one can conclude that there has been a strong focus on mean (e.g., Jacobbe & Carvahlo, 2011; Leavy & O'Loughlin, 2006), less focus on median (e.g., Lesser et al., 2014; Watson, 2013), and even less on mode (e.g., Groth & Bergner, 2006, 2013; Watson, 2013). One reason for less research about median, could be that the calculation is regarded as being easier than that for the mean (Lesser et al., 2014; Zawojewski & Shaughnessy, 2000). It is fair to assume that similar reasons apply to the research on mode.

Starting with the research on mean and median, previous research has had a strong emphasis on the reliance on computation of the mean (e.g., Cai & Moyer, 1995; Leavy & O'Loughlin, 2006; Watier et al., 2011). However, doing statistics is not equivalent to understanding statistics (Gal, 2000). Researchers have identified a need for knowledge other than of the procedures, for example, recognising contexts where pinpointing the mean is important (Leavy & Hourigan, 2016; Leavy & O'Loughlin, 2006; Watson & Moritz, 2000), or deciding whether mean or median is a valid measure (Groth, 2013; Leavy & O'Loughlin, 2006; Zawojewski & Shaughnessy, 2000). The latter depends on understanding median as a valid measure; an understanding that involves knowledge about the distribution and any extreme values in a dataset (Leavy & O'Loughlin, 2006). Other research findings show that median is hard to identify in a set of unordered data (Zawojewski & Shaughnessy, 2000), or in a graph (Friel & Bright, 1998). One possible explanation could be that having conceptual knowledge for the median involves, for example, an understanding in relation to the distribution—half of the values are higher, and half are lower (Schnell & Frischemeier, 2019).

In statistical literacy, the affective perspective is acknowledged as important (e.g., Carmichael et al., 2009; Carmichael et al., 2010; Gal, 2002). Related to affective perspective is the relationship between intuitive ideas and the understanding of measures. Sometimes, these intuitive ideas can affect understanding of measures in a positive way, for instance, the analogy of mean as a balancing point or as fair share (Leavy & O'Loughlin, 2006; Watier et al., 2011). However, an intuitive idea of median as the middle is not helpful in the same way (Schnell & Frischemeier, 2019). This is a good example of how the different use of words can have a strong impact on the understanding of the concept.

Looking at the few studies focusing on mode, we find studies investigating why it could be difficult to distinguish between the mode and the frequency (Groth & Bergner, 2013; Watson, 2014). This includes the understanding of the mathematical property that mode is the only measure suitable for nominal data where mean and median are not applicable.

2.7 Students' understanding and conceptions

Students' understanding of mean, median, and mode was the focus of an extensive review made by Garfield and Ben-Zvi (2007), who concluded that students' understanding of these measures often show a lack of conceptual understanding beyond the algorithm. For instance, many students know the algorithm for mean, but do not know, or pay attention to, the mathematical properties of the measure (Mathews & Clarke, 2003; Mokros & Russell, 1995). Students using an algorithmic approach without engaging the meaning of the measure could benefit from, for instance, an engagement with when the measures are most reasonably used (Mokros & Russell, 1995). Another option is to engage the balance metaphor in their way towards a definition of the measure (ibid.). One reason why students struggle with reasoning about the representativeness of measures is that they do not consider distribution (Groth, 2013; Leavy & Hourigan, 2016; Leavy & Middleton, 2011). Sometimes they consider the image of the data, where they look for clumps or clusters rather than the spread, leading to the most frequent value (Mokros & Russell, 1995).

Another aspect that has been identified is the focus on particular words being used: the word 'typical' or 'typicality' was used to develop students' notions of representativeness (Leavy & Middleton, 2011). However, it required an instruction about the representativeness of all values to move students' attention from clusters to variability. Research like this emphasis the need for research connecting the use of words to the possible understandings. Students often form intuitive understandings built upon colloquial interpretations of everyday words (Watier et al., 2011). These intuitive understandings are hard to change, and need to be unpacked with the students (Russell & Mokros, 1991). Looking at previous research, we see that one common intuitive idea built upon colloquial interpretation, is portrayal of the median as the centre/middle (Cooper & Shore, 2008; Mayén & Diaz, 2010; Schnell & Frischemeier, 2019). Mode is commonly presented as 'the most '(Mokros & Russell, 1995). As mentioned above, 'typical' is such a colloquial word. One study focusing on the word typical resulted in four different interpretations—that which: was reasonable, was the most common value, contrasted with atypical, and was typical beyond the context of their own classroom (Makar, 2014). When elaborating on the measure 'typical height', students paid attention to reasonable values for a known context, their own classroom. Then, they were able to link their concepts with the meaning by utilising the context. Situations or contexts are important when conceptualising the measures, as in the above example where students worked on connecting reasonable values to a context (Makar, 2014). Research has concluded that procedural knowledge alone is insufficient for using the measures in a meaningful way in different contexts (Konold & Pollatsek, 2002; Watson, 2013). Hence, being statistically literate means understanding the context and interpreting the results in relation to the context (Gal, 2019; Moore, 1990). This falls under the concept of context knowledge: to recognise suitable contexts for the different measures (Watson & Moritz, 2000).

Students also need to be afforded different visual aspects of distribution, for instance different graphs, to identify datasets (Schnell & Frischemeier, 2019). Here, students were helped to reason about the median using a graph as the starting point, something that became the basis for developing an understanding of the concept. Leavy and Hourigan (2016) found similar results. However, others have concluded that for students to interpret the measures, they need to learn to pay attention to different methods for different graphs (e.g., Cooper & Shore, 2008, 2010).

Besides mathematical and statistical knowledge and context knowledge, the affective part of statistical literacy also includes students' interest in statistics, for instance, how they express enjoyment and how they perceive their own competency (Carmichael et al., 2009; Carmichael et al., 2010). Previous studies report that positive attitudes are connected with higher achievement (Leavy & Hourigan, 2016; Ramirez et al., 2012), and that students' beliefs impact their attitude and conceptual understanding (Bond et al., 2012). On a more general level, it has also been reported that activities that require active participation and opportunities to engage in reasoning enhanced students' statistical literacy (Hourigan & Leavy, 2020). Together, these studies confirm that the elements in the theoretical model proposed by Gal (2004) are dependent on each other: knowledge elements depend on dispositional elements and vice versa.

2.8 Teachers and prospective teachers' understanding and conceptions

Over the years, there have been several reports on teachers' understanding and conceptions of the measures (e.g., Jacobbe & Carvalho, 2011; Jacobbe, 2012; Leavy & O'Loughlin, 2006; Peters et al., 2014; Russell

& Mokros, 1991), as well as prospective teachers (e.g., Groth & Bergner, 2006; Hannigan et al., 2013; Leavy 2010; Leavy & O'Loughlin, 2006; Leavy et al., 2021; Sorto, 2004). As I noted in section 2.7, regardless of the respondent group, the results signal the same lack of conceptual knowledge and high reliance on procedures as for students. As an example, one study compared prospective teachers' and students' knowledge about distribution when comparing two datasets (Canada, 2008). Both groups expressed limited conceptions with few considering the variability of the dataset. For median, results showed a confusion with mid-range (the arithmetic mean of the maximum and minimum values of the dataset)—students not knowing it could be a value not represented in the dataset, and trying to find a median of categorical data (Leavy & O'Loughlin, 2006; Leavy, 2015).

Further, results show most prospective teachers' descriptions of the mode are numerical (Groth & Bergner, 2006). This affects their knowledge about the mode, as this excludes knowledge about mode in qualitative datasets (Groth & Bergner, 2006, 2013). Results also show difficulties with identifying the mean and/or the median in graphs (Jacobbe, 2012; Leavy, 2010; Peters et al., 2014). Another example concerns insecurity about the representativeness of mean or median regarding outliers (Jacobbe, 2012; Sorto, 2004). For one teacher, it was not possible to reason about this, since they did not know what the median actually measures (Jacobbe, 2012). Other studies show that teachers do not use the mean to compare groups (Leavy & O'Loughlin, 2006; Makar & Confrey, 2004; Peters, 2009), and little knowledge is demonstrated about what the measures represent within a context (Jacobbe, 2012). Some of these studies involve training teachers to develop conceptual knowledge. One example shows how working with reflective activities and engagement in multiple representations succeeded in broadening their conceptions (Jacobbe, 2012; Peters et al., 2014). For instance, working with a dot plot as a representation was one way to experience the mean as a balancing point (Jacobbe, 2012).

There are fewer studies about dispositional elements compared to studies about cognitive issues (Estrada et al., 2011). Research in this field is about statistics in general, not specifically about the measures of central tendency (Groth & Meletiou-Mavrotheris, 2018). Here is a gap in the research body. In some of the main findings to date we see that respondents' attitudes are linked to their understanding of statistical concepts, which also affects their teaching of statistics (e.g., Chick & Pierce, 2008b; Estrada et al., 2005). This is in line with the results from

Hannigan and colleagues (2013) who found a connection between positive attitudes and conceptual knowledge, although not a strong correlation. It has also been shown that teachers' attitudes towards statistics affect their willingness to update both their own knowledge and their teaching (Martins et al., 2012). Other research findings show that experience affect teachers' conceptions and their knowledge of students' common ways of thinking (Cai & Gorowara, 2002). When comparing more and less experienced teachers' conceptions about the mean, the experienced teachers showed ability to predict both errors and representations common among middle-school students, whereas inexperienced teachers used and valued algebraic procedures.

Teachers' knowledge is not only about the content, but also about understanding the students and the ability to start remediation at the students' level (Watson & Callingham, 2013). Crucial factors affecting prospective teachers' attitudes to statistics include different ways of reasoning, ambiguous language, and how the content is connected with contexts (Leavy et al., 2013). If we want to frame some of these factors, we therefore require different theoretical underpinnings.

3 Theoretical considerations

> The time may not be very remote when it will be understood that for a complete initiation as an efficient citizen of one of the new great complex world wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write. (Wells, 1903, p. 204).

Given the research questions posed in Figure 1, which all focus on different aspect aspects of statistical literacy, further theoretical considerations had to be made. This section will present these considerations, in particular the theoretical units presented in the statistical literacy model in Figure 2.

3.1 Aspects of statistical literacy

The research questions posed in each of the studies span different aspects of the model developed by Gal (2004, 2019). In studies 1–3, the research questions use the terms conceptions or motivations as part of conceptions, for example, how an individual might think that one of the measures is more useful than the other. The decision here is to treat conceptions as a subjective truth, and these research questions fall in the category Dispositional elements. Questions that focus on mathematical aspects of the different measures, or how individual's relate different measures to different use of words is instead considered Knowledge elements. The last type of research question is about context, which is also considered a knowledge element. Therefore, in this thesis, an adapted version of the model in Figure 2 has been used, see Figure 3:

Figure 3

An adapted model concerning some aspects of statistical literacy

Knowledge elements	Dispositional elements
Mathematical knowledge Use of Words Context knowledge	Conceptions
Aspects of statistical literacy	

In the following section, I discuss some theoretical considerations regarding the knowledge elements and dispositional elements.

3.1.1 Mathematical knowledge

Starting with the knowledge elements in the adapted model (Figure 3), the decision is to look at mathematical knowledge and statistical knowledge as one unit, which is in line with the Swedish curriculum where statistics is part of Mathematics. I will here discuss mathematical knowledge, first in terms of mathematical properties, and then in terms of procedural or conceptual knowledge.

Mathematical properties

In this thesis, the focus is on the three concepts mean, median, and mode. They are then analysed with respect to their different mathematical properties. Here, a mathematical concept is understood as being composed of different dimensions of which mathematical properties is one—a dimension useful when evaluating mathematics learning (Usiskin, 2015a). Concepts in statistics are particularly interesting given that one must also consider the context (Gal, 2019). One example is being able to choose which measure is most appropriate in a certain dataset, and then being able to perform the transformation (the procedure). The answer must then be interpreted in relation to the context. It is therefore critical, according to Usiskin (2015b), to be able to identify relevant mathematical properties.

Lithner (2008) refers to such properties as intrinsic, and since there are several different types of mathematical arguments—identifying, predictive, verifying, and evaluating (Eriksson & Sumpter, 2021)—different mathematical properties will be intrinsic and thereby relevant depending on whether you are discussing why a certain strategy choice (transformation) will solve the task (predictive arguments) or why the conclusion is a relevant answer to the task (evaluating arguments).

Dealing with algorithms is just one aspect of many needed to understand the statistical concept (Usiskin, 2015a). The ability to identify and use relevant mathematical properties involves being able to deal with datasets, and understanding some fundamental ideas in statistics (Burrill & Biehler, 2011; Weiland, 2019a, 2019b; Wild & Pfannkuch, 1999). For instance, being aware of different types of variables, qualitative (categorical) and quantitative (numerical), and the relationships between them is important (Burrill & Biehler, 2011; Hall, 2011), as well as suitable graphs depending on the variable (Wild & Pfannkuch, 1999). In teaching, students first encounter qualitative data, then they move on to quantitative data, and in parallel they also encounter variation within a data set, between groups of data, as well as variation between data sets from different contexts (MacGillivray & Pereira-Mendoza, 2011). All these ways of changing data, like from calculations into different representations, has been termed transnumeration (Wild & Pfannkuch, 1999). It is not the mathematical calculation—the transformation—built upon the definition that signals conceptual knowledge, such as why a measure is representative in a dataset (Watson & Fitzallen, 2010). Key features like these must be understood in order to understand different ways of summarising statistical information, such as measures of central tendency (Wild & Pfannkuch, 1999). Also, one must pay attention to the different mathematical properties generated by a transformation (e.g., Lithner, 2008). Hence, examining an algorithm, for example based on how mathematical properties are generated through the transformation, provides a basis for developing conceptual understanding (e.g., Groth, 2014). Hence, understanding the context with respect to mathematical properties can also be on a micro-level and not just on a macro-level as described by McCright (2012) and Büscher (2019).

The following mathematical properties have been identified for mean (Strauss & Bichler, 1988): (1) The average is located between the extreme values; (2) The sum of the deviations from the average is zero; (3) The average is influenced by values other than the average; (4) The average does not necessarily equal one of the values that was summed; (5) The average can be a fraction that has no counterpart in physical reality; (6) When one calculates the average, a value of zero, if it appears, must be taken into account; and (7) The average value is representative of the values that were averaged. As stated before, these properties describe the mathematical concept mean, but offer no further mathematical analysis. For instance, that which is related to the different transformations and that which is related to the data are not explicit. Therefore, we cannot compare these categories to the other measures.

Median deals with quantitative values and sometimes ordinal values as well, depending on the dataset (e.g., Mayén & Diaz, 2010). Median has some properties in common with mean, for instance, that it might not equal a value in the dataset (Groth & Bergner, 2006), or be a value that does not correspond with physical reality. A joint property of mode and median is that they are not affected by extreme values (Groth & Bergner, 2006). When it comes to the properties for mode, there might be one mode, or several, or none; and it is unique applicable for nominal data (e.g., Groth & Bergner, 2006).

A mathematical property for one object could be equivalent to another object, but at the same time differ in other properties leading to different categorisations (Mason, 2011). An example of this is the square and rhombus that both have four corners and sides of equal length but differ in the size of the angles. This is how different levels of measures are categorised. All data can be grouped/classified, some data can be ranked ordered, and finally some data can be calculated with the algorithm for the arithmetic mean. As a consequence of these different transformations—group, rank order, or calculation—new properties arise (e.g., Mason, 2011). Different transformations evoke different properties. As nominal level is defined, data can only be grouped. Ordinal data can be grouped but also rank ordered. This means that it is the possible transformations of qualitative data that entail fewer properties compared to transformations of quantitative data (Byström & Byström, 2011; Groth & Bergner, 2006; Jacobbe & Carvahlo, 2011). The possible transformations also imply limitations to analysing qualitative data compared to quantitative data (Leavy, 2010).

Here, the theoretical framing of mathematical property comes from Lithner (2008). The starting point is the definition that central mathematics ideas of a concept are "built on a set of objects, transformations, and their properties" (Lithner, 2008, p. 261). Compared to Strauss and Bichler (1988), this definition allows one to separate different mathematical properties. Operationalising the three entities—objects, transformations, and concepts—one also takes into account their relationships: how mathematical properties of an input object (the data level of the dataset) affect what transformation is possible, as well as how these affect the result, the output object (e.g., Lithner, 2008; Mason, 2011). Since one of the aims of this thesis is to identify and analyse the characteristics of mathematical properties in tasks, the context, the transformation, and the calculated answer are relevant for the analysis. The chosen theoretical framing then makes numbers in context explicit (e.g., Nilsson et al., 2018), and helps to identify and provide insights into what learning the data in a task enables (e.g., Gal, 2019).

Conceptual and procedural knowledge

Another way to talk about mathematical knowledge is to apply the theoretical framing that uses the concepts conceptual and procedural knowledge (e.g., Hiebert & Lefevre, 1986). This involves different aspects of how we learn mathematics such as processes required to perform a task, as well as understanding the action taken (Murray & Rodney Sharp, 1986). Different theories of learning have dealt with this duality for more than a century, and the two kinds of knowledge are sometimes referred to as skill and understanding (Hiebert & Lefevre, 1986). One commonly used definition is provided by Hiebert and Lefevre (1986). In short, they define procedural knowledge either as composed of formal language and representations or as the algorithms and rules needed to solve a task. Conceptual knowledge is defined as rich in relationships, and because of this, cannot be an isolated piece of information. There are several similar descriptions/definitions of procedural and conceptual knowledge. There are descriptions of procedural knowledge concerning knowledge about the steps/procedures required to solve a task (Byrnes & Wasik, 1991; Canobi, 2009; Rittle-Johnson & Alibali, 1999). Conceptual knowledge can be thought of as abstract and general knowledge of concepts (Byrnes & Wasik, 1991; Canobi, 2009), or as an understanding of a domain and domain-specific knowledge (Rittle-Johnson & Alibali, 1999). However, it is not always easy to separate the two kinds of knowledge, and sometimes it is impossible (Hiebert & Lefevre, 1986; Rittle-Johnson & Alibali, 1999; Star, 2005).

There is an ongoing discussion if the definition by Hiebert and Lefevre (1986) means that the view of procedural knowledge is considered to be at a lower level of knowledge (e.g., Gal, 2014; Rittle-Johnson & Alibali, 1999; Star, 2005). One model argues that knowledge lies on a continuum, with procedural and conceptual knowledge at each end point (Rittle-Johnson & Alibali, 1999). Another model brings in the distinction between superficial and deep knowledge (Star, 2005), considering that procedural knowledge is also flexible (Rittle-Johnson et al., 2012). The background to this is that procedural knowledge is often considered to play a secondary role in students' learning, not linked to quality in the same way as conceptual knowledge is (e.g., Gal, 2014; Star, 2005). Gal (2014) concludes that procedural and conceptual knowledge seems to be intertwined, and that understanding the algorithm—having strong procedural knowledge—leads to increased opportunities to develop conceptual knowledge.

Another issue is that students show procedural fluency without any conceptual knowledge (Silver, 1986). The process of calculating a measure is just a way to gain information, but it does not demonstrate ability to understand any properties of the measure or how it is used (Rumsey, 2002). The theory of procedural and conceptual knowledge would explain this as better transfer requiring both procedural and conceptual knowledge (Silver, 1986). Procedural and conceptual knowledge benefit each other bidirectionally (Canobi, 2009; Rittle-Johnson et al., 2015). This approach builds on the notion that students develop knowledge about concepts through an iterative process where concepts and procedures influence each other (Canobi, 2009). Some studies have also concluded that procedural and conceptual knowledge depend on each other: that procedural knowledge is essential for conceptual knowledge (e.g., Bakker, 2004; Groth, 2014). Almost 20 years ago, Star (2005) brought this dependence to attention and concluded that we needed more research about procedural knowledge. In this thesis, I contribute by helping to address this gap using the definition of procedural and conceptual knowledge by Hiebert and Lefevre (1986) as presented above. In Paper 5, I discuss these concepts using the theoretical framework presented by Lithner (2008) to reveal which mathematical properties are intrinsic and relevant, and which are superficial.

3.1.2 Use of Words

The second knowledge element this thesis focuses on is the use of words. Below, I discuss theoretical considerations related to this concept—both the character of words and how to teach statistics with a focus on the words used.

Character of different frequently used words

Previous research has established that learning statistics involves linguistic challenges (e.g., Dunn et al., 2016; Richardson et al., 2013). Looking at the words that are used draws upon both general language and specific words (Dunn et al., 2016). Some words are lexically ambiguous, having several meanings apart from their everyday meanings (Barwell, 2005; Rubenstein & Thompson, 2002). Other words might be homonyms—in this case, two or more words spelled and pronounced the same but different in meaning (Rubenstein & Thompson, 2002), and sometimes students transfer a word's everyday meaning to the mathematical term (Kaplan et al., 2009, 2010). Using a colloquial word in a tautological way, like 'median is the middle', does not convey any statistical meaning (Clark et al., 2007). There are also terms only used in mathematics, and words having different meaning within different disciplines (Rubenstein & Thompson, 2002).

'Average' is a word with many connotations (Mokros & Russell, 1995; Watson, 2013). For a student, everyday situations where they encounter words like typical, normal, usual, middle, or most, convey different associations (Kaplan et al., 2014; Makar & McPhee, 2009; Mokros & Russell, 1995; Watson, 2013). Sometimes, average is used to refer to all measures of central tendency, sometimes to a particular measure (Kaplan et al., 2010). One study found that most students mentioned the mean specifically when they related average to one measure (Kaplan et al., 2010). Different intuitions about the term 'average' may arise from different uses of the term in the media (Watson, 2013).

Teaching on statistics and use of words

Since the term average has other meanings outside school mathematics, it is important for a teacher to be aware of the possibility that students already understand the term in another way (Watson, 2013). They rely on previous understandings and therefore often use a colloquial understanding in Mathematics (Kaplan et al., 2009; Konold, 1995; Makar & McPhee, 2009). It is hard to change students' adherence to such intuitive meanings, which may even contradict meanings a teacher wishes to convey (Konold, 1995). Students need to encounter the concepts in meaningful contexts to link the statistical concepts to their meaning (Kaplan et al., 2009; Makar & McPhee, 2009; Mokros & Russell, 1995). Allowing students to work with the meanings of a word, for instance, 'typical', in mathematical contexts helps them in terms of developing their conceptions of the word. This is a way to get them to reconsider the meaning of a colloquial word in the context of mathematics (Makar & McPhee, 2009; Mokros & Russell, 1995). Being unfamiliar with the meaning of a word hinders students from communicating understanding (e.g., Gal, 2002; Usiskin, 2012; Weiland, 2017). In general, the chances of remembering and understanding a word depend for instance on the amount of exposure to the word and manipulation of its properties (Schmitt, 2008). However, undeveloped notions of 'average' can also have been assimilated at school, since much time is spent on definitions and performing procedural calculations (Mokros & Russell, 1995). This could lead to confusion about the measures, and talking about the middle when explaining the mean (Mokros & Russell, 1995). It is as if the algorithms disconnect previous informal understandings (Russell & Mokros, 1991).

Another approach is deliberately working to decrease the use of colloquial words and increasing the use of the mathematical terms, for instance, replacing 'spread' with 'variability' (Kaplan et al., 2012). This is because colloquial words have meanings not related to the mathematical term, which promotes misunderstandings. Besides, a colloquial word does not carry the statistical meaning that the mathematical term does (Kaplan et al., 2012; Rubenstein & Thompson, 2002). Thus, according to Kaplan et al. (2012), even textbooks should follow this approach and teachers ought to refer to measures of centre or the name of the particular measure, instead of using the word average (Kaplan et al., 2010). However, if students do not learn to detect lexical ambiguity, they will not assimilate the correct meaning of ambiguous words (Kaplan, et al., 2010; Russell & Mokros, 1991). If students are allowed to use colloquial words, they will find the statistics easier and have the opportunity to understanding of the word correctly (Makar & Confrey, 2005). Using different terms for the same concept in different learning resources can also be helpful (Dunn et al., 2016). In teaching mathematics, we strive for learners to reach a context-independent generalisation of a mathematical concept (Mason, 2011), the meaning of which, as described in this section, is strengthened through working with contexts.

In this thesis, the aim is not to analyse the meaning of the concepts with respect to use of words. Instead, the study of the use of words is based on how the respondents express themselves, how they define concepts, and what lexical ambiguities occur (e.g., Dunn et al., 2016; Kaplan et al., 2009).

3.1.3 Context knowledge

The third knowledge element in the adapted model (Figure 3), is context knowledge. As stated earlier, statistics education theories need to address context as part of the content, in a different way compared to mathematics education (Nilsson et al., 2018). Here, statistics can be equated with numbers in context (Moore, 1990), but contexts in statistics can be both a source of meaning—which procedures to apply—and support for interpreting a result (Gal, 2019). Therefore, in task design, context knowledge is separated into different aspects of context, and the ability to place a concept in a specific context.

Different aspects of context in task design

Context knowledge can be related to different aspects (Greatorex, 2014), one of which is when context depends on principles for task design related to various philosophies of mathematics education (Kieran et al., 2015). Some examples are Anthropological Theory of Didactics (ATD) and Realistic Mathematics Education theory (RME), where contexts within ATD generate questions beyond school mathematics, while contexts within RME trigger solutions with the potential to be mathematised (Kieran et al., 2015). A common idea in the above examples, is that the context builds upon students' understanding (Da Ponte et al., 2013; Kieran et al., 2015). Within statistics, an underlying idea is that students learn statistics when they explore the context in the task (Büscher & Prediger, 2019; Da Ponte et al., 2013; Weiland, 2019a, 2019b). An example is how investigating real statistical information in a context like climate change develops statistical knowledge (e.g., Büscher & Prediger, 2019). However, many tasks in textbooks are dressed up bare tasks, where context has no connection to reality; they are, rather, fictitious contexts about peoples' age or favourite animal (e.g., Verschaffel et al., 2000; Weiland, 2019a, 2019b; Wijaya et al., 2015).

Another aspect of context knowledge is when a context relates to instructional aspects in a task, or how the mathematical concepts are engaged (Charalambous et al., 2010; Da Ponte et al., 2013; delMas, 2002; Watson & Mason, 2006). This could be how a solution can help students to develop mathematical concepts (Hong & Choi, 2018; Kieran et al., 2015; Reinke, 2019; Verschaffel et al., 2000; Watson & Thompson, 2015), or to focus on the mathematical properties of a concept in the context (e.g., Da Ponte et al., 2013; Mason, 2011). Hence, different approaches generate different opportunities to learn, and in this thesis the aim is towards mathematical content—more specific the mathematical properties (e.g., Kieran et al., 2015; Shahbari & Tabach, 2020). One example is directing students' attention towards classifying and characterising the mathematical content (Doyle, 1983; Mason, 2011). In these kinds of task, the text must help the student towards acquiring conceptual knowledge; the context must evoke some further engagement that goes beyond mere calculation (Kieran et al., 2015; Shahbari & Tabach, 2020; Watson & Mason, 2006). Design of tasks is often discussed from the perspective of cognitive or technical demand (what level of mathematics the task requires, from elementary to sophisticate) (e.g., Burkhardt & Swan, 2013; Stein & Smith, 1998). Non-contextual computational tasks generally have lower cognitive demands than many contextual tasks (Doyle, 1983).

Since the aim in Paper 4 and Paper 5 was to analyse tasks with respect to what mathematical properties they afford, the decision here is to categorise for context knowledge (contextual tasks) only if a task affords opportunities to learn about mathematical properties (e.g., Kieran et al., 2015; Mason, 2011). While this is a limited operationalisation, it may be necessary if we want to analyse many tasks in a short period of time.

Concepts in contexts

Students encounter definitions of a concept in school; however, these definitions do not on their own provide a deep understanding of the concept (Schmitt, 2008). Students need to be exposed to a specific word/concept several times in many different contexts to learn about the concept and how to use the word correctly (Schmitt, 2008). If the context is familiar, it promotes understanding of the term and its use (Ainley & Pratt, 2017; Usiskin, 2012). Placing a word in different contexts helps to draw attention to lexical ambiguity (Richardson et al., 2013). Similarly, placing a word in the right statistical context can offer an idea of the statistical meaning (Richardson et al., 2013). Another way is to use contexts where statistical meaning and colloquial meaning coincide (Durkin & Shire, 1991). It turns out that engagement with increasingly complex or unfamiliar contexts is associated with higher-level performance in terms of statistical literacy (Watson, 2013).

In Paper 3, the aim was to study context knowledge qualitatively, and then define it with respect to its being a source of meaning (e.g., Gal, 2019). The starting point is based on previous studies that show that students demonstrate context knowledge from own experiences, and how these descriptions relate to real-world contexts (e.g., Watson, 2013).

3.1.4 Conceptions

In this thesis, an adaption has been made for dispositional knowledge: to include both beliefs and attitude, and critical stance in the notion of conceptions. The reason is that different dispositional elements are used and defined—if they are defined at all—in various ways in different studies (e.g., Bond et al., 2012; Carmichael et al., 2010; Gal et al., 1997; Meletiou-Mavrotheris & Lee, 2003; Sproesser et al., 2016). Since it is difficult to separate different dispositional elements, such as how stable a belief is (e.g., Goldin, 2002), or how true it is (e.g., Philipp, 2007), the choice here is to use the overarching construction of conception, defined as "a general notion or mental structure encompassing beliefs, meanings, concepts, propositions, rules, mental images and preferences" (Philipp, 2007, p. 259). This broad definition encompasses both beliefs (e.g., Schoenfeld, 1992) and mental images (e.g., Tall & Vinner, 1981). Beliefs in this sense refer to an individual's engagement based on both understanding and feelings (Schoenfeld, 1992), while mental images or concept images refer to an individual's mental picture, including properties and processes (Tall & Vinner, 1981). Furthermore, self-efficacy can also be seen as part of the affective aspect of conceptions (e.g., Beswick et al., 2012).

3.2 Affordance and opportunities to learn

Analysing tasks requires a way of evaluating the opportunities to learn they might bring, for example, what mathematical knowledge they afford. Affordance as a noun was introduced to the English lexicon by Gibson (1979), with the initial intention of describing perception in a specific environment. The term has since come to be used within design, where the meaning of affordance is that the user easy percept how to interact with an object—to afford is to give a clue (Norman, 1988). In mathematics, affordance refers to the opportunities to learn that a task offers (Chick, 2007; Chick & Pierce, 2012; Liljedahl et al., 2007; Watson, 2007). A teacher's ability to predict possible affordances depends both on their mathematical knowledge and level of statistical literacy (Chick & Pierce, 2012); for example, their ability to recognise mathematical properties or methods (Watson, 2007). Although an affordance is not always visible, still it exists, and individuals may perceive different affordances depending on the need, interest, and situation, (Gibson, 1979; Kress & van Leeuwen, 2001; van Leeuwen, 2005). Using this definition, affordances can be seen as the potential usefulness of objects (van Leeuwen, 2005).

A task has inherent affordances—the possible opportunities it offers for learning specific mathematical knowledge (Chick, 2007). However, identifying potential affordances is not always straightforward in a task, but could change if the task was adapted to better focus on possible learning opportunities (Chick, 2007; Liljedahl et al., 2007; Shahbari & Tabach, 2020). If the student receives that help in formulating the task, the task will provide the student an opportunity to develop the afforded learning opportunity (e.g., Da Ponte et al., 2013; Chick, 2007; Liljedahl et al., 2007). Liljedahl et al. (2007) describes this refinement of a task as a way to liberate the affordances.

As a teacher, you need to have mathematical knowledge to understand and identify what possible affordances a task offers, and to select good tasks with potential to provide the intended learning (Ball, 2000; Chick & Pierce, 2008a, 2008b; Liljedahl et al., 2007; Watson, 2007). You also need to help learners by directing their attention, for instance, towards the mathematical properties (Da Ponte et al., 2013; Watson, 2007). One study showed that primary school teachers had trouble identifying possible knowledge inherent in a real-world dataset (Chick & Pierce, 2008b). A third of the teachers in the study suggested mode as a suitable measure to the dataset although it was not useful in the specific case, many gave examples using averages in a vague way, and only one teacher gave a meaningful example of the mean related to the context (Chick & Pierce, 2008b). Another example involves the ages of people, 5, 6, 8, 9, 9, 9, 10, 13, 37, and 71, attending a birthday-party (Chick & Pierce, 2008a). This dataset will become an issue of interest in teaching only if the teacher is aware of the potential affordance that the median is a suitable measure of central tendency.

As described above, affordances predict possible opportunities to learn (OTL), and research shows that what students learn is consistent with afforded opportunities offered to them (Cogan et al., 2001; Stein et al., 1996; Tarr et al., 2006). OTL as a concept was used for the first time to study whether a student had had the opportunity to study a certain content (Husén, 1967). It has also been used as a validity check in comparative studies like PISA and TIMMS when comparing taught content in different countries (Floden, 2002).

In study 5, I use the definition that an affordance of a task is either visible or invisible (e.g., Gibson, 1979; Kress & van Leeuwen, 2001; van Leeuwen, 2005). This is a rather common treatment of the concept and allows researchers to study what possible opportunities there are to learn a specific content (Hadar, 2017; Kieran et al., 2015; Tarr et al., 2006; Törnroos, 2005; Watson & Thompson, 2015).

4 Methodology and methods

> […] researchers in mathematics education should never become wedded to a single approach, epistemology, paradigm, means of representation, or method. All are partial and provisional, none can tell the whole story. (Kilpatrick, 1993, p. 17)

Because of the variety of theories linked to mathematical elements and dispositional elements, exploring statistical literacy is a challenging enterprise. Statistical literacy is an extensive research area, which is why certain limitations have to be made, such as focusing mathematical knowledge on one mathematical content, here measures of central tendency, and the choice to treat all dispositional elements as conceptions. Such limitations help to diminish the number of theories to be examined, but it also creates limitations with methods and methodologies. The width of content in statistical literacy entails many possible choices of research methods. In addition, statistical literacy as a phenomenon is contingent upon and embedded in different contexts. To study a phenomenon in various contexts provides the possibility of exploring the phenomenon cross-contextually (Mason, 2006). In this study, a questionnaire was used as a research method, to collect first-hand data from specific groups of interest for the context of the study. The second approach was to conduct a textbook analysis which gathered second-hand data that already existed within the context. Below, I present the different theories and methods to justify the research questions in relation to the choices and limitations of the study.

4.1 Methodological considerations

Qualitative research aims to provide rich information to the understanding of the world we live in (Creswell, 2013; Mason, 2018). One's aim and research questions guide the process (e.g., Mason, 2018). The ambition of this study was to explore different aspects of statistical literacy concerning mean, median, and mode using data from teachers and prospective teachers of years 4–6, students in year 6, and textbooks tasks for years 4–6. The narrative tells a story built upon data. I would here like to stress that if I had made a different decision, there would be another story to tell (see Kilpatrick, 1993). This section therefore discusses the strategic decisions made to achieve the intended outcome of this study.

Decisions related specifically to the model of statistical literacy (see figures 2 and 3) where various ontological properties are interconnected under the umbrella of statistical literacy. The study explores these ontological properties and how they are interconnected (e.g., Mason, 2018). A multi-dimensional approach is required to illuminate these different properties, which includes various philosophies that tell different parts of the story. Table 1 provides an overview of the theories used in the papers.

Table 1

Distribution of theoretical underpinnings in papers 1–5

Theoretical framework/ concepts	Papers				
	1	2	3	4	5
Procedural and conceptual knowledge (Hiebert & Lefevre, 1986)	x	(x)*	(x)*	x	x
Statistical literacy (Watson, 1997, Gal, 2004)	x	(x)**	x	(x)***	(x)***
Conceptions (Philipp, 2007)	x	x	x		
Mathematical properties (e.g., Lithner, 2008, 2017; Strauss & Bichler, 1988)	(x)*	(x)*	(x)*	x	x
Affordance (e.g., Gibson, 1979; van Leeuwen, 2005)				x	x
OTL (e.g., Jones & Pepin, 2016; Törnroos, 2005)					x
Task design (Floden, 2002; Husén, 1967; Kieran et al., 2015; Mason, 2011)					x

Note. *) Appears as a part of the result; **) Aspects of statistical literacy is part of the analytical framework; ***) Mathematical properties is part of mathematical knowledge, one aspect of statistical literacy

In the papers, statistical literacy is covered in different ways and to different extents. A more detailed compilation is found in Table 4 (see 4.4.2). In this chapter, the papers will be presented in two groups based on research method: questionnaire (1–3) and textbook analysis (4–5). A description of the theoretical underpinnings of the methods follows below.

4.1.1 Questionnaires

The research question in Paper 1: How do prospective teachers conceptualise the concepts of mean, median, and mode to a student in years 4–6? aimed to investigate prospective teachers' conceptions. A conception in this study is understood and theoretically operationalised as "a general notion or mental structure encompassing beliefs, meanings, concepts, propositions, rules, mental images and preferences" (Philipp, 2007, p. 259). The reason for the choice of this definition was its coherence with statistical literacy, including both knowledge elements and dispositional elements. It means that in the answers, both affective components can be expressed (e.g., 'I like mode') and more cognitive elements (e.g., 'The median is in the middle'). The next choice was to place these conceptions in a framework for statistical literacy. The first attempt used a framework that has three dimensions, the three-tiered hierarchy (e.g., Watson, 1997). It enables the distinguishing of different aspects of statistical literacy, framed as (a) a basic understanding of statistical terminology, (b) embedding of language and concepts in a wider context, and (c) questioning of claims. This framework was chosen because the overarching aim of the study was to investigate statistical literacy. However, it did not allow me to be more specific about cognitive levels of mathematical knowledge, so the framework was combined with procedural and conceptual knowledge (see Hiebert & Lefevre, 1986).

Paper 2 was a comparative study investigating both prospective teachers', and teachers' conceptions of averages. As with the first paper, it started from Philipp's definition of conception (2007) since this was sufficient for the objective of the study. Some further elaborations with the definition showed that one can interpret 'meanings, concepts, propositions, rules, and mental images' as having cognitive dimensions. These epistemological assumptions meant that the theoretical framing provided by Hiebert and Lefevre (1986), and Watson (1997), was omitted in the second study.

In Paper 3, the aim was to investigate aspects of students' statistical literacy regarding measures of central tendency, using the model in Figure 3 as a framework. The adaptions were made as a result of the two previous studies that had a more inductive approach. One ontological adaption was the inclusion of statistical knowledge in mathematical knowledge, a decision in line with the Swedish curriculum. I also decided to maintain the theoretical operationalisation of dispositional elements as in papers 1 and 2, since such elements, like beliefs or attitudes, have many different definitions, which makes it difficult to separate different affective constructs (e.g., Leder, 2019; Nolan et al., 2012). Because the aim here is not to separate or to define dispositional elements from the perspective of specific theories, I made the decision to include all kinds of dispositional elements in the wide definition of conceptions by Phillip (2007). With respect to the mathematical properties of the mathematical concepts mean, median, and mode, there has been a progression too over the different studies, and this will be further discussed in the next section. Finally, Literacy skills was removed in favour of Use of words since Literacy skills involves more aspects than this study covers. Here, the focus is on the specific wordings, which could be compared to Watson's (1997) concept 'statistical terminology'.

4.1.2 Textbook analysis

In Paper 4, a short paper, the decision was to focus on mode. The paper examines the distribution between afforded procedural and conceptual knowledge (e.g., Hiebert & Lefevre, 1986). The knowledge types were distinguished depending on whether mathematical properties were afforded. Affordance is here seen and operationalised as the existence of properties, visible or not, and it is up to the reader to perceive their existence (e.g., Gibson, 1979; van Leeuwen, 2005). Mathematical properties relate to possibilities of data on different levels (e.g., Lithner, 2017). For mode, it is of interest to understand the limitation of data on a nominal level. Paper 4 also studies descriptions of the concept. This entails both tasks, definitions, and the words used. The meaning of words can be a hurdle for students (e.g., Richardson et al., 2013), not least in definitions (Perrett, 2012). The point is how important afforded descriptions are for how students conceptualise the mode.

Paper 5 is a more comprehensive study, including both mean, median, and mode. Based on the results from Paper 4, I developed a framework consisting of mathematical properties in input values, transformations, and output values (e.g., Lithner, 2008). The framework provided a distinct way to analyse all afforded mathematical properties connected with these three components (i.e., input values, transformations, and output values). One limitation with using the theoretical concept 'mathematical property', is when properties are missing in the framework. So, to discuss tasks from the perspective of mathematical properties, I had to distinguish tasks as contextual or not contextual, depending on whether they contained mathematical properties, either in the context or in the solution. This is different to how mathematical properties has been treated before (e.g., Lithner, 2008; Strauss & Bichler, 1988). Another theoretical dilemma was that affordances can be explicit or implicit in the design of the task, and that these "design decisions can easily hinder or support affordances of a task with respect to the intended object of learning" (Kieran et al., 2015. p. 47). Therefore, the decision here was to use the concept ‘Opportunities to learn (OTL)’, (e.g., Jones & Pepin, 2016; Törnroos, 2005), which allowed me to discuss possible afforded learning.

4.2 Sampling and methods

Sampling is seen as a procedure that provides access to relevant data, presenting opportunities for gaining insights about a context (e.g., Mason, 2018), which in this thesis is statistical literacy. There are, of course, several appropriate sources, and here, prospective teachers, students, teachers, and textbooks were chosen. What they have in common is school years 4–6. The choice is intentional—these are settings where aspects of statistical literacy can be expected, and the different samples provide possibilities for making comparisons within the context of statistical literacy. The chosen data sources are then elements connected with the aim of the study, hence allowing a multidimensional vision of context (e.g., Mason, 2006).

All papers (1–5) follow a qualitative research paradigm, use mixed methods and aim for a descriptive and comprehensive result (e.g., Biesta, 2012). The papers then operate on the micro-level of the context. Statistical literacy, which is the macro level, can be interpreted as experiences from both school and daily life, which through different methods can be studied via the lens of the micro. Research questions were therefore framed to focus on different instances of the micro level, and methods were then chosen to generate data in response to the questions (e.g., Mason, 2006). The thesis then follows the trajectory of a problem-driven thesis (Schoenfeld, 1992), and will offer the possibility of exploring how chosen dimensions intersect including relating meaningful data to the theory of statistical literacy.

A summary of different aspects of data collection is provided in Table 2:

Table 2

Summary of data collection

Data	Paper	Sample School years 4–6	Data collection
First hand	1	29 prospective teachers	Online questionnaire
	2	27 prospective teachers 36 teachers	Online questionnaire
	3	130 students aged 12–13	Online questionnaire
Second hand	4	7 textbook series Total 17 books 276 tasks	All tasks and definitions about mode in the textbooks
	5	7 textbook series Total 17 books 1392 tasks	All tasks about mean, median, and mode in the textbooks

The data collection is presented in two sections, below: first, data generated through questionnaires; second, data generated through textbook analysis.

4.2.1 Questionnaires

Sampling

Questionnaires are useful for exploratory research, and the selection of groups was based on accessibility, convenience, purpose, and willingness to participate, (e.g., Denscombe, 2017; Patton, 2002). Here, the sampling was made to classify characteristics of conceptions expressed by different groups: teachers, prospective teachers, and students. Prospective teachers were contacted via email, and students were contacted through their teachers. Teachers were contacted via email and Facebook. The main limitation was that the sampling was made in different ways, meaning that it was difficult to make comparisons. However, creating a sampling process where all three groups could be considered similar (for instance, same Social Economic Status) turned out to be impossible and it boiled down to the question of getting some data to analyse. Sampling could be seen as a non-probability version of stratified sampling (e.g., Coe, 2012). It is a simple and useful method for qualitative research (Palys & Atchison, 2014), although it was important to consider possible bias regarding generalisability since we did not know how representative the samples were.

The sample represents several public and independent schools from different parts of a large city. Even though their teachers and parents were positive about their participation, the students could opt out if they wanted to. The response rate might therefore provide additional information about the validity of the results. The teachers chose to participate, which could entail a self-selection bias, and if, as a group, they had a strong interest in the subject, this could also affect the sample. The prospective teachers volunteered freely. All prospective teachers were studying to teach school years 4–6 and had not yet taken the course where statistics was included. This was a deliberate choice since I was to be one of the teachers of that course and I did not want my results to be influenced by their having done the course; I also wanted an understanding of what prospective teachers might bring to the course. Particularly interesting was that the prospective teachers were of different ages and represented a range of geographical locations. All respondents are treated anonymously. Even though there was no lecturer–student relationship, having students from the same university as the researcher could affect validity. The participation is tightly connected to the willingness of the respondents (e.g., Swedish Research Council, 2017). Researchers who may wish to replicate this study should note that I needed to use prompts to generate my data. The response rate was 63 percent (29 out of 46).

Methods

Using questionnaires in a first exploratory stage allows one to start collecting information (Tymms, 2012). In this study, questionnaires for papers 2 and 3 were developed using the results from Paper 1, where open questions were posed to offer possibilities for comparing individuals' descriptions, definitions, and understandings of conceptions (e.g., Coe, 2012). The 'how', 'explain' and 'years 4–6' elements of the questionnaire was chosen to frame the study's research context. The aim was to capture the respondents' conceptualisations of mean, median, and mode. As the results mainly reflected knowledge elements, the instrument was altered, and new items were added to capture more of the dispositional elements. Items about measures which were easiest/hardest to explain and most or least useful were added. The added questions aimed to capture expressed conceptions about confidence/self-efficacy. The focus on usefulness was expected to generate answers that expressed the mathematical side of the concept as well as context knowledge and conceptions. In this way, each item was constructed to sample different aspects of statistical literacy (e.g., Coe, 2012).

The questionnaires were also adapted for the two different groups of respondents. For teachers and prospective teachers, the questionnaire was given a teaching framing, while students got a student framing. This added to the difficulty of comparing the different results. Bias can be discussed from the perspective of representativeness and wording (Coe, 2012), and, depending on the sample, representativeness might affect validity. Wording is about how questions are interpreted. Major limitations with this instrument are that data cannot be developed with further questioning and respondents have no opportunity to query a question they do not understand.

4.2.2 Textbook analysis

Sampling

Textbooks are significant in mathematics education, since they provide teaching with content based on a syllabus (McCulloch, 2012). It has been found that examining textbooks contributes effectively to other methods of examining education (McCulloch, 2012). The sample of textbooks was based on convenience, commonality, and popularity. Several textbook series were chosen to increase representativeness (Silver, 2017; Son & Diletti, 2017), to capture the Swedish textbook signature regarding tasks about mean, median, and mode (e.g., Charalambous et al., 2010), and to make it possible to interpret the results on a general level.

Methods

The textbooks were examined by analysing afforded knowledge, since textbooks are regarded as being the source of the mathematical knowledge to be taught (e.g., Fan et al., 2013; Rezat & Straeßer, 2015). Textbooks generate second-hand data, which prospective teachers, students, and teachers, at some point, have all experienced in school. The first challenge was to find an appropriate definition of ‘task’. Here, a task was defined as the smallest division in a task/activity (e.g., Jones & Jacobbe, 2014). Most of the tasks were found within special chapters about averages, but some tasks were found in other chapters as well as in sections for homework. The various definitions of the three concepts could be found in various parts of chapters. In some books, they were found in the beginning of a chapter; in others, in the chapter where the concept was introduced; and in some books, in a summary at the end of a chapter. One challenge I encountered, was that in some books, there were definitions in more than one place, and in some series the definition was written differently in different places. Also, some textbook series only deal with a certain average in one school year, while others deal with it over several school years.

Depending on the underlying theories and research questions, data about tasks can be generated in different ways. Paper 4 examined the distribution among tasks about mode on a qualitative or quantitative level, and whether they were procedural or conceptual, while Paper 5 examined all three measures from the perspective of mathematical properties for input objects, transformations, and output objects, which is a more mathematical analysis than an analysis based on theories about procedural or conceptual knowledge. Paper 4 also examined wordings in definitions of mode. Although the textbooks afford a variety of possible content about the measures, my aim with this study did not extend to comparing the content between textbooks.

4.3 Data analyses

Given the different methods of data collection, different methods of analysis were chosen. The data in papers 1 and 2 were analysed inductively to explore and compare and build an understanding of statistical literacy. These two papers lay the theoretical and analytical foundation for Paper 3, where a deductive analysis was performed. Data for papers 4 and 5 were also deductively analysed.

The unit of observation is connected with the data collection method; what you actually observe in relation to the unit of analysis is what you will arrive at as a result (Sheppard, 2020). The unit of analysis is involved in the coding process, or in the development of the codes (Milne & Adler, 1999). In a mixed-method and multi-dimensional approach such as this, questions are framed to address various aspects of statistical literacy. Therefore, different units of analysis are chosen to be comparable with different units of observation, all affiliated to statistical literacy. Methods for both qualitative and quantitative analysis are used to provide opportunities for telling a story about statistical literacy, including comparisons of data. I used different methods to analyse the data descriptively: constant comparison, thematic analysis, and content analysis. In some papers, I have added some statistical analysis, and since I had nonparametric data, nonparametric methods are used. For smaller datasets, Fisher's Exact Test was used, and for larger datasets, the $\chi 2$ test was used (e.g., Gibbons & Chakraborti, 2014).

Table 3 provides a compilation of the units of analysis and observation, and methods of analysis for each paper.

Table 3

Summary of unit of analysis, of observation, and method of analysis

Paper	Unit of analysis	Unit of observation	Method of analysis
1	Group level: Prospective teachers	Answers to a questionnaire	Inductive: Constant comparative method
2	Group level: Prospective teachers and teachers	Answers to a questionnaire	Inductive: Thematic analysis Fisher's Exact Test
3	Group level: Students	Answers to a questionnaire	Deductive: Thematic analysis Chi-squared test
4	Documents: Textbooks	Definitions Tasks	Deductive: Thematic analysis Fisher's Exact Test
5	Documents: Textbooks	Tasks	Deductive: Content analysis Chi-squared test

In the thematic analysis and the content analysis, the data were quantified using coding guides.

4.3.1 Questionnaires

Qualitative analysis

A common purpose for papers 1 and 2 was to inductively examine possible themes to use for the deductive analysis in Paper 3.

The inductive analysis in Paper 1 was based on constant comparison (e.g., Glaser & Strauss, 1967). The purpose was to explore data through repeated coding (e.g., Waring, 2012). After coding all texts, comparisons were made by looking for the most frequently occurring codes (e.g., Thornberg, 2012). Three themes were generated. Alongside the selective coding, the codes were compared to basic skills for statistical literacy (e.g., Watson, 1997) and a definition of conception (e.g., Philipp, 2007).

In Paper 2, I used inductive thematic analysis. Thematic analysis is used to identify the meaning of the data (Braun & Clarke, 2006). The procedure is described in steps: first, it is the labelling of data in the codes, then looking for patterns and organising codes, and through these codes, developing themes, all with the intention of building a thematic framework. Also, frequencies of the themes are noted to allow quantitative analysis, since numbers help to show how strong themes are. Themes are seen as coherent patterns, with clear distinctions between themes (e.g., Braun & Clarke, 2006). Another argument for using thematic analysis was its theoretical freedom and the potential for providing rich and detailed data (e.g., Braun & Clarke, 2006). This was helpful, since one purpose was to explore whether the altered questionnaire would generate data about the dispositional elements, and whether the themes would also allow a comparison between prospective teachers' and teachers' conceptions.

In Paper 3, a deductive thematic analysis was made using Paper 1 and Paper 2, since the themes showed coherence to both knowledge elements and dispositional elements of statistical literacy. Therefore, the starting point in Paper 3 was the elements of statistical literacy chosen for this thesis (see Figure 3). A subcategory, homonyms, was added during analysis, making analysis of the category 'Use of words' abductive (Mason, 2018).

Quantitative analysis

Fisher's Exact Test (FET) was used in Paper 2, using a contingency table for each question. The test allows for the examining of two or more variables simultaneously (e.g., Keselman & Lix, 2012). FET is useful for small samples and provides exact p-values indicating the significance of the deviation from a null hypothesis. The p-value equals the sum of all probabilities from all tables having the same sums of row and column and with the same or smaller probability (Raymond & Rousset, 1995). Significant level is set to 5 percent in both papers. The null hypothesis indicates that there is no difference between the groups of respondents. However, if $p < .05$ the null hypothesis is rejected and the fact that the groups are different is statistically significant. Such a result indicates that the probability of getting this result is less than 5 percent compared to the total number of possible outcomes from the contingency table. The differences are then due to reasons other than chance. These results should be interpreted with care given the different samplings, something that will be discussed later in Section 4.4.

In Paper 3, with a more extensive sample, a homogeneity chi-squared test was used. Just as for FET, it was used to determine the significance of differences between distributions (e.g., Alm & Britton, 2008). Paper

3 results are also presented from the perspective of frequency and relative frequency. The purpose is to investigate the distribution of the themes in relation to mean, median, and mode.

4.3.2 Textbook analysis

Qualitative analysis

In Paper 4, a deductive thematic analysis was made (e.g., Braun & Clarke, 2006). This is an explicit analyst-driven approach, drawn from the research questions with an analytical focus on the tasks related to the mode. The starting point for the themes in Paper 4 (i.e., analysing tasks about the mode), was the different levels of measurement (i.e., qualitative [Ql] or quantitative [Qn]), and cognitive demands (i.e., procedural [P] or conceptual [C]). The definitions provided by the textbooks were coded for vocabulary (inductively) and levels of measurement. Qualitative and quantitative data were examined because of the mathematical property that mode exists for all levels of measure. My reason for examining for procedural or conceptual was that I wanted to be able to distinguish between tasks that had mathematical properties placed in the context of the task, and those that did not. In the final step, I analysed the textbooks' definitions of mode with respect to use of words, which was an inductive analysis, the aim of which was to examine how the different definitions captured different aspects concerning level of measures and mathematical properties, and what language was used.

My reason for performing a deductive content analysis (CA) in Paper 5 was so that I could have coding units not open for interpretation, while the method allows statistical analysis. In a CA, important factors in the research questions are identified and reflected in the data (White & Marsh, 2006). The starting point for the themes in Paper 5 was input objects, transformations, and output objects (e.g., Mayring, 2015; Schwarz, 2015). The aim was to draw inferences from the tasks to afford opportunities to learn mean, median, and mode. Here, data sets, most often presented as input objects, make levels of measure interesting because the level defines possible mathematical properties for the variable. The different mathematical properties show the distribution of the properties textbooks hold important to learn. Here, since the definition of contextual task requires that the task deals with mathematical properties in some way (as input object, transformation, or output object), means that we also have some information about how these tasks are designed.

Quantitative analysis

In Paper 4, the aim was to examine the distributions of procedural or conceptual tasks and quantitative or qualitative values. The choice was to use Fisher's Exact Test (FET) with a small sample. In Paper 5, the aim was to examine the distribution of contextual and non-contextual tasks between mean, median, and mode, as well as the distribution of levels of measure in the input objects between the three measures. The choice of statistical test was chi-squared since it allowed me to analyse a more extensive sample. Significant level was set to 5 percent for both papers. In paper 5, the tasks were also analysed from the perspective of transformations, and mathematical properties connected with the calculated measures. All results in both papers are also presented from the perspective of frequency and relative frequency.

4.4 Methodological reflections

> Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise (Tukey, 1962, p. 13–14).

For a chosen methodology to be valid, the study must measure and explain what is claimed (Djurfeldt et al., 2003; Mason, 2018). Since there are several notions and definitions of validity in content analysis (see Fitzpatrick, 1983), specific terms for validity will not be used. Instead, validity will be discussed in relation to different steps within the methods: as validity of data-generating methods and as validity of interpretation (e.g., Mason, 2018). Another way to discuss the validity and reliability of qualitative research is to use the concept trustworthiness, which consists of four criteria: credibility, transferability, dependability, and confirmability (Bryman, 2018; Cloutier & Ravasi, 2021). Credibility and dependability relate to how research is performed, transferability and confirmability to interpretation (Nowell et al., 2017). Demonstrating reliability is about accuracy in how we measure and replicability, hence how accurate the claims are (Djurfeldt et al., 2003; Mason, 2018). Finally ethical considerations are discussed.

4.4.1 Validity of data-generation methods

Generally speaking, using mixed methods results in the phenomenon, statistical literacy, being explored from different angles. Exploring more than one dimension of the phenomenon enhances validity since different methods entail different distinctiveness (Mason, 2006, 2018). To enhance analytical validity, samples are chosen where the phenomena is located (e.g., Mason, 2018)—in this study, students, prospective teachers, teachers, and textbooks. Operationalising concepts to concrete measurable indicators is important for avoiding systematic errors (Djurfeldt et al., 2003), so to avoid errors, two different instruments were used, namely, questionnaires and textbook analysis. The questionnaires were used to generate data from the groups involved, as they were expected to have had experience with mean, median, and mode. The textbook analysis was used to generate data on what is presented to teachers and students about the measures being the link between intended and implemented curriculum (e.g., Okeeffe, 2013).

One limitation with regard to validity is a lack of consensus on the theoretical definition (Nolan et al., 2012; Mason, 2018). Because of the confusion around dispositional elements, the broad notion of conceptions was used to avoid possible limitations with operationalising the concept. As mentioned earlier, this thesis is rigorous about mathematical knowledge, using mathematical properties as a framework, while conceptions, context knowledge, and use of words are discussed as broad concepts. This was made clear before developing the questionnaire items. The first analysis, in Paper 1, revealed little about conceptions, so items were added successively, generating data about conceptions for papers 2 and 3.

Validity is also about ensuring that the unit of analysis is on the correct level of data (DeCarlo, 2018). As a researcher, I should not make inferences on levels other than that of the unit of analysis because the unit of analysis indicates on which level data is analysed (Chu, 2017). Papers 1–3 are on a group level which is why a questionnaire is a suitable instrument (e.g., Tymms, 2012). For papers 4 and 5, textbook analysis was used. The unit of analysis was textbooks, since tasks of interest were to be found in different chapters and in different books. To avoid ecological fallacy and reductionism, claims in papers 1–3 are on a group level and for Paper 4–5 textbooks. To clarify, claims in Paper 4–5 do not relate to specific textbook series.

As indicated earlier, sampling might be biased regarding generalisability, and because of the size of the sample (e.g., Gal, 2002). Here, we have a main limitation with the sample of teachers and prospective teachers. The sample of the students provided another challenge: the difficulty of establishing the response rate. Low response rate in a questionnaire can threaten validity (e.g., Tymms, 2012), meaning that results should be interpreted with caution. However, we are not interested in generalising back to the entire population (e.g., Palys & Atchison, 2014). Instead, the different samples ensure that suitable groups are measured, thus providing meaningful data (e.g., Tymms, 2012). Having suitable groups and using mixed methods (including repeating methods), can enhance validity (Mason, 2006). Mixed methods can ensure validity when operationalising concepts, making it possible to analyse from the perspective of different units. Repeating a method can ensure validity, for instance, by making it possible to adapt items.

One bias when developing an instrument lies in the choice of words (Tymms, 2012). Online questionnaires do not allow respondents to comment on items. Respondents might interpret questions differently, resulting in responses that do not address the question. The decision was therefore to code such responses as ‘not relevant’, although this has the potential to increase limitations in the results when the category ‘not relevant’ is significantly large.

Despite the limitations, however, this study can be regarded as trustworthy regarding both the unit of analysis and unit of observation, which ensures validity for data generation (e.g., Chu, 2017).

4.4.2 Validity of interpretation

In the papers, there were various attempts to theoretically frame statistical literacy. Nonetheless, despite the different framings, all results could be interpreted within Gal’s (2004) framework (see Figure 2). Table 4 shows a compilation of the themes across papers 1–5.

Table 4

A compilation of themes within Gal's (2004) framework in relation to each paper

Elements of statistical literacy (Gal, 2004)	Paper 1	Paper 2	Paper 3
Literacy skills	Colloquial concepts		Use of words
Statistical and Mathematical knowledge	Procedural & Conceptual knowledge Statistics (mathematics) Didactics (teaching)	Mathematical Pedagogical	Mathematical knowledge Mathematical properties
Context knowledge	Context Usefulness	General	Context knowledge
Critical Questions			
Beliefs and Attitude	Usefulness Didactics (teaching)	Personal Pedagogical	Conceptions
Critical stance	Usefulness	Personal Not Relevant	Conceptions Not Relevant Do not know

Elements of statistical literacy (Gal, 2004)	Paper 4	Paper 5	
Literacy skills	Vocabulary in definitions		
Statistical and Mathematical knowledge	Qualitative Quantitative Procedural Conceptual Mathematical properties	Level of measure Transformation Mathematical properties	

I will now discuss possible interpretations of each paper.

Interpretation—Questionnaires

As described earlier, Paper 1 used an inductive approach using constant comparison where the idea was to interfere data with no other intention than to see what happened in the data (e.g., Thornberg, 2012). The results from initial coding were compared to the framework for statistical literacy presented by Watson (1997), in combination with the definitions of procedural and conceptual knowledge suggested by Hiebert and Lefevre, (1986), and the broad definition of conceptions proposed by Philipp (2007). This process, of constantly comparing newly collected data with existing theories ensures, according to Coe (2012), high validity in the interpretation of the results, which can be further increased when the process of constant comparison is performed together with another researcher, an approach used in Paper 1.

Paper 2 used an inductive approach using Thematic Analysis (TA), which resulted in five different themes. To ensure validity, one must be as transparent as possible when describing codes. I have tried to do this by providing examples, which helps the researcher, or other researchers, to identify how themes are compatible with the dataset, and reflects the meaning of the data as a whole (e.g., Alhojailan, 2012; Braun & Clarke, 2006). One example is provided in Table 5, showing a response coded as Pedagogical and Mathematical:

Table 5

Response to why mode is easiest to explain [PT5, about mode]

Personal	Pedagogical	Mathematical	General	Others
	A practical exercise can be to cut out numbers and then place/organise the different numbers	Calculate which number there are most of		

Another aspect of validity is validating themes both early and late in the process (Alhojailan, 2012; Braun & Clarke, 2006). In Paper 2, the second author validated the themes, both early in the process and when all data had been analysed.

The third paper used a deductive approach, namely, Thematic Analysis (TA). Each part of every response was coded using the established framework. The themes were elaborated and adapted to a framework of statistical literacy (see Figure 3), with a starting point being the themes generated in papers 1 and 2. This ensures validity in two ways. First, previous findings signal that the themes are reasonable. Second, words/phrases that are coded within a theme, for instance, mathematical knowledge, need to coincide with the theoretical underpinnings (see Gal, 2004; Watson & Callingham, 2003). In retrospect, the frameworks used in papers 1 and 2 appear to have captured the same epistemological phenomena as those used in Paper 3. Mason (2018) writes that some researchers struggle to admit that different framings may (or may not) provide similar results. In addition, we know that "all models are wrong, but some models are useful" (Box & Luceño, 1997, p. 61).

The one main strength of adapting and refining Gal's (2004) framework is that it allows me to compare my results with previous research. This model then becomes useful. It also makes it easier for other researchers interested in statistical literacy to engage with my results since it does not require the same theoretical puzzle as in the first attempts in papers 1 and 2.

Interpretation—Textbook analysis

The textbook analysis in papers 4 and 5 is research on actual textbooks, which enhances the presence of realistic data (e.g., Rezat & Straesser, 2015; Weninger, 2018). This means that the interpretation of data comprises inferences from texts, making them both replicable and valid (Rezat & Straesser, 2015). To increase the validity, I identified key concepts and compiled them into a coding guide followed by anchor examples (e.g., Mayring, 2015; Potter & Levine-Donnerstein, 1999). These are common techniques when analysing tasks.

4.4.3 Reliability

Reliability can be ensured in different ways; it can be measured by repeating a method for the same kind of data (Mason, 2018), which confirms reliability of the instrument (Gal, 2002). For me, reliability was an aspect of concern when processing the questionnaire and the themes for analysis in papers 1, 2, and 3. The instrument showed consistency for all items when comparing papers 1 and 2, to Paper 3. Adding both Content and Thematic analysis, both regarded as rigorous methodologies, ensures high reliability (Nowell et al., 2017; Weninger, 2018).

Another way to ensure reliability is the use of intercoder reliability, where independent coders evaluate the characteristics of codes (O'Connor & Joffe, 2020). This process was used for Paper 5. One problem with this process was the interraters' familiarity with the mathematical content-related codes. Here, this turned out to be crucial for the result, despite a coding scheme with explicit anchor examples. However, interraters' who were familiar with the mathematical content also could evaluate the codes in a proper way. The coding scheme is part of the content in Paper 5 to inform readers about the background to the interpretations. I also included supporting texts with visual evidence, like tables, to ensure trustworthiness and to make the research more reliable (e.g., Cloutier & Ravasi, 2021).

In all papers, the results of the analysis are reported both in text and in contingency tables, with statistical analysis at least on a frequency level. The statistical aspect helps to describe the thickness of the codes/themes. Results from papers 2 and 3 are also exemplified with excerpts, providing further understanding of the data (Cloutier & Ravasi, 2021). This is to increase the transparency of the results (e.g., Tracy, 2010).

4.4.4 Ethical considerations

Ethical considerations must be made continuously throughout the research process (Mason, 2018). While it is about protecting participants, it also involves the ethics of the various methods (Swedish Research Council, 2017; Tracy, 2010). Below, I discuss ethical considerations related to the studies conducted as part of this thesis.

Questionnaires

Asking individuals to answer questions to be analysed from the perspective of mathematical knowledge may be seen as sensitive: it is important that the participants feel both protected and willing to participate. The questionnaires started with information about the study, including rights to anonymity (e.g., Swedish Research Council, 2017). Collecting the data through an online questionnaire created a distance between the researcher and the respondents. The respondents were assured that the dates of the studies would not be mentioned and neither their names nor the names of their schools would be requested. While their responses were to constitute the data, no personal information would be shared and thus participants could not be traced. The data are stored securely according to Stockholm University rules and only I have access to the data.

The prospective teachers in Paper 1 were about to start a course where I was one of the teachers and were thus in an exposed position. However, the questionnaire was administered before they met me in class; they were assured that their responses would remain anonymous, and that participating would not in any way affect them in the course. For papers 2 and 3, the situation was different. The teachers were anonymous and were not connected with me or my teaching in any way. They agreed to participate using a web link to the online questionnaire. Prospective teachers in Paper 2 were not connected with me in any way and the participating students were anonymous. Since it was hard to make contact with teachers who would volunteer to let students participate, some of these teachers were familiar to me, but none of the students were in any way connected with me. There is no information about where they come from, and the only information given about sampling is that they come from different types of schools and from different areas surrounding a big city. Caregivers provided written consent for students who participated. I do acknowledge, however, that having a caregiver and a teacher approve their participation might have affected the students' attitude to such participation.

Textbook analysis

There are fewer ethical issues with conducting a textbook analysis. However, while you do not need permission to use existing texts in the public domain, it is still important to heed ethical considerations to substantiate the methodology and assure its quality (Mason, 2018). The sample was purposeful, but also based on convenience—the textbooks were available, or easy to get from the publishers. Here, where a publisher sent textbooks it is interpreted to mean that they consented to their analysis. Some publishers did not respond to the request. Although the sample included seven textbook series, this does not cover all available textbooks. That said, the most common are among these seven. The intention is not to evaluate the textbook series but rather to provide a picture of what opportunities to learn about the measures of central tendency the Swedish textbooks afford. The results have therefore been interpreted with 'textbooks' as the unit of analysis. I do not name the specific textbooks.

5 Results

The results are presented sequentially in two clusters: the first cluster is the papers about prospective teachers' and teachers' conceptions and students' expressed statistical literacy (papers 1–3), while the second cluster is the two papers that analyse textbooks (papers 4 and 5).

5.1 Paper 1—prospective teachers' conceptions of the concepts mean, median, and mode

The aim of the exploratory study in Paper 1 (Landtblom, 2018a) was to study prospective teachers' conceptions of mean, median, and mode in relation to explaining these concepts to students in years 4–6. The research question was: How do prospective teachers conceptualise the concepts of mean, median, and mode to a student in years 4–6? Data was collected through questionnaires with open questions in which the prospective teachers were to describe how they would explain the measures to a student.

The prospective teachers' answers were inductively analysed and generated seven codes: conceptual knowledge, procedural knowledge, context, colloquial concepts, usefulness, statistics (mathematics), and didactics (teaching). These codes were grouped into three tentative categories: Use of words, Understanding averages, and Teaching explanation.

In the responses, mean seemed to be the measure most familiar to the prospective teachers, as seen in both the categories Use of words and Understanding averages. The respondents used, besides definitions and numerical examples, more context in their explanations as well as examples of usefulness compared to the other averages. However, many colloquial words were used instead of the Swedish word 'medelvärde' (mean). When looking at median, the responses focused strongly on use of number lines to find what was in the middle, which was categorised as Teaching explanation. Mode was found to be the measure least familiar to the prospective teachers, since many could not give examples or responded that they did not know what it was. Another sign of unfamiliarity, categorised as Use of words, showed that the respondents emphasised the last part of the word, ‘värde’ (value) in the Swedish term ‘typvärde’ (mode). This was demonstrated in explanations where the respondents used numerical values.

The overarching result was that a high proportion of the prospective teachers demonstrated a profound procedural knowledge, but weaker conceptual knowledge, implying that teacher education must meet the need to enhance their conceptual knowledge in order to develop their statistical literacy. Even though the respondents had some connection to me, rendering the results somewhat limited in terms of generalisability, the results are nonetheless strongly coherent with earlier research.

5.2 Paper 2—teachers and prospective teachers’ conceptions about averages

In Paper 2 (Landtblom & Sumpter, 2021), the aim was to explore prospective teachers’ and teachers’ conceptions about averages. The research questions were: (1) What are the characteristics of the motivations given by prospective teachers and teachers to which of the averages is the easiest or hardest to explain? (2) What are their expressed conceptions about the usefulness of the averages? (3) How do prospective teachers and teachers differ in their responses?

Data were collected through questionnaires with open questions, given to teachers and prospective teachers who volunteered to participate. While the response rate was low at 35 percent, rich responses were generated, which was appropriate for the qualitative analysis. The first step in the inductive analysis was to identify signs of common conceptions across all questions. Five themes were generated: Personal, Pedagogical, Mathematical, General, and Not relevant. Responses categorised as Personal typically concerned their own knowledge, Pedagogical how to teach the averages, Mathematical definitions and calculations, and General daily use in society.

The result concerning how easy or hard a measure was considered to explain demonstrated that the teachers found mode to be the easiest to explain and mean and median the hardest. The prospective teachers, on the other hand, found mode hardest to explain. Furthermore, the motivations in the two respondent groups differed. While the teachers predominately based their motivations on pedagogical explanations, the prospective teachers based theirs on mathematical experiences, often procedures.

As concerns responses to the question about the usability of averages, both prospective teachers and teachers found mean to be most useful and median or mode to be least useful. Many motivations for most useful, for both groups, related to real situations and both groups provided real-life examples. Only teachers gave pedagogical motivations. For least useful, the motivations from both groups were mainly general or not relevant, and no real-life examples were given for mode.

The results indicated differences between the groups: the prospective teachers' motivations concerned their own knowledge while teachers predominantly gave pedagogical motivations, arguing from the perspective of teaching experience. The respondent groups' varied experiences can be seen as an explanation of the result. This difference was statistically analysed with Fisher's Exact Test, showing a significance. However, because of the sample, these results should only be regarded as indications.

5.3 Paper 3—which measure is most useful? Grade 6 students' expressed statistical literacy.

In Paper 3 (Landtblom & Sumpter, submitted), the aim was to study Grade 6 students' expressed statistical literacy about measures of central tendency. The research questions posed were: (1) What different knowledge elements do the students use when explaining measures of central tendency, and how are these knowledge elements expressed? (2) Which measure is, according to grade 6 students, easiest or hardest to explain, and what characterises their motivations? (3) Which measure is considered, according to grade 6 students, most or least useful, and what characterises their motivations?

The participating students in this study were contacted through their teachers who had volunteered to participate in the study. Data were collected through a questionnaire consisting of open questions. In this deductive study, the adapted framework of statistical literacy (see Figure 3) was used to categorise the data. The knowledge elements are Mathematical knowledge, Use of words, and Context knowledge and the dispositional element Conceptions.

Results for question 1 show that the most frequent knowledge elements used are Mathematical knowledge and Use of words. For all measures, Mathematical knowledge was mainly expressed using algorithms. However, many chose the wrong algorithm for the mean. A large proportion gave correct but incomplete algorithms for median and mode. Colloquial words were used mainly for mean, while descriptive words were used for median and mode. One interesting result for mode was that the Swedish homonym ‘typ’ in ‘typvärde’ (mode) meaning ‘approximately’ was used to express the meaning of mode.

The results for question 2, concerning how easy or hard the measures are to explain, show that many of the students found mode the easiest measure to explain and median the hardest. The characteristics of the motivations for easiest to explain fell mainly in the category Conceptions. Mathematical knowledge and Use of Words were also present in the data. The characteristics of the motivations for hardest to explain were mainly in the categories Conceptions, Mathematical knowledge, and Do not know. The results show that students think the algorithm for mode is easy, and median is hard to explain, including an expressed insecurity about the meaning of median.

The results for question 3, usefulness of the measures, show that the mean is thought of as the most useful, and median the least. Context knowledge was the most prominent category of characteristics of motivations for usefulness. The concepts mean and mode are considered most useful, which is shown with examples of suitable contexts. The median however was considered least useful. Students state it is a concept that only exist in a school context.

For the statistical analysis, a chi-squared test was used. The results for the first question indicated differences in the distribution of the knowledge elements between mean, median, and mode. The results for question 2 indicated differences in the distribution of the six categories: Mathematical knowledge, Context knowledge, Use of words, Conceptions, Do not know, and Not relevant, between the measures. The results for both easiest and hardest to explain were significant.

5.4 Paper 4—is data a quantitative thing? An analysis of the concept of the mode in textbook tasks for grades 4–6

In Paper 4 (Landtblom, 2018b), the aim was to investigate definitions of and tasks related to the concept of mode in Swedish textbooks. The research question posed was: What knowledge, procedural or conceptual, and what quantitative or qualitative context, do textbooks for years 4–6 afford Swedish students on the concept of mode? In this study, tasks from seven Swedish textbook series for school years 4–6 were analysed, considering whether data in tasks were qualitative or quantitative, and whether they provided procedural or conceptual knowledge. Conceptual knowledge was identified through mathematical properties. That is, that there might not be any mode or more than one mode and that it is applicable for nominal values. The analysis also focused on the wording of the definitions of mode presented in the textbooks.

The most prominent results concerned procedural knowledge and quantitative data. A high proportion of tasks afforded procedural knowledge (82%). The remaining tasks, affording conceptual knowledge, contained the mathematical property that data sets can have more than one mode. The results for quantitative and qualitative data showed that 75 percent of the contexts in the tasks was quantitative. In the tasks with qualitative data, colloquial phrases like ‘more popular’ or ‘the most’ was used in the question instead of asking explicitly for the mode. Only two percent of the tasks thus afford conceptual knowledge in a qualitative context.

In relation to the wording of the definitions, the results revealed a use of words related to quantitative values, of which one was the word number ‘tal’ instead of value in the definition. Only one textbook series exemplified with both qualitative and quantitative values. The distribution of tasks was tested by Fisher’s Exact Test and found statistically significant.

5.5 Paper 5—afforded opportunities to learn mean, median, and mode in textbook tasks

In Paper 5 (Landtblom, submitted), the aim was to draw inferences from the text in the tasks as to their effects, and the opportunities to learn they afford. The research questions posed were: (1) What is the distribution between non-contextual and contextual tasks? (2) What opportunities to learn about a) input objects, b) transformations, and c) output objects do textbook tasks afford, and what does the distribution look like? In this study, tasks from seven Swedish textbook series for school years 4–6 were analysed. In the analysis, it was not only the formulation of the task that was the focus of the analysis. I also looked at possible solutions and the output object (i.e., the answer). Input objects were analysed using the level of measure (i.e., nominal, ordinal, and quantitative). Transformations were analysed using the categories procedural (transformation close to the definition) or conceptual (for instance, reverse calculation). Output objects were analysed with respect to the mathematical properties.

The results of the analysis showed that, concerning the distribution of contextual and non-contextual tasks, 61 percent of the tasks were contextual, affording explicit or implicit affordances of mathematical properties. There was an uneven distribution among tasks about mean, median, and mode. Tasks about mean (62% of all tasks), afforded more contextual than non-contextual tasks. Tasks about median (19% of all tasks), afforded about the same number of each kind. Tasks about mode (18% of all tasks), afforded more non-contextual than contextual tasks. The distributions were tested using a homogeneity chi-squared test and found to be significant. These results signal that there are fewer mathematical properties among the tasks about mode.

The results for question 2a, about input objects, are also statistically significant according to a homogeneity chi-squared test. The textbooks afforded opportunities to learn about all three measures in quantitative contexts (92%). The result concerning transformations shows a high proportion of procedural transformations close to the definitions. For all tasks about mean, 67 percent were of this kind, for median 77 percent, and for mode 81 percent. Thus, there were more other transformations than procedural for the mean. Most frequent of these transformations were reverse calculation (22%). In the results for question 2c, output objects, the most frequent mathematical property in all tasks (48%) was when an average does not equal values from the dataset. The second most frequent mathematical property (20% of all tasks) was to consider zero as a value in the data set. The third most frequent mathematical property (10% of all tasks) was that an average can correspond to more than one dataset.

6. Discussion

In this section, I discuss the five papers. First, the results from papers 1–5 will be discussed using the selected aspects of statistical literacy, knowledge elements, and dispositional elements. Then, the results from papers 1–3 are compared with the results from papers 4 and 5. In the last section, the results are discussed in relation to the two overarching research questions: a) What aspects of statistical literacy do students, teachers, and prospective teachers express about mean, median, and mode? b) What do textbook tasks afford with respect to different aspects of statistical literacy regarding mean, median, and mode?

6.1 Mathematical knowledge

Previous research often points out that demonstrated knowledge is primarily procedural (Clark et al., 2007; Garfield & Ben-Zvi, 2007; Groth & Bergner, 2006; Hannigan et al., 2013; Jacobbe & Carvalho, 2011; Jacobbe, 2012; Leavy 2010; Leavy & O'Loughlin, 2006; Peters et al., 2014; Russell & Mokros, 1991; Sorto, 2004). Procedural knowledge is often thought of as consisting of rules or algorithms (Hiebert & Lefevre, 1986). The findings from my studies in papers 1–3, coincide with previous research, showing that both students, prospective teachers, and teachers primarily express procedural knowledge. Since procedural knowledge can be both productive and unproductive, the discussion will continue below, first on a general level and then, more specifically, for each of the different measures.

6.1.1 Procedural knowledge

Procedural knowledge can be connected with conceptual knowledge, and together create stable ground for further understanding of mathematical concepts (e.g., Bakker, 2004; Groth, 2014). Facts, rules, and calculations, however, are not enough for understanding what a measure represents (Gal, 2004; Mokros & Russell, 1995). However, without conceptual knowledge, procedural knowledge can also result in imitative reasoning where algorithms are used with little or no understanding of why or when (e.g., Gal, 2019: Lithner, 2008). The results from papers 1–3 show that respondents do not always remember the definition correctly or choose the correct definition. These are not new results—they have been reported before (e.g., Watson, 2007). As an example, in Paper 3, several students confused the algorithm for mean with the algorithm for median, and also struggled with the definitions for median and mode. Combining these results with previous research, using the theoretical underpinnings provided by Watson and Moritz (2000), the limited understanding of the different measures can affect their ability to structure the mathematical aspects (e.g., Büscher & Prediger, 2019). At the same time, one can only develop understanding of mathematical concepts based on what one is afforded (i.e., opportunity to learn) (e.g., Cogan et al., 2001; Tarr et al., 2006).

One way to connect procedural knowledge and conceptual knowledge is through the use of an intuitive analogy of mean as a balancing point or fair share (Leavy & O'Loughlin, 2006; Watier et al., 2011). The results from the textbook analysis in papers 4 and 5, show few such didactical presentations of the procedures in the tasks that were analysed. The vast majority of tasks only address procedural knowledge. The results are consistent with previous findings, both for mathematical content in general (e.g., Glasnovic Gracin, 2018) as well as for statistical content (e.g., Büscher, 2022a; Pickle, 2012). Here, in more than 70 percent of the tasks, the transformation is procedural, of which more than half are about the mean. The strong emphasis on computation of the mean is consistent with previous research (Cai & Moyer, 1995; Leavy & O'Loughlin, 2006; Watier et al., 2011). What is afforded to the students, their opportunity to learn, is therefore unbalanced. Since there is a body of research showing that there is a connection between afforded OTL and learning performance (e.g., Hadar, 2017; Kieran et al., 2015; Tarr et al., 2006; Törnroos, 2005; Watson & Thompson, 2015), the conclusion is that teachers need to make active choices if they want to provide more balanced teaching to their students. Some suggestions, based on the results from papers 4 and 5, are to use debug (e.g., Byrnes & Wasik, 1991) and reverse calculation (e.g., Groth & Bergner, 2006; Konold & Pollastek 2004), since they are transformations connecting the procedure with the statistical context, allowing the students to develop conceptual knowledge. The impact textbooks have, both on what is taught by the teacher (Glasnovic Gracin, 2018; Johansson, 2006; Jones & Pepin, 2016; Pepin & Haggarty, 2001; Remillard, 2005), and on what students learn (Cogan et al., 2001; Stein et al., 2007; Tarr et al., 2006), is well documented. The theoretical framework about afforded OTL allows us to identify whether tasks help to make sense of data beyond procedural knowledge (e.g., Büscher, 2022b; Huey & Jackson, 2015; Jones et al., 2015). However, if a more detailed analysis of the different measures is desired, then a more refined tool for describing mathematical concepts is needed. Hence, we move on to mathematical properties as part of procedural knowledge.

6.1.2 Mathematical properties

It is mainly in Paper 5 where the main contribution to the research field about mathematical properties in the concepts mean, median, and mode is made. Compared to Strauss and Bichler (1988), I differ with regard to the two main objects, input value and output value, and the different transformations to which the three concepts are connected. Another contribution is with respect to statistical literacy. As stated earlier, performing calculations is not equivalent to understanding statistics (Gal, 2000). In statistics, tasks without a context can therefore be interpreted as 'bare tasks' (e.g., Cobb & More, 1997; Mokros & Russell, 1995), but can still have intrinsic mathematical properties that function as an implicit context, as discussed in Paper 5. I have named these tasks 'implicit contextual tasks', and I build on the definition that non-contextual tasks do not deal with or contain any extra-mathematical elements (Reinke, 2019). This is something not discussed in earlier research, and I see this as a contribution to the field—a contribution to the lack of research on procedural knowledge identified by Groth (2014) and Star (2005).

6.2 Use of words

The important role of language as a precursor to understanding has been highlighted in previous research (Usiskin, 2012). Much research about the meaning of mean, median, and mode is connected with the English word 'average' (e.g., Watson, 2013), and the results from this thesis contribute to broadening this knowledge to include Swedish words.

My results show that both colloquial and descriptive words are used when explaining mean. Just as the English word 'average' has many colloquial connotations (e.g., Watson, 2013), the Swedish word for mean, 'medelvärde' (average value), has connotations. My results show that colloquial words are connected with the variable, like average age, or with the meaning of the word for mean, independent of the respondent group. On the other hand, the most common use of words for median were descriptive words describing the position of the median, like 'middle' or 'in between' (e.g., Kaplan et al., 2014; Makar & McPhee, 2009; Mokros & Russell, 1995; Watson, 2013). Given the definition of the median, one can see the reason why. However, when students use middle in a tautological way—saying that median is the middle—it does not provide any statistical meaning (e.g., Clark et al., 2007). Previous research confirms that this intuitive idea of median as the middle is not helpful if you do not consider middle in relation to the dataset (Schnell & Frischemeier, 2019). Despite 70 percent of the students knowing the definition, their inability to understand what the word median actually means could be one explanation for why so many of those participating in the Paper 3 study found median the hardest to explain.

Mode ('typvärde') as a mathematical concept does not have any colloquial connotations. Some students that never had heard the word 'typvärde', used a homonym based on the first syllable in the word 'typ', meaning 'about' or 'approximately'. The results relate to previous research which looked at everyday meanings of mathematical terms (e.g., Kaplan et al., 2009, 2010).

In papers 1–3, several respondents found the mathematical concept 'value' problematic. They expressed values as a quantitative entity. Their responses resonate with how most textbooks define the mathematical concept 'value'. When looking at previous research, it appears that little has been done with respect to value and how it is defined, both from a linguistic perspective and how it is used with respect to mathematical properties of the concept. I therefore suggest this as a possible area for further research.

Since students often use their colloquial understandings in Mathematics, intuitive meanings might be contradictory, and so students need opportunities to learn about how to express the meaning of measures through textbook tasks (Kaplan et al., 2009; Konold, 1995; Makar & McPhee, 2009; Russell & Mokros, 1991). Including both colloquial and mathematical concepts helps students grasp the correct meaning (Kaplan, et al., 2010; Russell & Mokros, 1991). Colloquial words also contribute positively to students' conceptions, as these make statistics easier to experience (Makar & Confrey, 2005). For teacher education, this implies a need to pay attention to knowledge about what words students use talking about statistical concepts. Prospective teachers need to develop knowledge about how students describe different measures, including value, and then especially median and mode that are not as familiar as mean.

6.3 Context knowledge

In this thesis, I follow the growing understanding of the importance of context in statistical knowledge (e.g., Büscher & Prediger, 2019; Gal, 2019; Weiland, 2019a, 2019b). Context knowledge includes the ability to provide examples of contexts where a measure is suitable, and this is seen as a different knowledge than that required to calculate the measure (Leavy & Hourigan, 2016; Leavy & O'Loughlin, 2006; Watson & Moritz, 2000). The results from papers 1–3 show that the respondents struggle to identify contexts where a measure is useful. The results are in line with previous studies (e.g., Richardson et al., 2013). For median, only one teacher of all the different respondents was able to provide a context. The results have several implications. First, one possible conclusion is that students disconnect the measures from their understanding of what a measure means since they spend a lot of time on procedures that involve calculating numbers without any contexts (e.g., Cobb & Moore, 1997; Mokros & Russell, 1995). Also, one can use a theoretical framework for mathematical reasoning (e.g., Lithner, 2008) to see if there are specific transformations involved in the process of identifying the task situation (e.g., Eriksson & Sumpter, 2021)—what is the task about with respect to the context (e.g., Gal, 2019)—and being able to evaluate the conclusions (e.g., Eriksson & Sumpter, 2021), also with respect to the context (e.g., Ben-Zvi & Garfield, 2004; Watson, 2013). As a teacher educator myself, I conclude that teacher education has room for improvement. Prospective teachers need to practise these specific kinds of knowledge needed to understand context knowledge (e.g., Leavy & Hourigan, 2016; Leavy & O'Loughlin, 2006; Watson & Moritz, 2000), which include the kind of reasoning suggested above.

6.4 Conceptions

Conceptions are an important aspect since they are, for instance, connected with engagement in mathematics (Carmichael et al., 2010). The choice of a broad view on conceptions, using the definition from Philipp (2007), allowed me to combine this broad definition with Dispositional elements from Gal (2002, 2004). Since Gal (2002, 2004) only focuses on beliefs, attitudes, and critical stance, other affective constructs such as mental images and preferences are excluded. In that sense, conceptions encompass more, so I could analyse a variety of conceptions. The results highlighted self-confidence as part of conceptions, especially connected with specific knowledge about measures (e.g., Beswick et al., 2012; Gal et al., 1997; Schoenfeld, 1992). The results are a contribution to the field, stressing that it is not just about knowing the different mathematical properties connected with the transformations and objects related to a concept, it is also about self-conception, that 'I am a person who knows these things'. Self-confidence is, according to Rumsey (2002), connected with different aspects of knowledge such as algorithms, language of statistics, and contexts. In this thesis, the respondents demonstrated more confidence about the mean in all the aspects raised by Rumsey (2002). The instrument used in papers 1–3 used several aspects of conceptions as defined by Philipp (2007), and the questions about what was considered easiest/hardest to explain generated lots of data. The results also showed that the category Use of words was tightly connected with conceptions. As an example, the respondents who knew a specific word that gave meaning to a measure also appeared to be confident answering questions about what was considered easiest to explain. Similar findings have been reported earlier (e.g., Makar & Confrey, 2005). There are several implications of the results regarding conceptions. First, since teaching measures involves affective aspects, teacher education then must address conceptions when teaching about measures. It also means that as a teacher educator, one must be aware that students' self-conceptions are an important aspect of their knowledge of measure.

6.5 Methodology

As stated earlier, this is a problem-driven thesis. The intellectual puzzle (Mason, 2018) has been different, depending on the different aims and research questions in the five papers. One can see this thesis as an ecological puzzle, where the aim is to study different aspects of statistical literacy, and how different categories of statistical literacy are interrelated. Therefore, different theories have been used, which is appropriate given that statistical literacy, as a theoretical concept, is a complex construction (Chance, 2002; Gal, 2004; Garfield, 2002; Rumsey, 2002). Different theories are needed to study the different aspects. It also means that the theories used in this thesis are not the only ones appropriate for studying statistical literacy: the theories need to be chosen with respect to the research questions, which is the core of problem-driven research (Schoenfeld, 1992). Also, each study must be argued for with respect to relevance so the reader can see that the chosen topic is meaningful (Tracy, 2010). In a problem-driven study, this is a key point (Schoenfeld, 1992) and I have therefore been careful to provide a variety of arguments for relevance in each of the studies.

The next step is the theoretical operationalisation. Mason (2018) writes that, as a researcher, one should "think openly and creatively about how you might investigate your intellectual puzzle" (Mason, 2018, p. 24). Here, openness and creativity are both present in the variety of theories and the different methods used. The limitations are that one might not achieve the same theoretical depth as one might have if only one theory had been applied. Also, it can be somewhat messy with several theories in action and thereby difficult for readers to follow your narrative. The results could also be interpreted as being superficial compared to what might have been achieved by focusing on one response group. For instance, the results from papers 1–3 might have been validated and expanded with qualitative interview studies, which could have increased the transparency of the results (e.g., Tracy, 2010).

Transparency is an important part of methods, both for generating data and methods of analysis (Tracy, 2010). I am aware that as a researcher, I am not bias free, but I have striven for rigour and sincerity. Papers 1, 2, and 4 have all been through a review process where those issues have been addressed. Also, I have aimed for thick descriptions, something that Mason (2018) recommends when conducting qualitative research, and which can increase credibility (Tracy, 2010).

Transparency is also connected with resonance, where two important aspects are generalisations and transferable findings (Tracy, 2010). With respect to the qualitative analysis, I have followed the steps suggested by Braun and Clarke (2006) when doing thematic analysis. For the content analysis, I had a second coder to establish intercoder reliability. This means that the coder scheme not only achieved a satisfying level, but also was transparent enough for another researcher to follow. This transparency then increases the probability that the findings are considered transferable (Tracy, 2010). With respect to the quantitative analysis, I chose appropriate statistical tests given the nature of the nonparametrical data. In hindsight, given the problem with the sample in Paper 2, it might have been enough to simply have conducted the thematic analysis and then presented descriptive statistics. However, the decision was to perform statistical analysis to show what was possible with the instrument, something the reviewers agreed on.

Looking at significant contribution (Tracy, 2010), I would like to raise two contributions besides the empirical results. First, I want to talk about a theoretical contribution. Although this has been problem-driven research (e.g., Schoenfeld, 1992), my research has contributed to theory about statistical literacy. The mathematical analysis, using the theoretical concepts from Lithner (2008), allowed me to explicitly separate different mathematical properties in different objects and transformations. The theoretical framing was far more detailed compared to Strauss and Bichler (1988) and Gal (2002). It was also far more detailed compared to theoretical framing that use the concepts procedural knowledge and conceptual knowledge (e.g., Hiebert & Lefevre, 1986). In that sense, the mathematical analysis that was conducted for Paper 5, could also be seen as a methodological contribution (e.g., Tracy, 2010).

Several dimensions can be focused on with respect to ethics (Mason, 2018; Tracy, 2010). I have, of course, followed the ethical rules provided by the Swedish Research Council (2017), and data has been managed and stored according to Stockholm University's data management plan. But ethics is also about how one views the respondents and data. Following Mason (2018), I see data as generated, not collected. Although I have used the word 'data collection', I still think that data is generated, and that different stimuli can generate different data. For instance, if I had used the instrument from papers 1–3 as an interview guide, the same questions could have generated different responses given the setup of the data collection process. I have tried to signal an awareness of the limitations related to having this ethical view. For instance, I use phrases such as "the respondents express" or "their responses indicate".

Lastly, meaningful coherence (Tracy, 2010). Again, problem-driven research must argue both for relevance and appropriate methods since the focus is on answering a problem, not developing a specific theory (Schoenfeld, 1992). I have also tried to create meaningful coherence by formulating two overarching research questions targeting statistical literacy to frame the empirical results from papers 1–5. The next section aims to answer these two questions.

6.6 Statistical literacy

In the beginning of this thesis, I raised two overarching questions. The first was: What aspects of statistical literacy do students, teachers, and prospective teachers express about mean, median, and mode? When combining the results from papers 1–3 and comparing these to the model of statistical literacy presented by Gal (2002, 2004), the overall conclusion is there is a big difference when talking about mean, median, or mode. The skewness in how much research there is about mean compared to the other two (e.g., Groth & Bergner, 2006, 2013; Watson, 2013) reflects to some degree the expressed statistical literacy articulated by the different response groups. Some examples of this, illustrated in the three different studies, were that the concept mean often had more productive conceptions connected with use of words and context knowledge. Although respondents found the procedure (i.e., the transformation) a bit more difficult than the other measures, they still regarded it as more useable than the other two. This can be contrasted with median—that respondents found the transformation easy to explain but struggled to connect the concept with a meaningful context.

When researchers discuss the need for developing statistical literacy as an overarching goal for teaching and learning statistics (Ben-Zvi & Garfield, 2004; Franklin et al., 2007; Franklin & Kader, 2010; Hannigan et al., 2013; Shaughnessy, 2007), the results from the present thesis highlight the need to deal with numbers with a context (e.g., Cobb & Moore, 1997). The implication of this, already raised by Ben-Zvi and Garfield (2004), is that we need to have knowledge about how to teach for statistical literacy. I would like to add, given that the results from my studies indicate that it was not only students who struggled with different aspects of statistical literacy but also the prospective teachers, that this implication is true for teacher education as well. Also, considering that Gal's (2002, 2004) model is designed to describe statistical literacy among adults, the results from papers 1–3 can function as an illustration of the development of statistical literacy as a process (e.g., Callingham & Watson, 2017; Gal, 2002; Sharma, 2017), where one can expect different levels depending on who and when you ask. Here, the teachers more often expressed aspects of statistical literacy that had a Pedagogical knowledge base, a result perhaps not surprising given that the teachers had had more experience of the measures in a didactical setting.

Another aspect of the ability to connect a concept to a meaningful context that was illuminated by the results, was the different mathematical properties that were part of the two objects, namely, input value and output value. With respect to input value, even in so-called bare tasks, the task can request identifying arguments (e.g., Eriksson & Sumpter, 2021) as part of understanding what transformation should be used and why. In that way, the task to some degree requires both procedural knowledge and conceptual knowledge (e.g., Bakker, 2004; Gal, 2019; Groth, 2014), but also a critical stance, an aspect of statistical literacy that the present thesis has not focused on. Similar reasoning can be applied to output value and evaluating argument (e.g., Eriksson & Sumpter, 2021). There is critical research that looks at the development of students' critical ability (Büscher, 2019; McCright, 2012; Zapata-Cardona & Martínez-Castro, 2021). The importance of combining a critical stance and the context as the core of statistical literacy is highlighted in Büscher's (2019) concept 'situativity of knowledge'. If we extend the implications raised in the paragraph above, teaching statistical literacy also means it is not only needed for learning the different mathematical properties and connecting these to a context, but also for critically assessing how mathematics is used and the impact of such use.

The second overarching research question was: What do textbook tasks afford with respect to different aspects of statistical literacy regarding mean, median, and mode? The results from papers 4 and 5 signal that what is afforded in these textbooks differs between the measures. Again, mean was overrepresented (e.g., Watson, 2013). Given that textbooks are an important part of mathematics teaching (Glasnovic Gracin, 2018; Jones & Pepin, 2016; Pepin & Haggarty, 2001; Remillard, 2005), the conclusion is that what is afforded to students with respect to statistical literacy, will be vastly different. The different aspects of statistical literacy in focus here were mathematical knowledge and use of words, but there are several more aspects that could be part of further research. For instance, given the discussion about conceptions and context above, one study could analyse how often students are asked to identify data in context, or evaluate answers (i.e., output value) with respect to context. Another issue, which has not been in focus here, is whether tasks allow students to develop their reasoning in several stages. Examples of such studies are the work of Leavy and Hourigan (2016) and Makar (2014) which looked at the area of inferences. In those tasks, it is not enough to follow the chain of input value-transformation-output value; the students need to use the output value for another cycle, to go beyond the data (e.g., Curcio, 1987;

Makar et al., 2011). It is therefore interesting to analyse textbook tasks with respect to how many of them challenge the students to reason in several cycles.

Combining the answers to these two questions, the low level of statistical literacy expressed by the different respondents is a reflection of the low level of statistical literacy afforded by the textbook tasks. The implications of such a conclusion are many. First, teachers have a great responsibility when choosing tasks from textbooks. However, given that teachers themselves appear to have difficulty with some aspects of statistical literacy, especially with median and mode, textbook authors have a great responsibility for balancing content. In the beginning, I described why this topic was relevant to me from a personal perspective. Because I see myself as a teacher educator, it is important to discuss implications for teacher education. The low level of statistical literacy expressed by the prospective teachers could be a direct result of them not yet having had their course in statistics so their responses may merely reflect the limitations of their previous schooling. This means that we as teacher educators must be aware of this and embrace the challenge.

7. Svensk sammanfattning

Bakgrunden till denna kvalitativa studie är ett intresse för att undersöka statistisk litteracitet med fokus på lägesmått. Forskningsområdet är stort internationellt, och denna studie bidrar till att kunna göra jämförelser mellan tidigare internationella forskningsresultat med studiens resultat som avser en svensk kontext. För att kunna belysa statistisk litteracitet ur olika perspektiv valdes fyra olika urval av datakällor: blivande lärare, elever, lärare och läroböcker. Inom alla ingående studier har fokus varit på lägesmåtten medelvärde, median och typvärde, årskurs 4–6. Skälet för dessa årskurser är att begreppen introduceras under dessa skolår.

Studien har genomförts med blandade metoder (jämför ’mixed methods’, t.ex., Mason, 2006). Skäl för metodval, att genomföra metoderna flera gånger samt att undersöka fenomenet i flera urval, utgör olika sätt för att öka studiens validitet (Mason, 2006, 2018).

Begreppet statistisk litteracitet, att vara statistiskt läskunnig, innehåller två ord, statistisk och litteracitet. Båda orden är viktiga i definitioner av begreppet som omfattar både färdigheter gällande beräkningar och begrepp samt förmågan att tolka data utifrån kritiskt tänkande och kommunikation. Utifrån denna bredd av möjliga innehåll har denna avhandling avgränsats till att fokusera på matematisk kunskap, kunskap om kontext, användning av ord samt affektiva element (t.ex. självförtroende och uppfattningar). Omfattning av ingående begrepp som studeras medför en stor variation av teorier applicerbara inom området. Två olika metoder har använts, dels enkät för att samla in förstahandsdata från specifika intressegrupper, dels läroboksanalys för att samla in befintlig andrahandsdata.

Avhandlingen består av problemdrivna studier (jämför, Schoenfeld, 1992), och målet är att undersöka och diskutera olika aspekter av statistisk litteracitet. För att möta detta mål ställs följande två övergripande forskningsfrågor: a) Vilka aspekter av statistisk litteracitet uttrycker elever, lärare och lärarstuderande om medelvärde, median och läge? b) Vilket möjligt lärande erbjuder läromedlens uppgifter angående medelvärde, median och typvärde?

De övergripande frågorna har besvarats med stöd av de fem ingående artiklarna. I första avsnittet nedan presenteras de tre artiklarna som genererat primärdata, därefter de två artiklarna som genererat sekundärdata. Slutligen avslutas med en sammanfattning av studiens övergripande forskningsfrågor samt implikationer.

8.1 Artikel 1–3

Syftet med den explorativa studien i Paper 1 (Landtblom, 2018a) var att studera blivande lärares uppfattningar om medelvärde, median och typvärde utifrån förklaringar av begreppen till elever i årskurs 4–6. Forskningsfrågan var: Hur förklarar blivande lärare begreppen medelvärde, median och typvärde för en elev i årskurs 4–6? Data samlades in genom en enkät med öppna frågor där de blivande lärarna skrev tänkta förklaringar till elevgruppen.

Lärarstudenternas svar analyserades induktivt och genererade sju koder: begreppskunskap, procedurkunskap, kontext, vardagliga begrepp, användbarhet, statistik (matematik) och didaktik (undervisning). Koderna grupperades i tre preliminära kategorier: Användning av ord, Förstå lägesmått och Undervisningsförklaring.

I svaren framstod medelvärde som det lägesmått som de blivande lärarna är mest bekanta med, vilket kan ses i kategorierna Användning av ord och Förstå medelvärden. Respondenterna använde, förutom definitioner och sifferexempel, mer kontext i dessa förklaringar samt exempel på användbarhet jämfört med de andra lägesmåtten. Däremot användes många vardagsord istället för det svenska ordet medelvärde. När man tittade på medianen fokuserade svaren starkt på användningen av tallinjer för att hitta talet i mitten, vilket kategoriserades som Undervisningsförklaring. Typvärde visade sig vara det lägesmått som de blivande lärarna var minst bekant med, eftersom många inte kunde ge exempel eller svarade att de inte visste vad det var. Ett annat tecken på att de inte var bekanta med typvärde, kategoriserat som Användning av ord, visades genom hur respondenterna tolkade den sista delen av ordet värde i ordet typvärde. Påföljden var en hög andel förklaringar där respondenterna använde uteslutande numeriska värden.

Det övergripande resultatet var att en hög andel av de blivande lärarna uppvisade en djupgående procedurkunskap, men svagare konceptuell kunskap, vilket innebär att lärarutbildningen måste möta behovet av att förbättra deras begreppskunskaper för att utveckla en högre nivå av statistisk litteracitet. Även om respondenterna hade en viss koppling till mig, vilket gör resultaten något begränsade i termer av generaliserbarhet, är resultaten ändå starkt överensstämmande med tidigare forskning.

I Paper 2 (Landtblom & Sumpter, 2021) var syftet att studera blivande lärares och lärares uppfattningar om lägesmått. Forskningsfrågorna var: (1) Vad kännetecknar de motiveringar som blivande lärare och lärare ger till vilket av lägesmåtten som är lättast eller svårast att förklara? (2) Vilka är deras uttryckta uppfattningar om användbarheten av lägesmåtten? (3) Hur skiljer sig blivande lärare och lärare i sina svar?

Data samlades in genom en enkät med öppna frågor, som gavs till lärare och blivande lärare som frivilligt ställde upp på att delta. Trots att svarsfrekvensen var låg, 35 procent, genererades rika svar lämpliga för den kvalitativa analysen. Det första steget i den induktiva analysen var att identifiera tecken på gemensamma uppfattningar i alla frågor. Fem teman genererades: Personligt, Pedagogiskt, Matematiskt, Allmänt och Ej relevant. Svar kategoriserade som Personligt gällde vanligtvis deras egna kunskaper, Pedagogiskt hur man lär ut lägesmått, Matematiskt innefattade definitioner och beräkningar och Allmänt berörde daglig användning i samhället.

Resultatet gällande hur lätt eller svårt ett lägesmått ansågs vara att förklara visade att lärarna ansåg att typvärde var lättast att förklara och att medelvärde och median var svårast. De blivande lärarna ansåg däremot det svårast att förklara typvärde. Dessutom skilde sig motivationerna i de två respondentgrupperna. Medan lärarna till övervägande del baserade sina motiveringar på pedagogiska förklaringar, baserade de blivande lärarna sin på matematiska erfarenheter, ofta procedurer.

När det gäller svar på frågan om användbarheten av lägesmått, fann både blivande lärare och lärare medelvärde vara det mest användbara och median eller typvärde vara de minst användbara. Många motiveringar för de mest användbara, för båda grupperna, relaterade till verkliga situationer och båda grupperna gav verkliga exempel. Endast lärare gav pedagogiska motiveringar. För minst användbara var motiveringarna från båda grupperna huvudsakligen generella eller inte relevanta, och inga verkliga exempel gavs för typvärde.

Resultaten indikerade skillnader mellan grupperna: de blivande lärarnas motivation gällde den egna kunskapen medan lärarna till övervägande del gav pedagogiska motiveringar och argumenterade utifrån ett undervisningserfarenhetsperspektiv. Respondentgruppernas varierande erfarenheter kan ses som en förklaring till resultatet. Denna skillnad analyserades statistiskt med Fishers Exact Test, vilket visade en signifikans. Men på grund av urvalet bör dessa resultat endast ses som indikationer.

I Paper 3 (Landtblom & Sumpter, submitted) var syftet att studera statistisk litteracitet uttryckt av elever i årskurs 6 med utgångspunkt från lägesmått. Forskningsfrågorna som ställdes var: (1) Vilka olika kunskapselement använder eleverna när de förklarar lägesmått, och hur uttrycks de olika kunskapselementen? (2) Vilket lägesmått är, enligt elever i årskurs 6, lättast eller svårast att förklara, och vad kännetecknar deras motiveringar? (3) Vilket lägesmått anses, enligt elever i årskurs 6, vara mest eller minst användbart, och vad kännetecknar deras motiveringar? De elever som deltog i denna studie kontaktades via sina lärare som frivilligt hade anmält sig att delta i studien. Data samlades in genom en digital enkät bestående av öppna frågor. I denna deduktiva studie användes det anpassade ramverket för statistisk litteracitet (se Figure 3) för att kategorisera data. Kunskapselementen i ramverket är Matematisk kunskap, Användning av ord och Kontextkunskap. Dessutom analyserades även elevernas Uppfattningar.

Resultat för fråga 1 visar att de vanligaste kunskapselementen som eleverna använder är Matematisk kunskap och Användning av ord. För alla lägesmått uttrycktes matematisk kunskap huvudsakligen med hjälp av algoritmer. Men många valde fel algoritm för medelvärdet. En stor andel gav korrekta men ofullständiga algoritmer för median och typvärde. Vardagliga ord användes främst för medelvärde, medan beskrivande ord användes för median och typvärde. Ett intressant resultat för typvärde var att homonymen typ användes för att uttrycka betydelsen av typvärde, vilket medförde beskrivningen av typvärde som ett ungefärligt värde.

Resultaten för fråga 2, om hur lätta eller svåra lägesmåtten är att förklara, visar att många av eleverna fann att typvärde var det enklaste lägesmåttet att förklara och median det svåraste. Kategorierna motiveringarna för lättast att förklara föll främst i kategorin Uppfattningar. Även kategorierna Matematisk kunskap och Användning av Ord förekom. Kategorierna för motiveringarna för svårast att förklara fanns främst inom Uppfattningar, Matematisk kunskap och Vet ej. Resultaten visar att eleverna tycker att algoritmen för typvärde är lätt medan median är svår att förklara. De uttrycker också en osäkerhet om innebörden av median.

Resultaten för fråga 3, lägesmåttens användbarhet, visar att medelvärdet anses vara det mest användbara och medianen som minst. Kontextkunskap var den mest framträdande kategorin hos motiveringar för användbarhet. Begreppen medelvärde och typvärde anses vara mest användbara, vilket visas med exempel på lämpliga kontexter. Medianen ansågs dock vara minst användbar. Eleverna anser att det är ett begrepp som bara finns i en skolkontext.

För den statistiska analysen användes ett chi-två test. Resultaten för den första frågan indikerade skillnader i fördelningen av kunskapselementen mellan medelvärde, median och typvärde. Resultaten för fråga 2 indikerade skillnader i fördelningen av de sex kategorierna: Matematisk kunskap, Kontextkunskap, Användning av Ord, Uppfattningar, Vet ej och Ej relevant, mellan lägesmåtten. Resultaten för både lättast och svårast att förklara var signifikanta.

8.2 Artikel 4–5

I Paper 4 (Landtblom, 2018b) var syftet att undersöka definitioner av och uppgifter om begreppet typvärde i svenska läroböcker. Forskningsfrågan som ställdes var: Vilken kunskap, procedurell eller konceptuell, och vilken typ av data, kvantitativa eller kvalitativa, erbjuds i uppgifter i läroböcker för åk 4–6 gällande begreppet typvärde? I denna studie analyserades uppgifter från sju svenska läroboksserier för årskurs 4–6, med hänsyn tagen till om uppgifterna i uppgifterna var kvalitativa eller kvantitativa och om de gav procedurkunskaper eller konceptuella kunskaper. Konceptuell kunskap identifierades genom matematiska egenskaper. Det vill säga att det kanske inte finns något typvärde eller fler än ett typvärde och att det är tillämpbart för nominella värden. Analysen fokuserade också på formuleringen av definitionerna av typvärde i läroböckerna.

Resultaten visar att uppgifterna framförallt erbjuder eleverna att träna på procedurkunskap och kvantitativa data. En hög andel av arbetsuppgifterna gav procedurkunskap (82 %). De återstående uppgifterna, som ger konceptuell kunskap, innehöll den matematiska egenskapen att det kan finnas fler än ett typvärde. Resultaten för kvantitativa och kvalitativa data visade att 75 procent av kontexterna i uppgifterna beskrev kvantitativa data. I uppgifterna med kvalitativa data användes vardagliga ord som "mer populär" eller "mest" i frågan istället för att explicit fråga efter typvärdet. Endast två procent av uppgifterna erbjuder att träna på konceptuell kunskap i ett kvalitativt sammanhang.

I förhållande till definitionernas ordalydelse visade resultaten en användning av ord relaterade till kvantitativa värden, varav ett var ordet tal i stället för värde i definitionen. Endast en läroboksserie exemplifierad med både kvalitativa och kvantitativa värden. Fördelningen av uppgifter testades med Fishers Exact Test och befanns statistiskt signifikant.

I Paper 5 (Landtblom, submitted) var syftet att dra slutsatser från texten i uppgifterna om vilka möjligheter till lärande de erbjuder. Forskningsfrågorna som ställdes var: (1) Vilken är fördelningen mellan icke-kontextuella och kontextuella uppgifter? (2) Vilka möjligheter att lära sig om a) indata objekt, b) transformationer och c) utdata objekt erbjuder läroboksuppgifterna, och hur ser distributionen ut? I denna studie har uppgifter från sju svenska läroboksserier för årskurs 4–6 analyserats. I analysen var det inte bara formuleringen av uppgiften som stod i fokus. Jag tittade också på möjliga lösningar och utdata objektet (d.v.s. svaret). Indata objektet analyserades med hjälp av datanivån (d.v.s. nominal, ordinal och kvantitativ). Transformationer analyserades med hjälp av kategorierna procedurell (transformation nära definitionen) eller konceptuell (till exempel omvänd beräkning). Utdataobjekt analyserades med avseende på matematiska egenskaper.

Resultaten av analysen visade att, när det gäller fördelningen av kontextuella och icke-kontextuella uppgifter, var 61 procent av uppgifterna kontextuella, som gav explicita eller implicita möjligheter att träna på matematiska egenskaper. Det var en ojämn fördelning mellan uppgifter om medelvärde, median och typvärde. Uppgifter om medelvärde (62 % av alla uppgifter), genererade fler kontextuella än icke-kontextuella uppgifter. Uppgifter om median (19 % av alla uppgifter), genererade ungefär lika många av varje slag. Uppgifter om typvärde (18 % av alla uppgifter) genererade fler icke-kontextuella än kontextuella uppgifter. Fördelningarna, som testades med användning av ett chi-två test, befanns vara signifikanta. Dessa resultat signalerar att det genereras färre uppgifter om typvärde som hanterar matematiska egenskaper.

Resultaten för fråga 2a, om indata objekt, är också statistiskt signifikanta enligt ett chi-två test. Läroböckerna gav möjligheter att lära sig om alla tre lägesmåtten i kvantitativa sammanhang (92 %). Resultatet avseende transformationer, fråga 2b, visar på en hög andel procedurella transformationer som ligger nära definitionerna. För alla uppgifter om medelvärde var 67 procent av detta slag, för median 77 procent och för typvärde 81 procent. Det fanns alltså fler andra transformationer än procedurmässiga för medelvärdet. Den vanligaste av dessa transformationer var omvänd beräkning (22 %).

I resultaten för fråga 2c, utdata objekt, var den mest frekventa matematiska egenskapen i alla uppgifter (48 %) när ett medelvärde inte är lika med något av värdena från datamängden. Den näst vanligaste matematiska egenskapen (20 % av alla uppgifter) var att betrakta noll som ett värde i datamängden. Den tredje vanligaste matematiska egenskapen (10 % av alla uppgifter) var att ett lägesmått kan korrespondera till fler än en datamängd.

8.3 Sammanfattning

I början av denna uppsats tog jag upp två övergripande frågor. Den första var: Vilka aspekter av statistisk litteracitet uttrycker elever, lärare och lärarstuderande om medelvärde, median och läge? När man kombinerar resultaten från Paper 1–3 och jämför dessa med modellen för statistisk litteracitet som presenteras av Gal (2002, 2004), är den övergripande slutsatsen att det finns en stor skillnad när man talar om medelvärde, median eller typvärde. Skevheten i hur mycket forskning det finns om medelvärde jämfört med de andra två (t.ex. Groth & Bergner, 2006, 2013; Watson, 2013) speglar i viss mån den uttryckta statistiska litteracitet som uttrycks av de olika respondentgrupperna. Några exempel på detta, illustrerade i de tre olika studierna, var att begreppet medelvärde ofta visade mer procedurella föreställningar med koppling till både vilka ord som användes och gällande kunskap om kontexter. Även om respondenterna tyckte att proceduren (dvs. transformationen) var lite svårare än de för de andra lägesmåtten, ansåg de fortfarande att medelvärdet var mer användbart än de övriga två. Detta kan jämföras med median – där respondenterna tyckte att transformationen var lätt att förklara men kämpade för att koppla begreppet till ett meningsfullt sammanhang.

Forskare diskuterar behovet av att utveckla statistisk litteracitet som ett övergripande mål för undervisning och lärande i statistik (Ben-Zvi & Garfield, 2004; Franklin et al., 2007; Franklin & Kader, 2010; Hannigan et al., 2013; Shaughnessy, 2007), och kopplat till detta visar resultaten från denna avhandling behovet av att hantera tal med en kontext (t.ex. Cobb & Moore, 1997). Innebörden av detta, som redan tagits upp av Ben-Zvi och Garfield (2004), är att vi behöver ha kunskap om hur man undervisar för statistisk litteracitet. Jag skulle vilja tillägga, med tanke på att resultaten från mina studier tyder på att det inte bara var elever som kämpade med olika aspekter av statistisk litteracitet utan även de blivande lärarna, att denna implikation gäller även för lärarutbildningen.

Med tanke på att Gals (2002, 2004) modell är utformad för att beskriva statistisk litteracitet bland vuxna, kan resultaten från artiklarna 1–3 fungera som en illustration av utvecklingen av statistisk läskunnighet som en process (t.ex. Callingham & Watson, 2017; Gal, 2002; Sharma, 2017), där man kan förvänta sig olika nivåer beroende på vem och när man frågar. I avhandlingens resultat uttryckte lärarna oftare aspekter av statistisk litteracitet kategoriserat som Pedagogisk kunskap, ett resultat som kanske inte är förvånande med tanke på att lärarna hade mer erfarenhet av lägesmåtten i en didaktisk miljö.

En annan aspekt av förmågan att koppla ett begrepp till en meningsfull kontext som belystes av resultaten, var de olika matematiska egenskaperna som kopplas till datanivån för ingångende data och utgångende objekt. När det gäller ingångsvärden, även i så kallade nakna uppgifter, kan uppgiften efterfråga identifierande argument (t.ex. Eriksson & Sumpter, 2021) som en del av att förstå vilken transformation som ska användas och varför. På det sättet kräver uppgiften till viss del både procedurkunskap och konceptuell kunskap (t.ex. Bakker, 2004; Gal, 2019; Groth, 2014), men också ett kritiskt ställningstagande, en aspekt av statistisk läskunnighet som denna avhandling inte har fokuserat på. Liknande resonemang kan appliceras på utgående objekt och utvärderande argument (Eriksson & Sumpter, 2021). Det finns kritisk forskning som tittar på utvecklingen av elevers kritiska förmåga (t.ex. Büscher, 2019; McCright, 2012; Zapata-Cardona & Martínez-Castro, 2021). Vikten av att kombinera ett kritiskt ställningstagande och kontexten som kärnan i statistisk litteracitet lyfts fram i Büschers (2019) begrepp ’kunskapens situativitet’. Om vi utvidgar de implikationer som tas upp i stycket ovan, innebär undervisning i statistisk litteracitet också att det inte bara behövs för att lära sig de olika matematiska egenskaperna och koppla dessa till en kontext, utan också för att kritiskt bedöma hur matematik används och effekten av sådan användning.

Den andra övergripande forskningsfrågan var: Vilket möjligt lärande erbjuder läromedlens uppgifter angående medelvärde, median och typvärde? Resultaten från artiklarna 4 och 5 signalerar att det som erbjuds i dessa läroböcker skiljer sig åt mellan lägesmåtten. Återigen var medelvärdet överrepresenterat (t.ex. Watson, 2013). Med tanke på att läroböcker är en viktig del av matematikundervisningen (t.ex. Glasnovic Gracin, 2018; Johansson, 2006; Jones & Pepin, 2016; Pepin & Haggarty, 2001; Remillard, 2005), är slutsatsen att det som erbjuds elever med avseende på statistisk litteracitet, är framförallt procedurell matematisk kunskap samt det inte berör alla möjliga matematiska egenskaper. De olika aspekterna av statistisk litteracitet i fokus här var matematisk kunskap och användning av ord, men det finns flera aspekter som skulle kunna ingå i vidare forskning. Till exempel, med tanke på diskussionen om uppfattningar och kontext ovan, kan en studie analysera hur ofta eleverna uppmanas att identifiera data i sitt sammanhang, eller utvärdera svar (d.v.s. det utgående objektet) med avseende på kontext.

En annan fråga, som inte har varit i fokus här, är om uppgifter gör att eleverna kan utveckla sina resonemang i flera steg. Exempel på sådana studier är genomförda av Leavy och Hourigan (2016) och Makar (2014) som tittade på området för statistisk inferens. I de uppgifter som hanterar inferens räcker det inte med att följa kedjan av ingångs värde-transformation-utgångs värde; eleverna behöver även använda utgångsvärdet för en ny cykel för att resonera bortom ingående data (t.ex. Curcio, 1987; Makar et al., 2011). Det är därför intressant att analysera läroboksuppgifter med avseende på hur många av dem som utmanar eleverna att resonera i flera cykler.

References

Ainley, J., & Pratt, D. (2017). Computational modelling and children's expressions of signal and noise. *Statistics Education Research Journal*, *16*(2), 15–36.

Alhojailan, M. I. (2012). Thematic analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences*, *1*(1), 39–47.

Alm, S. E., & Britton, T. (2008). *Stokastik: Sannolikhetsteori och statistikteori med tillämpningar*. Liber.

Arnold, P. (2008). 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* (pp.1–6). https://doi.org/10.52041/SRAP.08507

Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education*, *11* (1). https://doi.org/10.1080/10691898.2003.11910694

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*(2), 64–83. https://doi.org/10.52041/serj.v3i2.552

Bakker, A., & Gravemeijer, K. P. (2006). An historical phenomenology of mean and median. *Educational Studies in Mathematics*, *62*(2), 149–168. https://doi.org/10.1007/s10649-006-7099-8

Ball, D. L. (2000). Bridging practices: Intertwining content and pedagogy in teaching and learning to teach. *Journal of Teacher Education*, *51*, 241–247. https://doi.org/10.1177/0022487100051003013

Barwell, R. (2005). Ambiguity in the mathematics classroom. *Language and Education*, *19*(2), 117–125. https://doi.org/10.1080/09500780508668667

Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi, & J. B. Garfield. (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 121–145). Springer. https://doi.org/10.1007/1-4020-2278-6_6

Ben-Zvi, D., & Garfield, J. B. (2004). Statistical Literacy, Reasoning and Thinking: Goals, Definitions, and Challenges. In D. Ben-Zvi, & J. B.

Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-16). Kluwer Publishers. https://doi.org/10.1007/1-4020-2278-6_17
Ben-Zvi, D., & Makar, K. (2016). International perspectives on the teaching and learning of statistics. In D. Ben-Zvi, & K. Makar. (Eds.), *The teaching and learning of statistics* (pp. 1–10). Springer. https://doi.org/10.1007/978-3-319-23470-0_1
Beswick, K., Callingham, R., & Watson, J. (2012). The nature and development of middle school mathematics teachers' knowledge. *Journal of Mathematics Teacher Education*, *15*(2), 131–157. https://doi.org/10.1007/s10857-011-9177-9
Bibby, J. (1986). 1786-1986: Two centuries of teaching statistics. In *Proceedings of the Second International Conference on Teaching Statistics, ICOTS 2)* Victoria, Canada. (pp. 478–493). International Statistical Institute.
Biesta, G. (2012). Mixed methods. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education* (pp. 147–152). Sage.
Bond, M. E., Perkins, S. N., & Ramirez, C. (2012). Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. *Statistics Education Research Journal*, *11*(2), 6–25. https://doi.org/10.52041/serj.v11i2.325
Box, G. E. P., & Luceño, A. (1997). *Statistical control: By monitoring and feedback adjustment*. John Wiley & Sons.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
Bryman, A. (2018). *Samhällsvetenskapliga metoder*. Liber.
Burkhardt, H., & Swan, M. (2013). Task design for systemic improvement: principles and frameworks. In C. Margolinas (Ed.), *Task design in mathematics education*. *Proceedings of ICMI Study 22*, (pp. 431–440). Available from
Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 57–69). Springer. https://doi.org/10.1007/978-94-007-1131-0_10
Büscher, C. (2019). Students' development of measures. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research. International perspectives* (pp. 27–50). Springer. https://doi.org/10.1007/978-3-030-03472-6_2
Büscher, C. (2022a). Learning opportunities for statistical literacy in German middle school mathematics textbooks. In *Proceedings of the Twelfth Congress of the European Society for Research in Mathematics Education*

(*CERME12*). Büscher, C. (2022b). Design principles for developing statistical literacy in middle schools. *Statistics Education Research Journal, 21*(1), 1–8. https://doi.org/10.52041/serj.v21i1.80
Büscher, C., & Prediger, S. (2019). Students' reflective concepts when reflecting on statistical measures: A design research study. *Journal Für Mathematik-Didaktik*, *40*(2), 197–225. https://doi.org/10.1007/s13138-019-00142-2
Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, *27*(5), 777–786. https://doi.org/10.1037/0012-1649.27.5.777
Byström, J., & Byström, J. (2011). *Grundkurs i statistik*. (7th ed.). Natur och kultur.
Cai, J., & Gorowara, C. C. (2002). Teachers' conceptions and constructions of pedagogical representations in teaching arithmetic average. In M. G. Ottaviani (Ed.), *Proceedings of the sixth International Conference on Teaching of Statistics (ICOTS6)*. International Statistical Institute.
Cai, J., & Moyer, J.C. (1995). *Middle school students' understanding of average: A problem solving approach*. [Paper presentation] Annual Meeting if the North American Chapter of the International Group for the Psychology of Mathematics Education.
Callingham, R., & Watson, J.M. (2011). Measuring levels of statistical pedagogical content knowledge. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education: A joint ICMI/IASE study* (pp. 283–293). Springer. https://doi.org/10.1007/978-94-007-1131-0_28
Callingham, R., & Watson, J. M. (2017). The development of statistical literacy at school. *Statistics Education Research Journal*, *16*(1), 181–201. https://doi.org/10.52041/serj.v16i1.223
Callingham, R., Carmichael, C., & Watson, J. M. (2016). Explaining student achievement: The influence of teachers' pedagogical content knowledge in statistics. *International Journal of Science and Mathematics Education*, *14*(7), 1339–1357. https://doi.org/10.1007/s10763-015-9653-2
Canada, D.L. (2008). Conceptions of distribution held by middle school students and preservice 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.* https://doi.org/10.52041/srap.08314
Canobi, K. H. (2009). Concept–procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology*, *102*(2), 131–149. https://doi.org/10.1016/j.jecp.2008.07.008
Carmichael, C., Callingham, R., Watson, J., & Hay, I. (2009). Factors influencing the development of middle school students' interest in statistical literacy. *Statistics Education Research Journal*, *8*(1), 62–81. https://doi.org/10.52041/serj.v8i1.459

Carmichael, C., Callingham, R., Hay, I., & Watson, J. (2010). Statistical literacy in the middle school: The relationship between interest, self-efficacy and prior mathematics achievement. *Australian Journal of Educational & Developmental Psychology*, *10*, 83–93.
Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3), 1–14. https://doi.org/10.1080/10691898.2002.11910677
Charalambous, C. Y., Delaney, S., Hsu, H. Y., & Mesa, V. (2010). A comparative analysis of the addition and subtraction of fractions in textbooks from three countries. *Mathematical Thinking and Learning*, *12*(2), 117–151. https://doi.org/10.1080/10986060903460070
Chick, H. L. (2007). Teaching and learning by example. *Mathematics: Essential Research, Essential Practice*, *1*, 3–21.
Chick, H. L., & Pierce, R. (2008a). Issues associated with using examples in teaching statistics. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda, Proceedings of the *Joint meeting of the PME 32 and PME-NA XXX* (Vol. 2, pp. 321–328). Universidad Michoacana de San Nicolás de Hidalgo.
Chick, H. L., & Pierce, R. U. (2008b). Teaching statistics at the primary school level: Beliefs, affordances and pedagogical content knowledge. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education, Proceedings of the ICMI study 18 and 2008 IASE Round Table Conference*. https://doi.org/10.52041/srap.08303
Chick, H. L., & Pierce, R. (2012). Teaching for statistical literacy: Utilising affordances in real-world data. *International Journal of Science and Mathematics Education*, *10*(2), 339–362. https://doi.org/10.1007/s10763-011-9303-2
Chick, H., & Pierce, R. (2013). The statistical literacy needed to interpret school assessment data. *Mathematics Teacher Education and Development*, *15*(2), n2.
Chu, Y. (2017). Twenty years of social studies textbook content analysis: still "decidedly disappointing"?. *The Social Studies*, *108*(6), 229–241. https://doi.org/10.1080/00377996.2017.1360240
Clark, J., Kraut, G., Mathews, D., & Wimbish, J. (2007). The fundamental theorem of statistics: Classifying student understanding of basic statistical concepts. *Unpublished manuscript.*
Cloutier, C., & Ravasi, D. (2021). Using tables to enhance trustworthiness in qualitative research. *Strategic Organization*, *19*(1), 113–133. https://doi.org/10.1177/1476127020979329
Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, *104*(9), 801–823. https://doi.org/10.1080/00029890.1997.11990723

Coe, R. J. (2012). Conducting your research. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education* (pp. 41–52). Sage.

Cogan, L. S., Schmidt, W. H., & Wiley, D. E. (2001). Who takes what math and in which track? Using TIMSS to characterize US students' eighth-grade mathematics learning opportunities. *Educational Evaluation and Policy Analysis*, *23*(4), 323–341. https://doi.org/10.3102/01623737023004323

Cooper, L. L., & Shore, F. S. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education*, *16*(2). https://doi.org/10.1080/10691898.2008.11889559

Cooper, L. L., & Shore, F. S. (2010). The effects of data and graph type on concepts and visualizations of variability. *Journal of Statistics Education*, *18*(2). https://doi.org/10.1080/10691898.2010.11889487

Creswell, J.W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*, (3rd ed.). Sage.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education*, *18*(5), 382–393. https://doi.org/10.2307/749086

Danielsson, T. (1964). *Sagor för barn över 18 år*. Wahlström & Widstrand.

Da Ponte, J., Pereira, J., Enriques, A., & Quaresma, M. (2013). *Designing and using exploratory tasks*. Paper presented at ICMI study, 22, Task design in mathematics education, Oxford, United Kingdom. 491–500.

DeCarlo, M. (2018). *Scientific inquiry in social work*. Open Social Work Education.

Denscombe, M. (2017). *The good research guide: For small-scale social research projects*. Open University Press

delMas, R. C. (2002). Statistical literacy, reasoning, and learning: A commentary. *Journal of Statistics Education*, *10*(3), 1–11. https://doi.org/10.1080/10691898.2002.11910674

diSessa, A. A. (2018). Computational literacy and "the big picture" concerning computers in mathematics education. *Mathematical Thinking and Learning*, *20*(1), 3–31. https://doi.org/10.1080/10986065.2018.1403544

Djurfeldt, G., Larsson, R., & Stjärnhagen, O. (2003). *Statistisk verktygslåda-samhällsvetenskaplig orsaksanalys med kvantitativa metoder*. Studentlitteratur.

Doyle, W. (1983). Academic work. *Review of Educational Research*, 53, 159–199. https://doi.org/10.3102/00346543053002159

Dunn, P. K., Carey, M. D., Richardson, A. M., & McDonald, C. (2016). Learning the language of statistics: Challenges and teaching approaches. *Statistics Education Research Journal*, *15*(1), 8–27. https://doi.org/10.52041/serj.v15i1.255

Durkin, K., & Shire, B. (1991). Lexical ambiguity in mathematical contexts. In K. Durkin & B. Shire (Eds.), *Language in mathematical education: Research and practice* (pp. 71–84). Open University Press.

Engel, J. (2019). Statistical literacy and society: What is civic statistics? In J.M. Contreras, M.M. Gea, M.M. López-Martín, & E. Molina-Portillo (Eds.*), Actas del Tercer Congreso International Virtual de Educación Estadística* (pp. 1–17).

Eriksson, H., & Sumpter, L. (2021). Algebraic and fractional thinking in collective mathematical reasoning. *Educational Studies in Mathematics, 108*(3), 473-491.https://doi.org/10.1007/s10649-021-10044-1

Estrada, A., Batanero, C., Fortuny, J. M., & Díaz, C. (2005). A structural study of future teachers' attitudes toward statistics. In M. Bosch (Ed.), *Proceedings of the 4th Congress of the European Society for Research in Mathematics Education* (pp. 508–517). IQS Fundemi.

Estrada, A., Batanero, C., & Lancaster, S. (2011). Teachers' attitudes towards statistics. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 163–174). Springer. https://doi.org/10.1007/978-94-007-1131-0_18

Fan, L., Zhu, Y., & Miao, Z. (2013). Textbook research in mathematics education, development status and directions. *Zentralblatt für Didaktik der Mathematik*, *45*(5), 633–646. https://doi.org/10.1007/s11858-013-0539-x

Fitzpatrick, A. R. (1983). The meaning of content validity. *Applied Psychological Measurement*, *7*(1), 3–13. https://doi.org/10.1177/014662168300700102

Floden, R. E. (2002). The measurement of opportunity to learn. In A. C. Porter, & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement*, 231–266. National Academies Press. https://www.researchgate.net/publication/270585005

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K-12 curriculum framework*. American Statistical Association.

Franklin, C., & Kader, G. (2010). Models of teacher preparation designed around the GAISE framework. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8)*. International Statistical Institute.

Freeman, J. V., & Campbell, M. J. (2007). The analysis of categorical data: Fisher's exact test. *Scope*, *16*(2), 11-12.

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). Lawrence Erlbaum Associates, Inc.

Gal, I. (1995). Statistical tools and statistical literacy: The case of the average. *Teaching Statistics*, *17*(3), 97–99. https://doi.org/10.1111/j.1467-9639.1995.tb00720.x
Gal, I. (ed.), (2000). *Adult Numeracy Development: Theory, Research, Practice*, Hampton Press.
Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1–25. https://doi.org/10.2307/1403713
Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47–78). Springer. https://doi.org/10.1007/1-4020-2278-6_3
Gal, I. (2019). Understanding statistical literacy: About knowledge of contexts and models. In J. M. Contreras, M. M. Gea, M. M. López-Martín & E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística.*
Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. *The Assessment Challenge in Statistics Education, 12*, 37–51.
Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education, 10*(3), 58–69. https://doi.org/10.1080/10691898.2002.11910676
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. https://doi.org/10.1111/j.1751-5823.2007.00029.x
Gibbons, J. D., & Chakraborti, S. (2014). *Nonparametric statistical inference*. CRC press. https://doi.org/10.1007/978-3-642-04898-2_420
Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin. https://doi.org/10.4324/9781315740218
Glaser, B.G., & Strauss, A.L. (1967). *The discovery of grounded theory*. Aldine.
Glasnovic Gracin, D. (2018). Requirements in mathematics textbooks: A five-dimensional analysis of textbook exercises and examples. *International Journal of Mathematical Education in Science and Technology, 49*(7), 1003–1024. https://doi.org/10.1080/0020739X.2018.1431849
Goldin, G. A. (2002). Affect, meta-affect, and mathematical belief structures. In G. C. Leder, E. Pehkonen & G. Törner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 59–72). Kluwer academic publishers. https://doi.org/10.1007/0-306-47958-3_4
González, O. (2016). A framework for assessing statistical knowledge for teaching based on the identification of conceptions of variability held by teachers. In: D. Ben-Zvi & Makar, K (Eds.), *The teaching and learning of statistics* (pp. 315–325). Springer. https://doi.org/10.1007/978-3-319-23470-0_37

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Greatorex, J. (2014). Context in mathematics examination questions. *Research Matters: A Cambridge Assessment Publication*, *17*, 18–23.

Groth, R. E. (2007). Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, *38*(5), 427–437. https://doi.org/10.2307/30034960

Groth, R.E. (2013). Characterizing key developmental understandings and pedagogical powerful ideas within statistical knowledge for teaching framework. *Mathematical Thinking and Learning*, *15*(2), 121–145. https://doi.org/10.1080/10986065.2013.770718

Groth, R. E. (2014). Prospective teachers' procedural and conceptual knowledge of mean absolute deviation. *Investigations in Mathematics Learning*, *6*(3), 51–69. https://doi.org/10.1080/24727466.2014.11790335

Groth, R. E., & Bergner, J.A. (2006). Preservice elementary teachers' conceptual and procedural knowledge of mean, median and mode. *Mathematical Thinking and Learning*, *8*(1), 37–63. https://doi.org/10.1207/s15327833mtl0801_3

Groth, R. E., & Bergner, J. A. (2013). Mapping the structure of knowledge for teaching nominal categorical data analysis. *Educational Studies in Mathematics*, *83*(2), 247–265. https://doi.org/10.1007/S10649-012-9452-4

Groth, R., & Meletiou-Mavrotheris, M. (2018). Research on statistics teachers' cognitive and affective characteristics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *Springer international handbook of research in statistics education* (pp. 327–355). Springer. https://doi.org/10.1007/978-3-319-66195-7_10

Gutstein, E. (2006). *Reading and writing the world with mathematics*. Routledge.

Hadar, L. L. (2017). Opportunities to learn: Mathematics textbooks and students' achievements. *Studies in Educational Evaluation*, *55*, 153–166. https://doi.org/10.1016/j.stueduc.2017.10.002

Hall, J. (2011). Engaging teachers and students with real data: Benefits and challenges. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 335–346). Springer. https://doi.org/10.1007/978-94-007-1131-0_32

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education*, *16*(6), 427–449. https://doi.org/10.1007/s10857-013-9246-3

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education*, *5*(1), 35–59. https://doi.org/10.1023/A:1013886730487

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–28). Lawrence Erlbaum Associates, Inc.

Hong, D. S., & Choi, K. M. (2018). A comparative analysis of linear functions in Korean and American standards-based secondary textbooks. *International Journal of Mathematical Education in Science and Technology*, *49*(7), 1025–1051.https://doi.org/10.1080/0020739X.2018.1440327

Hourigan, M., & Leavy, A. M. (2020). Using integrated STEM as a stimulus to develop elementary students' statistical literacy. *Teaching Statistics*, *42*(3), 77–86. https://doi.org/10.1111/test/.12229

Huey, M. E., & Jackson, C. D. (2015). A framework for analyzing informal inferential reasoning tasks in middle school textbooks. *North American Chapter of the International Group for the Psychology of Mathematics Education*.

Husén, T. (1967). *International study of achievement in mathematics: A comparison of 12 countries* (Vol. 2). Almquist & Wicksell.

Jacobbe, T. (2012). Elementary school teachers' understanding of the mean and median. *International Journal of Science and Mathematics Education*, *10*(5), 1143–1161. https://doi.org/10.1007/s10763-011-9321-0

Jacobbe, T., & Carvahlo, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 199–209). Springer. https://doi.org/10.1007/978-94-007-1131-0_21

Johansson, M. (2006). *Teaching mathematics with textbooks: A classroom and curricular perspective* [Doctoral dissertation]. Department of Mathematics, Luleå University of Technology.

Jones, D.L., & Jacobbe, T. (2014). An analysis of the statistical content in textbooks for prospective elementary teachers. *Journal of Statistics Education*, *22*(3), 1–17. https://doi.org/10.1080/10691898.2014.11889713

Jones, D. L., Brown, M., Dunkle, A., Hixon, L., Yoder, N., & Silbernick, Z. (2015). The statistical content of elementary school mathematics textbooks. *Journal of Statistics Education*, *23*(3). https://doi.org/10.1080/10691898.2015.11889748

Jones, K., & Pepin, B. (2016). Research on mathematics teachers as partners in task design. *Journal of Mathematics Teacher Education*, *19*(2–3), 105–121. https://doi.org/10.1007/s10857-016-9345-z

Juter, K. (2005). Students' attitudes to mathematics and performance in limits of functions. *Mathematics Education Research Journal*, *17*(2), 91-110. https://doi.org/10.1007/bf03217417

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature climate change*, *2*(10), 732-735. https://doi.org/10.1038/nclimate1547

Kaplan, J. J., Fisher, D. G., & Rogness, N. T. (2009). Lexical ambiguity in statistics: What do students know about the words association, average, confidence, random and spread?. *Journal of Statistics Education*, *17*(3). https://doi.org/10.1080/10691898.2009.11889535

Kaplan, J., Fisher, D. G., & Rogness, N. T. (2010). Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education*, *18*(2). https://doi.org/10.1080/10691898.2010.11889491

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2012). Lexical ambiguity: Making a case against spread. *Teaching Statistics*, *34*(2), 56–60. https://doi.org/10.1111/j.1467-9639.2011.00477.x

Kaplan, J. J., Rogness, N. T., & Fisher, D. G. (2014). Exploiting lexical ambiguity to help students understand the meaning of random. *Statistical Education Research Journal*, *13*(1), 9–24. https://doi.org/10.52041/serj.v13i1.296

Kieran, C., Doorman, M., & Ohtani, M. (2015). Frameworks and principles for task design. In A. Watson & M. Ohtani (Eds.), *Task design in mathematicseEducation* (pp. 19–81). Springer. https://doi.org/10.1007/978-3-319-09629-2_2

Kilpatrick, J. (1993). Beyond face value: Assessing research in mathematics education. In G. Nissen & M. Blomhoj (Eds.), *Criteria for scientific quality and relevance in the didactics of mathematics* (pp. 15–34). Roskilde University.

Kitto, K., Williams, C., & Alderman, L. (2019). Beyond average: Contemporary statistical techniques for analysing student evaluations of teaching. *Assessment & Evaluation in Higher Education*, *44*(3), 338–360. https://doi.org/10.1080/02602938.2018.1506909

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1), 1–11. https://doi.org/10.1080/10691898.1995.11910479

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. https://doi.org/10.1007/1-4020-2278-6_8

Konold, C., & Pollatsek, A. (2004). Conceptualizing an average as a stable feature of a noisy process. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 169–199). Springer. https://doi.org/10.1007/1-4020-2278-6_8

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and the media of contemporary communication*. Hodder Arnold.

Landtblom, K. K. (2018a). Prospective teachers' conceptions of the concepts mean, median and mode. In H. Palmér & J. Skott (Eds.), *Students' and teachers' values, attitudes, feelings and beliefs in mathematics classrooms* (pp. 43–52). Springer. https://doi.org/10.1007/978-3-319-70244-5_5

Landtblom, K. (2018b). Is data a quantitative thing? An analysis of the concept of the mode in textbooks for grade 4–6. In M. A. Sorto, A. White &

L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10*). International Statistical Institute.
Landtblom, K. (submitted). Afforded opportunities to learn mean, median, and mode in textbook tasks.
Landtblom, K., & Sumpter, L. (2021). Teachers and prospective teachers' conceptions about averages. *Journal of Adult Learning, Knowledge and Innovation*, *4*(1), 1–8. https://doi.org/10.1556/2059.03.2019.02
Landtblom, K., & Sumpter, L. (submitted). Which measure is most useful? Grade 6 students' expressed statistical literacy.
Leavy, A. M. (2010). The challenge of preparing preservice teachers to teach informal inferential reasoning. *Statistics Education Research Journal*, *9*(1), 46–67. https://doi.org/10.52041/serj.v9i1.387
Leavy, A. (2015). Looking at practice: Revealing the knowledge demands of teaching data handling in the primary classroom. *Mathematics Education Research Journal*, *27*(3), 283–309. https://doi.org/10.1007/s13394-014-0138-3
Leavy, A., & Hourigan, M. (2016). Crime scenes and mystery players! Using driving questions to support the development of statistical literacy. *Teaching Statistics*, *38*(1), 29–35. https://doi.org/10.1111/test.12088
Leavy, A. M., & Middleton, J. A. (2011). Elementary and middle grade students' constructions of typicality. *The Journal of Mathematical Behavior*, *30*(3), 235–254. https://doi.org/10.1016/j.jmathb.2011.03.001
Leavy, A., & O'Loughlin, N. (2006). Preservice teachers' understanding of the mean: Moving beyond the arithmetic average. *Journal of Mathematics Teacher Education*, *9*(1), 53–90. https://doi.org/10.1007/s10857-006-9003-y
Leavy, A. M., Hannigan, A., & Fitzmaurice, O. (2013). If you're doubting yourself, whats' the fun in that? An exploration of why prospective secondary mathematics teachers perceive statistics as difficult. *Journal of Statistics Education*, *21*(3), n3. https://doi.org/10.1080/10691898.2013.11889684
Leavy, A. M., Hourigan, M., Murphy, B., & Yilmaz, N. (2021). Malleable or fixed? Exploring pre-service primary teachers' attitudes towards statistics. *International Journal of Mathematical Education in Science and Technology*, *52*(3), 427–451. https://doi.org/10.1080/0020739X.2019.1688405
Leder, G. C. (2019). Mathematics-related beliefs and affect. In M.S. Hannula, G.C. Leder, F. Morselli, M. Vollstedt, & Q. Zhang (Eds.), *Affect and mathematics education* (pp. 15–35). Springer. http://doi.org/10.1007/978-3-030-13761-8_2
Lesser, L. M., Wagler, A. E., & Abormegah, P. (2014). Finding a happy median: Another balance representation for measures of center. *Journal of Statistics Education*, *22*(3), 1–27. https://doi.org/10.1080/10691898.2014.11889714

Liljedahl, P., Chernoff, E., & Zazkis, R. (2007). Interweaving mathematics and pedagogy in task design: A tale of one task. *Journal of Mathematics Teacher Education*, *10*(4), 239–249. https://doi.org/10.1007/s10857-007-9047-7

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, *67*(3), 255–276. https://doi.org/10.1007/s10649-007-9104-2

Lithner, J. (2017). Principles for designing mathematical tasks that enhance imitative and creative reasoning. *ZDM Mathematics Education*, *49*(6), 937–949. https://doi.org/10.1007/s11858-017-0867-3

MacGillivray, H., & Pereira-Mendoza, L. (2011). Teaching statistical thinking through investigative projects. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 109–120). Springer. https://doi.org/10.1007/978-94-007-1131-0_14

Makar, K. (2014). Young children's explorations of average through informal inferential reasoning. *Educational Studies in Mathematics*, *86*(1), 61–78. https://doi.org/10.1007/s10649-013-9526-y

Makar, K., & Confrey, J. (2004). Secondary teachers' statistical reasoning in comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 353–374). Kluwer Academic. https://doi.org/10.1007/1-4020-2278-6_15

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal*, *4(1)*, 27–54. https://doi.org/10.52041/serj.v4i1.524

Makar, K., & McPhee, D. (2009). Young children's explorations of average in an inquiry classroom. In R. Hunter, B. Bicknell & T. Burgess (Eds.), *Crossing divides: Proceedings of the 32nd Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 347–354).

Martins, J. A., Nascimento, M. M., & Estrada, A. (2012). Looking back over their shoulders: A qualitative analysis of Portuguese teachers' attitudes towards statistics. *Statistics Education Research Journal*, *11*(2), 26–44. https://doi.org/10.52041/serj.v11i2.327

Mason, J. (2006). Mixing methods in a qualitatively driven way. *Qualitative research*, *6*(1), 9–25. https://doi.org/10.1177/1468794106058866

Mason, J. (2018). *Qualitative researching*. Sage.

Mason, J. (2011). Classifying and characterising: Provoking awareness of the use of a natural power in mathematics and in mathematical pedagogy. In O. Zaslavsky & P. Sullivan (Eds.), *Constructing knowledge for teaching secondary mathematics* (pp. 39–55). Springer. https://doi.org/10.1007/978-0-387-09812-8_3

Mathews, D., & Clark, J. (2003). *Successful students' conceptions of mean, standard deviation, and the Central Limit Theorem* [Unpublished paper]. Retrieved October, *20*, 2007.

Mayén, S., & Díaz, C. (2010). Is median an easy concept? Semiotic analysis of an open-ended task. In C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8*). International Statistical Institute.

Mayring P. (2015) Qualitative content analysis: Theoretical background and procedures. In A. Bikner-Ahsbahs, C. Knipping & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education. Advances in mathematics education* (pp. 365–380). Springer. https://doi.org/10.1007/978-94-017-9181-6_13

McCright, A. M. (2012). Enhancing students' scientific and quantitative literacies through an inquiry-based learning project on climate change. *Journal of the Scholarship of Teaching and Learning*, *12*(4), 86-101.

McCulloch, G. (2012). Documentary methods. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education* (pp. 210–216). Sage.

Meletiou-Mavrotheris, M., & Lee, C. (2003). *Studying the evolution of students' conceptions of variation using the transformative and conjecture-driven research design* [Conference paper]. 3rd International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-3), Lincoln, USA.

Milne, M. J., & Adler, R. W. (1999). Exploring the reliability of social and environmental disclosures content analysis. *Accounting, Auditing & Accountability Journal*, *12*(2), 237–256. https://doi.org/10.1108/09513579910270138

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, *26*(1), 20–39. https://doi.org/10.2307/749226

Moore, D. S. (1990). Uncertainty. In L. Arthur (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp.95–137). Mathematical Science Education Board

Murray, F.B., & Rodney Sharp, H. (1986). Foreword. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. xi–xiii). Lawrence Erlbaum Associates, Inc.

Nilsson, P., Schindler, M., & Bakker, A. (2017). The nature and use of theories in statistics education. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education*, (pp. 359–386). Springer. https://doi.org/10.1007/978-3-319-66195-7_11

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103–123. https://doi.org/10.52041/serj.v11i2.333

Norman, D. A. (1988). *The Psychology of Everyday Things.* Basic Books

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, *16*(1). https://doi.org/10.1177/1609406917733847

O'Connor, C., & Joffe, H. (2020). Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406919899220

Ogburn, W. F. (1940). Statistical trends. *Journal of the American Statistical Association, 35*(209b), 252–260. https://doi.org/10.1080/01621459.1940.10500563

Okeeffe, L. (2013). A framework for textbook analysis. *International Review of Contemporary Learning Research, 2*(1), 1–13. https://doi.org/10.12785/irclr/020101

Palys, T. S., & Atchison, C. (2014). *Research decisions: Quantitative, qualitative, and mixed method approaches*. Nelson Education.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Sage.

Pearson, K. (1895). X. Contributions to the mathematical theory of evolution.—II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London. (A.)*, (186), 343–414. https://doi.org/10.1098/rsta.1895.0010

Pepin, B., & Haggarty, L. (2001). Mathematics textbooks and their use in English, French and German classrooms. *Zentralblatt für Didaktik der Mathematik*, *33*(5), 158-175. https://doi.org/10.1007/bf02656616

Perrett, J. J. (2012). A case study on teaching the topic "experimental unit" and how it is presented in advanced placement statistics textbooks. *Journal of Statistics Education*, *20*(2). https://doi.org/10.1080/10691898.2012.11889640

Peters, S. A. (2009). *Developing an understanding of variation: AP Statistics teachers' perceptions and recollections of critical moments* [Unpublished doctoral dissertation]. Pennsylvania State University, Pennsylvania.

Peters, S. A., Watkins, J. D., & Bennett, V. M. (2014). Middle and high school teachers' transformative learning of center. In K. Makar, B. de Sousa & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the 9th International Conference on Teaching Statistics (ICOTS9)*. International Statistical Institute.

Petocz, P., Reid, A., & Gal, I. (2018). Statistics education research. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 71–99). Springer. https://doi.org/10.1007/978-3-319-66195-7_3

Pickle, M. C. C. (2012). *Statistical content in middle grades mathematics textbooks* [Unpublished doctoral dissertation]. University of South Florida, Tampa, Florida.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, *13*(1–2), 27–46. https://doi.org/10.1080/10986065.2011.538302

Pfannkuch, M. (2018). Reimagining curriculum approaches. In D. Ben-Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 387–413). Springer. https://doi.org/10.1007/978-3-319-66195-7_12

Pfannkuch, M., & Ben-Zvi, D. (2011). Developing teachers' statistical thinking. In C. Batanero, G. Burrill & C. Reading (Eds.), *Teaching statistics in school mathematics: Challenges for teaching and teacher education* (pp. 323–333). Springer. https://doi.org/10.1007/978-94-007-1131-0_31

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and Learning*, *1* (pp. 257–315). Information Age.

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*, *27*, 258–284. https://doi.org/10.1080/00909889909365539

Ramirez, C., Schau, C., & Emmioğlu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, *11*(2), 57–71. https://doi.org/10.52041/serj.v11i2.329

Ranney, M. A., & Clark, D. (2016). Climate change conceptual change: Scientific information can transform attitudes. *Topics in cognitive science*, *8*(1), 49-75. https://doi.org/10.1111/tops.12187

Raymond, M., & Rousset, F. (1995). An exact test for population differentiation. *Evolution*, *49*(6), 1280–1283. https://doi.org/10.2307/2410454

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201–226). Springer. https://doi.org/10.1007/1-4020-2278-6_9

Reinke, L. T. (2019). Toward an analytical framework for contextual problem-based mathematics instruction. *Mathematical Thinking and Learning*, *21*(4), 265–284. https://doi.org/10.1080/10986065.2019.1576004

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211-246. https://doi.org/10.3102/00346543075002211

Rezat, S., & Straeßer, R. (2015). Methodological issues and challenges in research on mathematics textbooks. *Nordic Studies in Mathematics Education*, *20*(3–4), 247–266.

Richardson, A. M., Dunn, P. K., & Hutchins, R. (2013). Identification and definition of lexically ambiguous words in statistics by tutors and students. *International Journal of Mathematical Education in Science and Technology*, *44*(7), 1007–1019. https://doi.org/10.1080/0020739X.2013.830781

Ridgway, R., Nicholson, J., & Gal, I. (2018). Understanding statistics about society: A framework of knowledge and skills needed to engage with civic statistics. In M. A. Sorto, A. White & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)*. International Statistical Institute.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other?. *Journal of Educational Psychology*, *91*(1), 175-189. https://doi.org/10.1037/0022-0663.91.1.175

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures?. *British Journal of Educational Psychology*, *82*(3), 436–455. https://doi.org/10.1111/j.2044-8279.2011.02037.x

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, *27*(4), 587–597. https://doi.org/10.1007/s10648-015-9302-x

Rubenstein, R. N., & Thompson, D. R. (2002). Understanding and supporting children's mathematical vocabulary development. *Teaching Children Mathematics*, *9*(2), 107–112. https://doi.org/10.5951/tcm.9.2.0107

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10* (3), 6–13. https://doi.org/10.1080/10691898.2002.11910678

Russell, S. J., & Mokros, J. R. (1991). What's typical? Children's ideas about average. In j. Landwehr & D. Vere-Jones (Eds.). *Proceedings of the Third International Conference on Teaching Statistics (ICOTS3*), Dunedin New Zealand, (pp. 307–313). International Statistical Institute.

Schield, M. (2017). GAISE 2016 promotes statistical literacy. *Statistics Education Research Journal*, *16*(1), 50–54. https://doi.org/10.52041/serj.v16i1.214

Schnell, S., & Frischemeier, D. (2019). Primary school students reasoning about and with the median when comparing distributions. In U. T. Jankvist, M. Van den Heuvle-Panhuizen & M. Veldhuis *Eleventh Congress of the European Society for Research in Mathematics Education (CERME11)*. Freudenthal Group; Freudenthal Institute; ERME.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Schoenfeld, A. (1992). On paradigms and methods: What do you do when the ones you know don't do what you want them to? Issues in the analysis of data in the form of videotapes. *The Journal of the Learning Sciences*, *2*(2), 179–214. https://doi.org/10.1207/s15327809jls0202_3

Schwarz, B. (2015). A study on professional competence of future teacher students as an example of a study using qualitative content analysis. In A. Bikner-Ahsbahs, C. Knipping & N.C. Presmeg (Eds.), *Approaches to*

qualitative research in mathematics education (pp. 381–399). Springer. https://doi.org/10.1007/978-94-017-9181-6_14
Shahbari, J. A., & Tabach, M. (2020). Making sense of the measure concept through engagement in model-eliciting activities. *International Journal of Mathematical Education in Science and Technology*, 1–18. https://doi.org/10.1080/0020739X.2020.1740803
Sharma, S. (2017). Definitions and models of statistical literacy: A literature review. *Open Review of Educational Research*, *4*(1), 118–133. https://doi.org/10.1080/23265507.2017.1354313
Shaughnessy, J.M. (2007). Research on statistical learning and reasoning. In F.K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp.957–1009). Information age Publishing.
Sheppard, V. (2020). *Research methods for the social sciences: An introduction.* BC Campus.
Silver, E.A. (1986) Using conceptual and procedural knowledge: A focus on relationships. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 181–198). Lawrence Erlbaum Associates, Inc.
Skolverket (2011). *Läroplan för grundskolan, förskoleklassen och fritidshemmet.* Skolverket.
Skolverket (2022). *Läroplan för grundskolan, förskoleklassen och fritidshemmet. Lgr22*. Skolverket.
Skolöverstyrelsen (1951). *Timplaner och huvudmoment för studieplaner för skolor av a- och b- form vid försöksverksamhet i anslutning till 1946 års skolkommissions principförslag*. Kungliga Skolöverstyrelsen.
Skolöverstyrelsen (1955a). *Undervisningsplan för rikets folkskolor den 22 januari 1955*. Norstedt.
Skolöverstyrelsen (1955b). *Timplaner och huvudmoment vid försöksverksamhet med nioårig enhetsskola: [fastställda av Kungl. Skolöverstyrelsen för läsåren 1955/59]*. Svenska bokförlaget.
Skolöverstyrelsen (1962). *Läroplan för grundskolan*. Kungliga Skolöverstyrelsen.
Skolöverstyrelsen (1969a). *Läroplan för grundskolan, Allmän del: Mål och riktlinjer, kursplaner, timplaner*. Utbildningsförlaget.
Skolöverstyrelsen (1969b). *Läroplan för grundskolan 2, Supplement: kompletterande anvisningar och kommentarer*. Utbildningsförlaget.
Skolöverstyrelsen (1980). *Läroplan för grundskolan. Allmän del: Mål och riktlinjer, kursplaner, timplaner*. Liber Läromedel/Utbildningsförlaget.
Son, J. W., & Diletti, J. (2017). What can we learn from textbook analysis? In J.W. Son, T. Watanabe & J.J. Lo (Eds.), *What matters? Research trends in international comparative studies in mathematics education* (pp. 3–32). Springer. https://doi.org/10.1007/978-3-319-51187-0_ 1
Sorto, M. A. (2004). *Prospective middle school teachers' knowledge about data analysis and its application to teaching* [Doctoral dissertation]. Michigan State University. Department of Mathematics.

Sproesser, U., Engel, J., & Kuntze, S. (2016). Fostering self-concept and interest for statistics through specific learning environments. *Statistics Education Research Journal*, *15*(1). https://doi.org/10.52041/serj.v15i1.256
Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, *36*(5), 404–411. https://doi.org/10.2307/30034943
Stein, M. K., & Smith, M. S. (1998). Mathematical task as a framework for reflection: From research to practice. *Mathematics Teaching in the Middle School*, *3*(4), 268–275. https://doi.org/10.5951/MTMS.3.4.0268
Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488. https://doi.org/10.3102/00028312033002455
Stillman, G., Brown, J., Faragher, R., Geiger, V., & Galbraith, P. (2013). The role of textbooks in developing a socio-critical perspective on mathematical modelling in secondary classrooms. In G. A. Stillman (Ed.), *Teaching mathematical modelling: Connection to research and practice. International perspectives on the teaching and learning of mathematical modelling* (pp. 361–371). Springer Science & Business Media. https://doi.org/10.1007/978-94-007-6540-5_30
Strauss, S., & Bichler, E. (1988). The development of childrens' concepts of the arithmetic average. *Journal for Research in Mathematics Education*, *19*(1), 64–80. https://doi.org/10.2307/749111
Swedish Research Council (2017). Good research practice. *Swedish Research Council.*
Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, *12*(2), 151–169. https://doi.org/10.1007/BF00305619
Tarr, J. E., Reys, B. J., Barker, D. D., & Billstein, R. (2006). Selecting high-quality mathematics textbooks. *Mathematics Teaching in the Middle School*, *12*(1), 50–54. https://doi.org/10.5951/mtms.12.1.0050
Thornberg, R. (2012). Grounded theory. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education* (pp. 85–93). Sage.
Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851. https://doi.org/10.1177/1077800410383121
Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, *33*(1), 1–67. https://doi.org/10.1214/aoms/1177704711
Tymms, P. (2012). Questionnaires. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education*, 231–240.

Törnroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*, *31*(4), 315–327. https://doi.org/10.1016/j.stueduc.2005.11.005
UNESCO (2005). *Education for all: Literacy for life*. UNESCO.
UNESCO (2013). *Second global report on adult learning and education: Rethinking literacy*. UNESCO Institute for Lifelong learning.
UNESCO (2017). *Literacy rates continue to rise from one generation to the next. Fact sheet no. 45.* Unesco Institute for Statistics.
UPL (1920). *Undervisningsplan för rikets folkskolor den 31 oktober 1919*. Norstedt.
Usiskin, Z. (2012). What does it mean to understand some mathematics?. In *Selected regular lectures from the 12th International Congress on Mathematical Education.* Springer International Publishing. https://doi.org/10.1007/978-3-319-17187-6_46
Usiskin, Z. (2015a). What does it mean to understand some Mathematics?. In *Selected regular lectures from the 12th international congress on mathematical education* (pp. 821-841). Springer International Publishing. https://doi.org/10.1007/978-3-319-17187-6_46
Usiskin, Z. (2015b). Mathematical modeling and pure mathematics. *Mathematics Teaching in the Middle School*, *20*(8), 476-482. https://doi.org/10.5951/mathteacmiddscho.20.8.0476
Utbildningsdepartementet (1994). *Kursplaner för grundskolan.* Utbildningsdepartementet.
Utts, J. M. (2015). *Seeing through statistics*, (Fourth edition). Cengage Learning.
van Leeuwen, T. (2005). *Introducing social semiotics.* Routledge.
Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Swets & Zeitlinger.
Walker, H. M. (1951). Statistical literacy in the social sciences. *The American Statistician*, *5*(1), 6–12. https://doi.org/10.1080/00031305.1951.10481912
Waring, M. (2012). Grounded theory. In J. Arthur, M. Waring, R. Coe & L.V. Hedges (Eds.), *Research methods and methodologies in education*, 297–308.
Watier, N. N., Lamontagne, C., & Chartier, S. (2011). What does the mean mean?. *Journal of Statistics Education*, *19*(2), 1–20. https://doi.org/10.1080/10691898.2011.11889615
Watson, A. (2007). The nature of participation afforded by tasks, questions and prompts in mathematics classrooms. *Research in Mathematics Education*, *9*(1), 111–126. https://doi.org/10.1080/14794800008520174
Watson, A., & Mason, J. (2006). Seeing an exercise as a single mathematical object: Using variation to structure sense-making. *Mathematics Thinking and Learning*, *8*(2), 91–111. https://doi.org/10.1207/s15327833mtl0802_1

Watson, A., & Thompson, D. R. (2015). Design issues related to text-based tasks. In A. Watson & M. Ohtani (Eds.), *Task design in mathematics education* (pp. 143–190). Springer. https://doi.org/10.1007/978-3-319-09629-2_5

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

Watson, J. M. (2013). *Statistical literacy at school: Growth and goals*. Routledge.

Watson, J. (2014). What is typical for different kinds of data. *Australian Mathematics Teacher*, *70*(2), 33–40.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46. https://doi.org/10.52041/serj.v2i2.553

Watson, J., & Callingham, R. (2013). PCK and average. In V. Steinle, L. Ball & C. Bardini (Eds.), *Mathematics Education: Yesterday, today and tomorrow, Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia*. Melbourne, VIC: MERGA.

Watson, J., & Fitzallen, N. (2010). *The development of graph understanding in the mathematics curriculum*. State of New South Wales through the Department of Education and Training.

Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, *2*(1–2), 11–50. https://doi.org/10.1207/S15327833TL0202_2

Weiland, T. (2016). Towards a framework for a critical statistical literacy in high school mathematics. In M.B. Wood, E. E. Turner, M. Civil & J.A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. The University of Arizona.

Weiland, T. (2019a). Critical mathematics education and statistics education: Possibilities for transforming the school mathematics curriculum. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and Trends in Current Statistics Education Research: International Perspectives* (pp. 391–411). Springer. https://doi.org/10.1007/978-3-030-03472-6_18

Weiland, T. (2019b). The contextualized situations constructed for the use of statistics by school mathematics textbooks. *Statistics Education Research Journal*, *18*(2), 18–38. https://doi.org/10.52041/serj.v18i2.138

Wells, H.G. (1903). *Mankind in the making: by H.G. Wells*. Leipzig: Tauchnitz.

Weninger, C. (2018). Textbook analysis. In C.A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics.* Hoboken, NJ: Wiley & Sons. https://doi.org/10.1002/9781405198431.wbeal1489

White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library trends*, *55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053

Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics*, *89*(1), 41–65. https://doi.org/10.1007/s10649-015-9595-1

Wild, C. J. (2017). Statistical literacy as the earth moves. *Statistics Education Research Journal*, *16*(1), 31–37. https://doi.org/10.52041/serj.v16i1.211

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–248. http://www.jstor.org/stable/1403699

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics?. In D. Ben-Zvi, K. Makar & J. Garfield (Eds). *International handbook of research in statistics education* (pp. 5–36). Springer. https://doi.org/10.1007/978-3-319-66195-7_1

Zapata-Cardona, L., & Marrugo Escobar, L. M. (2019). Critical citizenship in Colombian statistics textbooks. In G. Burrill & D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research. International perspectives* (pp. 373–389). Springer. https://doi.org/10.1007/978-3-030-03472-6_17

Zapata-Cardona, L., & Martínez-Castro, C. A. (2021). Statistical modeling in teacher education. *Mathematical Thinking and Learning*, 1–15. https://doi.org/10.1080/10986065.2021.1922859

Zawojewski, J. S., & Shaughnessy, J. M. (2000). Take time for action: Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School*, *5*(7), 436–440. https://doi.org/10.5951/mtms.5.7.0436

Appendix

Appendix I: Information to prospective teachers

Hej
Du har fått den här enkäten eftersom du utbildar dig till lärare för åk 4-6 i matematikämnets didaktik.
Frågorna i enkäten handlar om statistik och mer specifikt lägesmåtten medelvärde, median och typvärde.
När du svarar på enkäten kommer du att vara anonym då enkätens anonymitetsfunktion används.

Tack för din medverkan!
Har du frågor får du gärna höra av dig till mig!

Karin Landtblom
doktorand och universitetsadjunkt
MND, Stockholms universitet
karin.landtblom@mnd.su.se

Tillägg: Studenterna blev kontaktade via mail med information. De godkände sitt deltagande genom att öppna länken till enkäten och frivilligt svara på frågorna. Anonymitet försäkrades.

Appendix II: Information to principals for approval of teachers' participation

Stockholms universitet

Informationsbrev om forskningsstudie

Mitt namn är Karin Landtblom och jag är doktorand på Stockholms universitet. Efter påsk genomför jag en enkätundersökning med lärare som undervisar i matematik i årskurs 4-6 och vill härmed fråga om jag får inkludera lärare från er skola.

Syftet med min forskningsstudie är att skapa en bild av elevers, lärares och lärarstudenters uppfattningar om lägesmåtten medelvärde, median och typvärde med inriktning mot årskurserna 4-6. Statistik är ett viktigt område inom matematiken men är eftersatt vad det gäller forskning. Därför är det viktigt att forskning genomförs och att resultaten kommer lärarutbildningen till del. Denna undersökning kommer att bli en del av mitt avhandlingsarbete.

Datainsamlingen kommer att ske via en digital enkät som fylls i av lärarna. Lärarnas deltagande i undersökningen är helt frivilligt och läraren kan när som helst avbryta sitt deltagande utan närmare motivering. En hög svarsfrekvens är viktig för studiens tillförlitlighet och vi är därför tacksamma för om era lärare kan delta.

Inga personuppgifter samlas in. Anonymitet garanteras; enkäten innehåller varken namn, nummer eller annan möjlighet till identifiering. I rapportering av resultat kommer varken ort eller skola att nämnas. Vi kommer att följa de etiska regler som anges av Vetenskapsrådet genom Codex (se codex.vr.se för mer information). Endast jag, mina handledare och examinatorer kommer att ha tillgång till materialet.

Du som skolledare tillfrågas härmed om lärare i årskurs 4-6 från din skola kan delta i denna undersökning.

Ytterligare upplysningar kan fås antingen via e-post eller via telefon.

Tack på förhand!

Stockholm 11/4 2017

Mvh

Karin Landtblom
Universitetsadjunkt/doktorand

Tel: 0812076616, Karin.landtblom@mnd.su.se

Handledare:

Lovisa Sumpter
Lovisa.sumpter@mnd.su.se

Appendix III: Information to principals for approval of students' participation

Stockholms universitet

Informationsbrev om forskningsstudie

Mitt namn är Karin Landtblom och jag är doktorand på Stockholms universitet. Efter påsk genomför jag en enkätundersökning med elever i årskurs 6 och vill härmed fråga om jag får inkludera elever från er skola.

Syftet med min forskningsstudie är att skapa en bild av elevers, lärares och lärarstudenters uppfattningar om lägesmåtten medelvärde, median och typvärde med inriktning mot årskurserna 4-6. Statistik är ett viktigt område inom matematiken men är eftersatt vad det gäller forskning. Därför är det viktigt att forskning genomförs och att resultaten kommer lärarutbildningen till del. Denna undersökning kommer att bli en del av mitt avhandlingsarbete.

Datainsamlingen kommer att ske via en digital enkät och fylls i av eleverna själva. Frågorna handlar om hur eleverna förklarar de olika lägesmåtten samt hur de beskriver användningsmöjligheter för dessa. Undersökningen sker i klassrumsmiljö under en matematiklektion och beräknas ta 30 minuter i anspråk. Elevernas deltagande i undersökningen är helt frivilligt och eleven kan när som helst avbryta sitt deltagande utan närmare motivering. En hög svarsfrekvens är viktig för studiens tillförlitlighet och vi är därför tacksamma för om era elever kan delta.

Inga personuppgifter samlas in. Anonymitet garanteras; enkäten innehåller varken namn, nummer eller annan möjlighet till identifiering. I rapportering av resultat kommer varken ort eller skola att nämnas. Vi kommer att följa de etiska regler som anges av Vetenskapsrådet genom Codex (se codex.vr.se för mer information). Endast jag, mina handledare och examinatorer kommer att ha tillgång till materialet.

Du som skolledare tillfrågas härmed om elever i årskurs 6 från din skola kan delta i denna undersökning.

Ytterligare upplysningar kan fås antingen via e-post eller via telefon.

Tack på förhand!

Stockholm 11/4 2017

Mvh

Karin Landtblom
Universitetsadjunkt/doktorand

Handledare:

Lovisa Sumpter
Lovisa.sumpter@mnd.su.se

Appendix IV: Consent, parents and caretakers

Stockholms universitet

Informationsbrev om forskningsstudie gällande begreppsuppfattning om lägesmått

Du tillfrågas härmed om Ditt barns deltagande i denna undersökning.

Mitt namn är Karin Landtblom och jag är doktorand på Stockholms universitet. Syftet med min forskningsstudie är att skapa en bild av elevers, lärares och lärarstudenters uppfattningar om lägesmåtten medelvärde, median och typvärde med inriktning mot årskurserna 4-6. Statistik är ett viktigt område inom matematiken men är eftersatt vad det gäller forskning. Därför är det viktigt att forskning genomförs och att resultaten kommer lärarutbildningen till del. Denna undersökning kommer att bli en del av mitt avhandlingsarbete.

Datainsamlingen kommer att ske via en digital enkät och fylls i av eleverna själva. Frågorna handlar om hur eleverna förklarar de olika lägesmåtten samt hur de beskriver användningsmöjligheter för dessa. Undersökningen sker i klassrumsmiljö under en matematiklektion och beräknas ta 30 minuter i anspråk. Ditt barns deltagande i undersökningen är helt frivilligt och ditt barn kan när som helst avbryta sitt deltagande utan närmare motivering. En hög svarsfrekvens är viktig för studiens tillförlitlighet och vi är därför tacksamma för om ditt barn kan delta.

Inga personuppgifter samlas in. Anonymitet garanteras; enkäten innehåller varken namn, nummer eller annan möjlighet till identifiering. I rapportering av resultat kommer varken ort eller skola att nämnas. Vi kommer att följa de etiska regler som anges av Vetenskapsrådet genom Codex (se codex.vr.se för mer information). Endast jag, mina handledare och examinatorer kommer att ha tillgång till materialet.

Du som vårdnadshavare tillfrågas härmed om ditt barn kan delta i denna undersökning. Det skulle vara mycket värdefullt för mitt avhandlingsarbete om ditt barn deltar. Ytterligare upplysningar kan fås antingen via e-post eller via telefon. Tack på förhand!

Stockholm 11/4 2017

Mvh

Karin Landtblom
Universitetsadjunkt/doktorand

Tel: 0812076616
Karin.landtblom@mnd.su.se

Handledare:

Lovisa Sumpter
Lovisa.sumpter@mnd.su.se

Barnets namn: ______________________________

Vårdnadshavares underskrift: ______________________________

Namnförtydligande: ______________________________