The development of middle school children's interest in statistical literacy

Colin Stuart Carmichael

M.Phil, M.Ed, B.Sc(hons)

Submitted in fulfilment of
the requirements for the Degree of
Doctor of Philosophy
University of Tasmania

November 2010

Certification of dissertation

I certify that the ideas, experimental work, results, analyses, software and conclusions reported in this dissertation are entirely my own effort, except where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

Colin Carmichael

Colin Stuart Carmichael 3/11/2010

Statement of authority of access

This thesis may be made available for loan and limited copying in accordance with the *Copyright Act 1968*.

Colin Carmichael

Colin Stuart Carmichael 3/11/2010

Statement of ethical conduct

The research associated with this thesis abides by the international and Australian codes on human experimentation as set out in the "National Statement of Ethical Conduct in Human Research (2007)" and interpreted by the Human Research Ethics Committee of the University.

Colin Carmichael

Colin Stuart Carmichael 3/11/2010

Abstract

The focus of the study is interest and its influence as a motivating factor on adolescent children. Interest has a pivotal role in determining the extent to which students choose to re-engage in learning material. The dissertation describes the development of an instrument that is suitable for measuring middle school children's interest in statistical literacy, which is an ability to interpret messages containing statistical elements.

The "Statistical Literacy Interest Measure" (SLIM) is based on theoretical models that are embedded in the motivational literature. From these models, a bank of items was written, reviewed, and tested on a pilot sample of Australian middle school children. Testing and selection of items was undertaken using the Rasch Rating Scale Model (Andrich, 1978). Based on the outcomes of this process, further development of items occurred and they were subsequently retested on a larger sample of Australian middle school students. As a result of the process, 16 self-descriptions were deemed to be suitable for inclusion in the instrument.

Students' responses to SLIM and the "Self-Efficacy for Statistical Literacy" (SESL) scale, a measure of students' self-efficacy also developed in the study, were used to generate interest and self-efficacy logit scores. A number of statistical models were applied to these scores, as well as achievement and demographic data that were also collected during the study.

The results of the study indicate that interpretations based on SLIM will be valid. The measure explained approximately two thirds of the variance in students' responses and reported satisfactory reliability coefficients. The placement of items on the one interest continuum confirmed that there is a meaningful hierarchy associated with the interest construct, in that it commences with the low levels of interest that are associated with task-mastery and increases up to those high levels of interest that are associated with a desire

to re-engage with the domain.

The modelling process confirmed that in a middle school context, students' self-competency beliefs were a strong predictor of their interest but that interest itself was not a strong predictor of achievement. The inclusion of some teacher and school-related variables in the models suggested that teachers and schools have a greater influence on students' achievement than on their interest.

Given the increased emphasis that statistics education now appears to have in the proposed Australian curriculum, SLIM is a timely addition to the repertoires of researchers seeking to explore the development of middle school students' statistical literacy.

Acknowledgements

The completion of this project would not have been possible without the generous support of the following people.

- My supervisors Rosemary Callingham, Ian Hay and Jane Watson, for their encouragement and patience.
- My wife Patricia and son Stephen for their continued support.
- The diligent teachers who allowed me to intrude upon their lessons and time in order to study their students.
- The students who participated in the study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This study broadly concerns the interest that children have towards learning. Unfortunately there are aspects of twenty-first century schooling, such as rewards, sanctions and evaluative grades, that dampen a child's interest in learning (Ryan & Deci, 2000b). A more recent aspect concerns the global trend towards a national school-testing regime, evident in the United States of America, the United Kingdom, and now Australia. Such testing is known to narrow the curriculum, in that schools and teachers emphasise the attainment of cognitive outcomes associated with these national tests (Thomas, 2005). The concern is that such an emphasis should not come at the expense of students' affective growth, where affect in this instance is regarded as "a broad rubric that refers to all things emotional" (Rosenberg, 1998, p. 247). There is the risk that students may be able to do certain tasks but they will not want to do them.

At a more specific level, the study describes the development of an instrument designed to measure middle school children's interest in statistical literacy. The level of a child's interest in the learning of a specific task or in the learning of a component of the curriculum is an important measure of his or her affective development. The use of a suitable instrument that provides a valid measure of children's interest is one way to assess their affective development and more importantly to assess the efficacy of learning programs that emerge from the syllabus documents.

The study is aligned with a larger teacher professional development research study, titled Statsmart, details of which can be found in Callingham and Watson (2007). Statsmart seeks, in part, to determine the influence of teacher professional development on middle school students' cognitive development in statistical literacy. The study thus seeks to explore students' affective development and in particular, the development of their interest in statistical literacy. The context for the study is therefore middle school statistical literacy development.

In this introductory chapter, the rationale for the study is explored further. In doing so, the discussion focusses on the need to encourage statistical literacy through programs that also impact upon the affective development of students, and in particular those in adolescence. The discussion then presents the research aims for the study and concludes with an overview of the dissertation.

1.1 Rationale

Scientific assertions that anthropogenic activity is causing global warming currently generate much public debate. Indeed the very presence of such warming is contended, with claims that "Earth's temperatures continue a chilling trend that began 11 years ago" (Murdock, 2009). Although the underlying science on global warming is undoubtedly complex and outside this study, the counterclaims commonly used by sceptics are of relevance. These counterclaims invariably fail to appreciate simple statistical concepts, including the nature of variation and sampling error. In order for any citizen to contribute in an informed way to such debates, he or she must have some knowledge of these statistical concepts. Indeed, Wallman (1993, p. 1) argued that "statistical questions suffuse the fabric of our society at almost all folds." Statistical literacy, the focus of this study, concerns the ability to interact with such messages in a meaningful way. It is considered to be an essential life skill, so much so that Rumsey (2002) has described it as "statistical citizenship." Despite the importance of this literacy as a key life-skill, reports suggest that given a choice, students at the university level will not take statistics courses unless they have to (Schield, 2004). Arguably this reticence to engage with statistical literacy is not restricted to university students and has resulted in reported skill shortages in the field of statistics (Trewin, 2005).

The reluctance of adults to engage with statistical content is likely to have its genesis in their educational experiences with the domain. Certainly in the mathematics domain, researchers have found that students' re-engagement, as measured through subject enrolment rates, is predicted by their interest and liking for mathematics, their previous mathematics achievement, their mathematics self-concept, and their perceptions regarding its usefulness and difficulty (McPhan, Morony, Pegg, Cooksey, & Lynch, 2008). Other reports suggest that students' affect, in this case their interest and liking for the domain, is the strongest predictor of their subsequent re-engagement (Wigfield, Tonks, & Eccles, 2004; Watt, 2005).

Given the importance of affect as a motivator for re-engagement with a domain and ultimately subject choice, its positive development in students is an important outcome. Unfortunately there is a dearth of research concerning the positive development of affect, including interest, in the statistical literacy context and especially during the key phase of adolescence. This is not the case in mathematics education, where there is a significant body of research relating to affect and its development (e.g. Goldin, 2002; McLeod, 1992; Schiefele & Csikszentmihalyi, 1995). Statistical literacy, as a domain of knowledge, is sufficiently different from mathematics to warrant a separate investigation. Statistical literacy should be acquired across the various secondary school subjects encountered by adolescents and not just in the mathematics curriculum. Middle-school students should encounter statistical concepts in subject domains such as the natural and social sciences.

Although research into middle school students' statistical literacy has noted the importance of affect (Watson, 2006), it has not yet explored its influence. The study, therefore, seeks to address this gap in the current research literature through the development of an instrument that assesses middle school students' interest in statistical literacy.

1.2 Research aim and objectives

Given the importance of statistical literacy as a key life-skill, and the role of interest in explaining student re-engagement with the domain, the broad aim of this study is to develop a valid measure of middle school students' interest in statistical literacy that can subsequently be used to explore their statistical literacy development. The specific objectives of study are:

1. To develop a measure of middle school students' interest in statistical literacy.
2. To validate this instrument against theoretical models of interest and in particular: internal, external and developmental models.
3. To use this instrument to explore both the antecedents and precedents of middle school students' interest in statistical literacy.

1.3 Dissertation Outline

In Chapter 2 the context of this study is described, that is statistical literacy development during adolescence. The discussion in the chapter commences with a description of those aspects of statistics, as a domain of knowledge, that differentiate it from mathematics. It then examines the concept of literacy and in particular statistical literacy, with a subsequent review of current models that describe students' development of this literacy. The discussion in the chapter then examines adolescent development. It argues that the formation of identity during this period and with it the establishment of individual interests, ensures that adolescence, as opposed to other key human phases of development, is important in the development of enduring interests.

Chapter 3 reviews the literature as it relates to interest. In particular the chapter commences with a review of the theoretical interest-based literature and describes both the significance of interest and how interest is thought to develop during adolescence. The chapter then reports a review of the empirical interest-based literature and in particular notes the absence of any interest-based studies in the current context. Based on these reviews, the discussion presents a model to describe the development of middle school students' interest in statistical literacy. The final section of the chapter details the specific research questions that are addressed in this study.

Chapter 4 describes the methodology used in the study. It provides details of the subjects who participated in the study, the instruments used, and the methods used to answer each of the study's research questions. Within this chapter a theoretical background to Rasch models is provided, as they form the basis for much of the analysis in this study. The discussion in the chapter concludes with a description of the specific procedures used to analyse the data, including: the pooling of data, and the treatment of outliers and missing values.

Chapter 5 reports the initial development of the study's instruments, the Statistical Literacy Interest Measure (SLIM), and the Self-Efficacy for Statistical Literacy (SESL) scale. The discussion commences by detailing the types of evidence necessary to establish the validity of interpretations that are to be made from these instruments. It develops theoretical models of interest and self-efficacy in statistical literacy that were subsequently used as the basis for item development. The discussion then reports the procedures that were used to develop these instruments, including the panelling of items and their subsequent piloting on a group of students from Queensland. It concludes by presenting preliminary validity evidence for the two proposed instruments.

The results of the study are reported in Chapter 6. Evidence related to the validity of the proposed instruments, but based on a pooled sample of students, is presented at the commencement of this chapter. The chapter then reports the results of the study as they relate to each of the specific research questions.

Chapter 7 provides a discussion of the study's results and addresses the study's research questions. In particular it commences with a review of the results of the study and then discusses the implications of these findings. It concludes by suggesting further research that could emanate from the findings of the study.

Chapter 2

Study context

The ability to ask the “right” questions about statistics, or more specifically about messages that contain statistical elements, is a critical aspect of statistical literacy. Given the proliferation of information that is the World Wide Web, such ability is becoming increasingly important. In this chapter the discussion focuses on the development of statistical literacy during adolescence, which is the context for the study.

Traditionally statistics is taught by mathematicians as a part of the mathematics syllabus, and the practice continues in secondary schools today (Holmes, 2003; National Curriculum Board, 2009). The discussion in the chapter commences with a description of statistics as a domain of knowledge. It describes the distinctive features of the domain and in particular those that differentiate it from mathematics. This distinction is particularly relevant in the study, especially as several studies have examined interest in a secondary mathematics context. The study, however, seeks to examine the development of middle school students’ interest in statistical literacy, rather than statistics per se. The discussion, therefore, continues with a description of statistical literacy and includes a review of models related to the development of this literacy during the middle school years. The review reveals that although current models describe the cognitive development of middle school students’ statistical literacy, there is gap in the literature related to their accompanying affective development.

Having described the knowledge domain at focus in the study, the discussion in the chapter concludes with a description of the middle school context. As is argued, this period of development is particularly relevant for the study, which examines interest development. In the first instance, it is a period when children establish their identities, and with these their individual interests. It is also the period when Australian students decide on subject choices for senior secondary school. Students' interest in a given subject domain is a key predictor of their desire to re-engage with that domain and hence their motivation to pursue further study or a career in that domain.

2.1 Statistics as a branch of knowledge

A statistic can be regarded colloquially as a number about something, in other words one that is associated with a particular context. Unfortunately statistics in any natural context display inherent variability, so much so that Charles Darwin (cited in Holmes, 2003, p. 439) expressed the hope that "the inaccuracies and uncertainties of the world will be recognized as one of its essential features." The domain of knowledge that has developed to accommodate this variability in data is known as statistics. It is broadly defined as "information gathering and information processing" (Rao, 1975, p. 152) and is "concerned with finding out about the real world by collecting, and then making sense of, data" (Wild, 1994, p. 164).

Due to the quantification of most data, statistics as a domain of knowledge deals to a large extent with numerical data, and consequently has a close connection with mathematics. Arguably, most of the general public would regard statistics as a sub-domain of mathematics. It is traditionally taught as a part of the mathematics school curriculum (Holmes, 2003) and at the tertiary level by mathematicians (Moore, 1988). How then does statistics as a domain of knowledge differ from mathematics?

Moore (1988) in an essay titled "Should mathematicians teach statistics?" argued that the two knowledge domains were sufficiently different to answer no to this question. Statistics educators have identified a number of such differences. In the first instance, statistics as a domain of knowledge originated from the study of census data and subsequent major developments have occurred in non-mathematical domains such as agriculture and the social-sciences (Moore, 1988). Compared to mathematics, statistics is a relative late-comer. As Moore and Cobb (2000, p. 261) then argued, "it was coalesced in this century from beginnings in many fields."

Statistics is "a methodological discipline rather than a core substantive area" (Moore & Cobb, 2000, p. 620). As a result context is paramount in statistics. It provides meaning in statistics, yet obscures structure in mathematics (Cobb & Moore, 1997). The major aims of statistics deal with the inherent variability of data (Rao, 1975). Consequently statistical investigations are quite distinct from their mathematical counterparts. The former often result in an opinion that is supported by the data, whereas the latter typically result in a solution (Garfield, 2003).

Given the apparent differences between statistics and mathematics, the extent to which mathematics education research is applicable to statistics education is of particular relevance for this study, especially given the emphasis that is now placed on statistics education in Australia, where it is one of only three content strands in the proposed national mathematics curriculum (National Curriculum Board, 2009) . This emphasis is intended to enable school leavers to "comprehend, interpret, and critically evaluate messages with statistical elements" (Gal, 2003, p. 80). Such a facility with statistics embedded in messages is termed "statistical literacy" and is discussed in some depth in the next section.

2.2 Statistical literacy

The Oxford dictionary defines literacy as "an ability to read and write" (Coulson, 1969, p. 311). Thus the term "statistical literacy" suggests an ability to interpret statistical messages and where necessary communicate such messages using the written or spoken word. Such a view although not incorrect is too narrow. Ramdas (1990, p. 31) argued that “literacy is to be conceived of as a political, human and cultural process of consciousness raising and liberation.” Within this paradigm, statistical literacy becomes more than an outcome: It becomes an enabling process and one of several “multiliteracies” (Lo Bianco, 2000) that have arisen in response to globalisation. This enabling process then includes opportunities for learners to engage with data, as well as opportunities for them to interact with statistical messages. Hence the term statistical literacy, as conceptualized in this study, encompasses aspects of “doing” as well as communicating statistics. Overarching this doing and communicating of statistics, is the ongoing development of positive affect towards statistics as a domain of knowledge.

A statistically literate person, the outcome of statistical literacy, is one who has the ability to understand and critically evaluate the statistical messages that permeate daily life, together with an appreciation of the contributions that statistical thinking can make in decision making processes (Wallman, 1993). Such ability lies on a continuum, with some people having a much greater insight into statistical concepts than others. Gal (2002) discussed the concept of “functional” as opposed to “true” statistical literacy. A functionally literate person should be familiar with concepts and be able to communicate them; a truly literate person would also have a deep understanding of underlying theories and concepts. Such an understanding is in some texts termed “statistical reasoning” (Ben-Zvi & Garfield, 2004) and regarded as being at a cognitively higher level than statistical literacy. In this study, it is assumed that statistical literacy requires some statistical reasoning and “statistical thinking” (Chance, 2002). Given the middle school context, however, the primary focus in this study is on the development of functional literacy, which presumes an ability level at which consumers as opposed to producers of statistics can successfully operate.

Gal (2002) outlined a model of statistical literacy that included both knowledge and dispositional elements. Dispositional features in his model included: a critical stance, which he defined as a propensity to question messages of a quantitative nature; and the necessary beliefs and attitudes to support such a stance. The knowledge elements of Gal's model included: the ability to read and interpret text, knowledge of statistical processes and terminology, a facility with mathematics, knowledge of the associated context, and an ability to ask the right questions regarding the data or the message. In her model of statistical literacy, Watson (2006) concurred with the elements identified by Gal but regarded a knowledge of variation to be so fundamental to statistical literacy as to warrant its inclusion as a separate element in the model. Watson also regarded an ability to work with differently formatted tasks as essential for statistical literacy, although this may be of more importance in the school context, which is the focus of Watson's research. Watson too acknowledged the importance of dispositions, including scepticism, curiosity and imagination, for the positive development of statistical literacy.

In regard to the cognitive demands of the statistical literacy domain, Watson (1997) proposed that a person becomes statistically literate through a cyclic process that encompasses three broad stages. In the first, the student must be familiar with terminology used in the everyday reporting of statistics. The interpretation of such terms in a variety of contexts is a next necessary step towards statistical literacy. Finally, the student must be able to question the reports of others critically. More recently, and based on quantitative methods involving a large group of school students, Watson and Callingham (2003) identified six hierarchical stages associated with statistical literacy.

1. Idiosyncratic. In the early stage students are unable to engage with the context and their responses to simple tasks are typically idiosyncratic.

2. Informal. In the second stage, students demonstrate little engagement with context and their responses to tasks are typically "unistructural"

(Biggs & Collis, 1982) in that only one aspect of the task is addressed.

3. Inconsistent. In the third stage students demonstrate some engagement with the context, however this is inconsistent. Their responses are typically "multistructural" (Biggs & Collis, 1982) in that two or more aspects of the task are considered.

4. Consistent (non-critical). In the fourth stage, students demonstrate consistent engagement with context but are unable to appraise the work of others critically.

5. Critical. In the fifth stage, students are able to engage in the context and criticize the reports of others provided such criticism does not rely on proportional reasoning.

6. Critical (mathematical). In the final stage students are able to engage in the context and critically analyze statistical reports including through the use of proportional reasoning.

The content domain on which these levels are based included a consideration of: context, data collection, data representation, data reduction, probability, inference, variation, and, mathematical and statistical skills.

The models of statistical literacy discussed in this section acknowledge the importance of affect. Both Watson and Gal have included dispositional elements in their respective models. Yet, in a sense, both researchers have paid lip-service to the influence of affect, with Watson and Callingham's detailed model of the statistical literacy hierarchy examining only cognitive outcomes. The research reported in the dissertation seeks to address this short-coming of current statistical literacy models through an investigation of middle school students' interest in statistical literacy.

2.3 Middle-school education

In the Australian context, the concept of a middle school as a distinct entity is unusual. In the study, therefore, the term middle school is used to encompass that developmental period in children that includes puberty, and with it the physical and emotional changes that this brings. Consequently the ages of students in the "middle school" typically range from 11 to 15 years. The following discussion examines the significance of the middle school period, as a human developmental phase.

Adolescence is one of the key phases of human development yet it coincides with low levels of affect for learning. Several authors agree that one of the main tasks of adolescence is that of identity formation (Hay & Ashman, 2003; Low & Rounds, 2007; Oyserman, 2004). It is perhaps because the adolescent is so preoccupied with his or her identity that several longitudinal studies report declines in students' affect for learning during the middle school period (Dotterer, McHale, & Crouter, 2009; Fredricks & Eccles, 2002; Watt, 2004, 2008). In a mathematics education context, Fredricks and Eccles (2002) reported a steady decline in levels of interest over the entire period that students attend school. Also in a mathematics context, Watt (2008) reported similar findings, yet found that the greatest fall in intrinsic valuing for mathematics occurred during Year 7, which in her study coincided with the first year of high school. Based on academic interests in general, Dotterer et al. (2009) reported that levels of interest in learning reach a minimum at the age of 16 and previous declines in interest are more pronounced for boys than for girls.

Adolescence is a human developmental period in which affect plays a relatively pronounced role. Wigfield, Byrnes, and Eccles (2006) cited evidence that suggest physical changes to adolescents' brains are likely to result in more affective activity during this period. In addition to this, a range of evidence suggests that students' emotional stability increases during adolescence. In their study of 220 students from Years 5 to 12, Larson, Moneta, Richards, and Wilson (2002) reported that younger students show a greater variability in their emotions than older students. As another example, Köller, Baumert, and Schnabel (2001) conducted a longitudinal study of 602 students from Years 7 to 12 and concluded that junior secondary students were more sensitive to achievement feedback than their older peers. These results suggest that affect is more prominent for students in early adolescence, than in later adolescence.

The evidence suggests that the role of affect as a motivator for learning is also dominant during adolescence. In their meta-analysis of 113 studies across the entire school period, Ma and Kishor (1997) reported that the highest correlation between attitudes for mathematics and achievement in mathematics occurred for students in Years 7 to 9, although in a later study Ma and Xu (2004) reported that it occurred for students in Years 9 and 10. Consequently the influence of affect on learning appears to be more pronounced for students in the middle school than for those in earlier and later developmental periods, although there is the suggestion that this influence stabilizes somewhat before the onset of adolescence (Marcoulides, Gottfried, Gottfried, & Oliver, 2008).

The middle school is also a period when students are required to consider their subject choices for senior secondary school. Ma (2006) found that the change in attitude towards mathematics during lower secondary school was the strongest predictor of subsequent choice of mathematics courses in the senior secondary school. McPhan et al. (2008) reported a similar finding in the Australian context. Such results align with research based on the Expectancy-Value (EV) model of learning (Wigfield & Eccles, 2000) that suggests students' valuing of a learning task is predictive of their desire to re-engage, whereas their expectancy of success is predictive of their actual performance. These results are confirmed empirically with Köller et al. (2001) reporting that although students' interest does not predict their achievement in mathematics it does predict their choice of mathematics course.

2.4 Chapter summary

In this chapter the discussion focussed on the study context, in particular the concept of statistical literacy development in a middle school. The discussion commenced with a review of the differences between mathematics and statistics as distinct domains of knowledge. It was noted that although statistics relies on mathematical procedures, it is a methodological subject for which context is important. The discussion then examined the concept of statistical literacy, regarded as an enabling process allowing students the opportunity to interact in a meaningful way with messages containing statistical elements.

Models of statistical literacy were also reviewed in the chapter, specifically those proposed by Gal (2002) and Watson (2006). Although both of these models acknowledge the importance of dispositional elements, it was noted that research into the development of statistical literacy in children has focussed primarily on their cognitive growth. Consequently there is a significant gap in the literature as it relates to the development of middle school students' statistical literacy.

In the last section of the chapter the discussion highlighted the importance of affective development during adolescence and thus the middle school period. The research cited in Section 2.3 points to generally low levels of affect for learning in the middle school, yet heightened affective development. This research also suggests that younger adolescents are more likely to vary their affective state than older adolescents: They are less emotionally stable. In addition to this, correlations between students' attitudes towards mathematics and their achievement in mathematics suggest that the influence of affect on learning is strongest during adolescence. The middle school period, which encompasses early adolescence, is thus painted as particularly important to a study aimed at exploring affect. In addition to this, and probably of greater relevance to the rationale for this study, it was noted that early adolescence is particularly important because it is at the conclusion of this phase in life that students choose their senior school subjects, with such choices being governed by their affect for learning.

The discussion thus far has examined the influence of affect, although occasionally elements of affect including attitudes, emotions, values and interest have been mentioned. The study, however, specifically concerns interest, which is a key element of affect. The discussion in the next chapter, therefore, explores the concept of interest. It provides a review of both the theoretical and empirical interest literature and in doing so develops the theoretical basis for the study.

Chapter 3

Interest and learning

The term interest is widely used; however its exact nature is seldom explicated. The discussion in this chapter seeks to address this contradiction. It commences with a review of interest-based theories, which is used to describe the complex nature of interest and the processes that influence its positive development. The review then describes how interest influences learning. In particular it describes the Model of Domain Learning (Alexander, Jetton, & Kulikowich, 1995), a learning model that includes interest as the key motivational construct.

Having reviewed the theoretical basis for interest development the discussion then reviews empirical evidence related to interest. Little research has actually investigated middle school children's interest in statistics, although one study has examined senior secondary school students' interest in statistics. As a result, the review examines related research in the secondary school mathematics and tertiary statistics contexts. The review concludes by identifying specific factors that should influence the development of middle school students' interest in statistical literacy. As part of the review, a theoretical model of interest development is proposed, which is subsequently used in the study to establish the external validity of the proposed interest measure. The specific research questions for the study are presented in the last section of the chapter.

3.1 Interest and interests

The Macquarie Dictionary defines interest as "the feeling of one whose attention or curiosity is particularly engaged by something" (Delbridge, Bernard, Blair, & Ramson, 1987, p. 910). The term itself is derived from the Latin *inter-esse*, which means "to be between." Dewey (1910, p. 91) argued that interest “marks the annihilation of the distance between the person and the materials and results of his action.”

In the study interest is regarded as an affect, which is considered to be hierarchical (Rosenberg, 1998) with emotional states at the bottom of the hierarchy and temperament traits at the top. These temperament traits are regarded as “stable predispositions toward certain types of emotional responding” (Rosenberg, 1998, p. 249) and are thought to have an organising role in the activation of the transient states, which include moods and emotions. The state/trait property of affect can accommodate the complexity inherent in descriptions of interest. At a state level, interest manifests itself in the feelings described in the dictionary definition, yet at a trait level it is strongly associated with the self, as alluded to by Dewey (1910) and argued more recently by Renninger (2009).

Interest is a positive affect; however, it is directed specifically towards an object that is termed the *object of interest.* The term *interests* as opposed to interest refers to a collection of such objects. At the trait level *individual interest* is described as “a person’s relatively enduring predisposition to re-engage particular content over time” (Hidi & Renninger, 2006, p. 113). Interest at the state level is more transitory and is typified by positive emotions. This state can be induced by aspects of the environment and in such instances is termed *situational interest* or it can be induced from the individual’s predisposition to engage with the object and in such instances is termed *actualized interest.*

In the next two sections the trait and state like properties of interest are discussed further and in particular their influence as a motivator for student learning.

Interest as a trait

Individual interest is a close personal attachment to, or valuing of, an interest object. The value that an individual places on an interest object is distinct and far more important to the self than the utility of the object. As an example, many adults in western countries value their cars. For some people cars are indispensable: They are valued for their utility. For other people the replacement cost of the car is considerable, so the value placed on the car is related to its cost. There are some people, however, who are *interested* in cars. For these motor enthusiasts the car could be considered to be an extension of the self and thus it is highly valued. In such cases there is an emotional attachment to the car.

Many individuals have clearly defined interests; even very young children can have highly focussed interests (Hidi & Harackiewicz, 2000). Interests are regarded as important for psychological health, with Hunter and Csikszentmihalyi (2003) reporting a general malaise experienced by adolescents who do not have clearly defined interests.

Considerable research has explored the association between individual interest and learning. Schiefele (1991) found that controlling for ability and intelligence, individual interest could predict:

1. The type of learning, with higher levels of interest predicting deeper levels of text processing;

2. The use of learning strategies, with higher levels of interest predicting elaboration strategies rather than rehearsal strategies; and,

3. The quality of the learning experience, with higher levels of interest predicting higher levels of potency, intrinsic motivation and self-esteem.

Further, students with high levels of individual interest are more likely to enjoy their learning (Schiefele & Csikszentmihalyi, 1995), which is itself a desirable outcome. There is a statistically significant association between learning achievement and individual interest: In a meta-analysis of 121 studies involving school-aged children in a number of subject contexts, Schiefele, Krapp, and Winteler (1992) reported that the average correlation between the level of interest in the subject and achievement in the subject was $r = 0.31$ (the 95% confidence interval for the population correlation coefficient was 0.05 to 0.57). This association tended to be higher for males than for females.

Interest as a state

People who are actively engaged with an object of interest typically experience and display the emotion of interest, one of several fundamental human emotions (Izard, 1977). In some circumstances such people may even experience "flow" (Csikszentmihalyi, 2002), a state of such involvement that a person typically loses all sense of time. In a learning context the emotion of interest is often experienced with the emotion of enjoyment, so much so that some researchers regard the two emotions as synonymous (Schiefele, 1991; Marsh, Trautwein, Ludtke, Köller, & Baumert, 2005). Evidence suggests, however, that the two emotions are quite distinct (Reeve, 1989), with feelings of enjoyment emerging from successful encounters with learning tasks and feelings of interest emerging from interactions with novel learning tasks. Silvia (2001, p. 277) regarded the emotion of interest as being akin to the "lay usage of the word curiosity – a motivational state aimed at understanding." Such a view is taken in this study, where the interest emotion could be described as the "aha" factor: The emotion felt when curiosity has been satisfied.

Features of the learning task may evoke situational interest. In the mathematics context, for example, Mitchell (1993) identified puzzles, group work and computers as learning strategies that trigger situational interest. In the reading context, which may also be of importance in statistical literacy context, Schraw, Flowerday, and Lehman (2001) argued that a well written and coherent text would be more likely to create situational interest than one which was not well written. In some cases, just being with another person in the learning task can create situational interest in an otherwise disinteresting activity (Isaac, Sansone, & Smith, 1999). It is believed that students who repeatedly experience situational interest will come to value the interest object and thus acquire an individual interest in the object (Hidi & Renninger, 2006; Krapp, 2002, 2007; Mitchell, 1993). The development of interest, including that of situational interest into individual interest, is discussed in the next section.

The development of interest

Research on child development has shown a decline in interest in learning over the entire school period (Fredricks & Eccles, 2002) with evidence suggesting that more pronounced declines occur as students enter their teenage years (Watt, 2008). Krapp (2002) offered a number of explanations for this decline in interest. He argued that individual interests dominate a child's activities during early childhood but as the child ages he or she increasingly becomes aware of the interests of important others. At the age of approximately four years, for example, children become aware of sex stereotypical interest objects, such as dolls, and therefore consciously ignore some of these objects. Further, Krapp (2002) maintained that by early adolescence, children become more aware of their immediate social context and accordingly alter their interest structure. As an example, students may lose their interest in learning if such an interest is perceived to be incompatible with the interests of their peers. Ryan and Deci (2000b) attributed this general decline in interest in learning to the structure of western schools. They reported that the use of extrinsic motivators in schools, such as rewards, sanctions, and evaluative grades, ultimately reduces intrinsic motivation in students and hence their general interest in learning. Given this

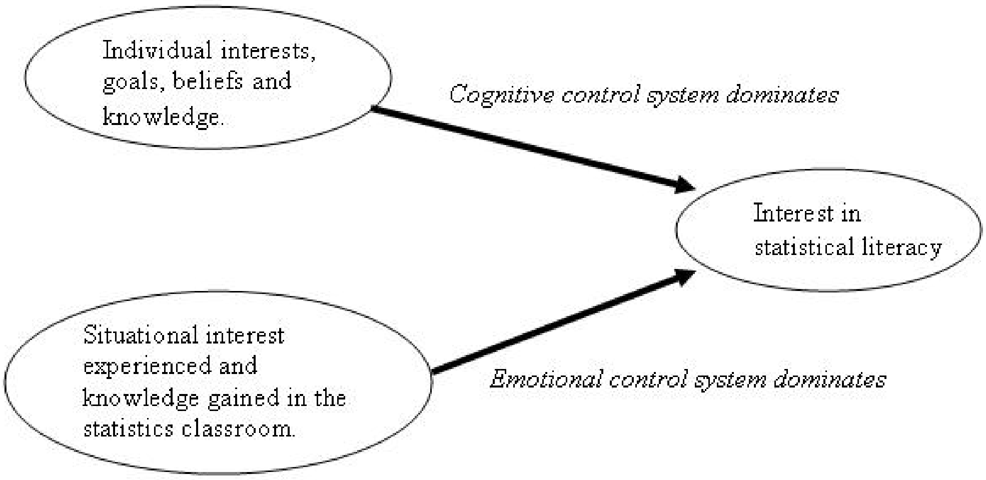


Figure 3.1. Suggested major paths to interest in statistical literacy

bleak picture of adolescent interest in learning it is imperative to explore ways in which an individual's interest can develop positively.

Positive interest development in statistical literacy may follow one of two, although not necessarily distinct, paths. As shown in Figure 3.1, the first path is the emergence of interest from individual interests, goals, beliefs and knowledge. On this path factors unique to the individual dominate the direction of the path. The second path is the emergence of individual interest from situational interest. On this path factors related to the situation dominate the direction. Krapp (2007) proposed that interest development is directed by two psychological control systems, the cognitive and the emotional. Along the first path, the cognitive control system dominates, so that a student consciously directs his or her attention to tasks that satisfy goals or perceived needs. Along the second path the emotional control system dominates, in that a student's need to experience positive emotions provides his or her motivation for engagement in tasks similar to those in which he or she has experienced interest (Pekrun, 2006). In a learning context, interest development will most likely follow both paths. The following discussion expands upon both of these paths.

In relation to the cognitive path, students may consciously direct their attention to a certain task at the expense of others in order to achieve personal goals. Such strategic choices may in fact lead to the development of interest. In a school context, students are often extrinsically motivated to engage in the various learning activities that they encounter. Ryan and Deci (2000a) proposed a hierarchy of motivation with such extrinsic motivation at its lowest levels. Within the mid-levels of this hierarchy students may be motivated because they see the activity as personally important or useful. At this level students' motivation is internalized to the extent that no obvious external motivating factors are present, yet their motivation is not intrinsic as they seek to satisfy external goals: They are motivated by "impure interest" (Dewey, 1910). Students at this level of motivation, however, may make the next step to truly intrinsic motivation and develop an interest in the task. Fox (1982), for example, reported a positive association between students' perception of the utility of mathematics and their interest in choosing mathematics related careers. In other words, if students see statistical literacy as important to them personally, then it is likely that their motivation to re-engage in related tasks will be internalized and that true interest may eventually follow. More recent research has demonstrated a positive association between students' mastery goal orientation – their reason for completing tasks relates primarily to one of task-mastery – and their interest in learning (Harackiewicz, Durik, Barron, Linnenbrink-Garcia, & Tauer, 2008; Hulleman, Durik, Schweigert, & Harackiewicz, 2008; Pekrun, Elliot, & Maier, 2009).

Students may also consciously choose tasks that align with their current personal interests. Krapp (2002) identified three models to explain how interest may emerge from current interests; these models tacitly acknowledge the close link between interest and knowledge.

1. Growth model. As an example, an individual may initially be interested

in mathematics. As his or her subject knowledge increases he or she becomes aware of specific interesting details encountered during mathematics, say chance and data. Further knowledge in these areas may reveal new interests in displaying data or calculating chance. The interest in mathematics and thus statistical literacy grows as the subject knowledge becomes more differentiated.

2. Channelling model. As an example, an individual initially interested in mathematics, might develop an interest in one aspect of mathematics (say probability) so that this becomes his or her main interest. In other words, his or her interest is channelled into new areas as knowledge differentiation occurs.

3. Overlap model. As an example, an individual may initially be interested in mathematics; he or she then may develop an interest in computers. An overlap between these two subjects may be the use of computers to analyse data and consequently the person may direct his or her interests towards this specific aspect.

In relation to the emotional path for interest development, the repeated experience of situational interest should lead to a more enduring individual interest. Mitchell (1993), for example, was able to demonstrate that under certain circumstances interest that was "caught" in learning environments high in situational interest could be maintained. As individuals are motivated to engage in tasks in which they are likely to experience positive emotion, the emotional control system is likely to dominate interest development along this path. But this does not preclude the operation of the cognitive control system. Indeed Silvia (2001) proposed that interest development is essentially the result of the individual resolving the cognitive conflict that occurs when he or she interacts with the object of interest. Based on the work of Berlyne (1960), he argued that during the person-object interaction, incoming stimuli are collated with current personal information on the basis of a number of *collative variables* that are associated with the learner's response to the stimuli. These collative variables include: novelty, uncertainty, and complexity. During the person-object interaction, the learner will fail to engage in any significant way with stimuli that are considered routine, that is have low levels of novelty. Similarly the learner will fail to engage when the stimuli are too unknown or frightening: They contain high levels of novelty. Berlyne (1960) argued that for optimal levels of these variables a state of curiosity will be evoked that is characterized by high levels of arousal. In this state the learner will be motivated to resolve the conflict created by the particular collating variable. If this conflict cannot be resolved quickly, the learner will be motivated to persist with the object, even return to it at later times. Such persistence with the object may uncover further stimuli that in turn create a conflict in need of resolution. In such a way it is hypothesised that both knowledge and interest in the object will develop, with the learner losing interest in simple objects and pursuing those with more complex associated knowledge. Consequently it is believed that knowledge and interest are closely related: One cannot have interest without knowledge. Indeed, Renninger (2000) argued that knowledge was in fact a dimension of interest. Alexander (2003), on the other hand, regarded both as inter-related components of a model of learning.

The previous discussion has focussed on the development of individual interest. Such development is likely to occur as a result of existing individual factors and/or as a result of situational factors. Further, this development will be directed by both cognitive and emotional control systems. Irrespective of whether individual interest emerges from current interests or situational interest, students' knowledge of the domain will be intricately linked to their interest. As they become more interested in statistical literacy they must become more knowledgeable about concepts related to this literacy. The Model of Domain Learning (Alexander, 2003), which is discussed in the next section, is an interest-based model of learning that recognizes this close link between interest and knowledge development.

3.2 The Model of Domain Learning

Unlike other motivational models of learning, such as the Expectancy Value Model (Wigfield & Eccles, 2000), the Model of Domain Learning (MDL) examines learning from a developmental perspective. Rather than seeking to explain specific student learning behaviour, it describes the development of knowledge over a more sustained period of learning. The MDL has three major components: the knowledge that is acquired, the learning strategies that are employed during this knowledge acquisition, and the motivation behind a student's learning. The last is assessed through their interest in the domain. Although several empirical studies have confirmed the utility of the MDL (Alexander, Sperl, Buehl, Fives, & Chiu, 2004; Lawless & Kulikowich, 2006; Murphey & Alexander, 2002), most have been restricted to a tertiary context.

The MDL identifies three major stages through which knowledge acquisition occurs. During the *acclimation stage* knowledge is typically fragmented and incomplete, although students may have deep levels of knowledge on small areas of the domain. These novice learners typically rely on general learning strategies that are often applied inefficiently and inappropriately. Moreover during this stage they have little domain knowledge and rely on situational rather than individual interest. During the *competency stage*, however, learners' knowledge becomes deeper and broader. They use more topic specific learning strategies and use them more efficiently. During this stage they rely equally on individual and situational interest for motivation. In the *expertise stage*, knowledge is highly developed and coherent. Further, it is sufficiently broad and deep that experts are contributing to knowledge in the domain. Consequently, it is unlikely that few, if any, high school students would reach this stage of development (Alexander, 2003). Experts typically have high levels of individual interest in the domain and seldom rely on situational interest for motivation.

Within the context of statistical literacy development, Watson and Callingham (2003) identified the presence of six hierarchical stages of knowledge development that were described in Section 2.2. The development of this hierarchy, however, was based on studies involving school students. Given Alexander's assertion that few school students reach the expertise stage in any domain; it is likely that these six stages will correspond with the acclimation and competency stages of the MDL. Therefore it is expected that most middle school students will be heavily reliant on situational interest for their motivation. In fact even students near the top of the statistical literacy hierarchy are likely to be equally reliant on situational and individual interest.

3.3 Review of interest based studies

The review of the interest theories has identified two paths for the development of interest, one influenced by the individual and the other by the situation. In this section, a review of empirical educational studies is undertaken in order to identify the specific factors that feature on each of these paths. Accordingly, the review reported in this section seeks to answer the following question: What are the factors documented in the literature that may influence middle school students' interest in statistical literacy?

The literature review, reported in this section, was conducted in three phases commencing with a search on the specific question and then generalising the search to encompass interest development in secondary mathematics contexts and then the development of positive attitudes in the tertiary statistics context. This broadening of the review was a result of the dearth of research relating specifically to the middle school context. After retrieving relevant research articles from all phases of the search, a content analysis (Krippendorff, 1980) identified common outcomes related to the research question but also differences in the way the interest construct was operationalised.

Common outcomes

The identified themes, as shown in Table 3.1, suggest that factors contributing towards interest in statistical literacy can be broadly classified into those that are situational and those that are individual. The former include pedagogical strategies and aspects of the learning environment, whereas the latter include the prior experiences and beliefs of the learners.

Table 3.1

Common study themes

Factor	Details and studies involved
Situational factors:	
Pedagogical practices	• Influence interest in mathematics/statistics (Bikner-Ahsbahs, 2004; Mitchell, 1997; Mitchell & Gilson, 1997; Sciutto, 1999; Trautwein, Ludtke, Köller, Marsh, & Baumert, 2006).
	• Promote positive attitudes towards statistics (Allredge, Johnson, & Sanchez, 2006; D'Andrea & Waters, 2002; Leong, 2006).
Technological	• Technology-enhanced classrooms can promote positive attitudes towards statistics (Meletiou-Mavrotheris, Lee, & Fouladi, 2007; Schou, 2007; Suanpang, Petocz, & Kalceff, 2004).

Continued on next page

Factor	Details and studies involved
Social climate	• A positive social climate can promote positive attitudes towards statistics (Cobb & Hodge, 2002; Mvududu, 2003).
Teacher support	• Classroom management strategies and the views of significant others can promote interest in mathematics (Fox, 1982; Kunter, Baumert, & Köller, 2007).
Individual factors:	
Prior knowledge, self-concept and age	• Individual factors are associated with interest in mathematics (Fox, 1982; Köller et al. 2001; Lawless & Kulikowich, 2006; Lopez, Brown, Lent, & Gore, 1997; Marsh et al. 2005; Trautwein et al. 2006).
Prior knowledge, attributional and competency beliefs	• Individual factors are associated with positive attitudes for statistics (Budé et al. 2007; Carmona, 2004; Estrada, Batanero, Fortuny, & Diaz, 2005; Finney & Schraw, 2003; Perney & Ravid, 1990; Sorge & Schau, 2002) .

Situational factors that promote interest in statistics. Pedagogical practices, including the types of learning experiences that students encounter and the classroom management strategies used by their teachers, have been shown to promote interest. Several studies from the mathematics education literature provide supporting evidence (Trautwein et al., 2006; Mitchell, 1997; Mitchell & Gilson, 1997). Mitchell, for example, was able to provide some evidence to suggest that the individual interest of students in environments high in situational interest will increase in both a mathematics (Mitchell &

Gilson, 1997) and statistics (Mitchell, 1997) secondary school context.

In the statistics education context pedagogical strategies were shown to promote positive attitudes towards statistics and presumably situational interest in statistics. These include: using video clips that demonstrate real-life applications of statistics (Allredge et al., 2006); embedding statistical activities in stories (D'Andrea & Waters, 2002); and using real-life and person-based scenarios (Leong, 2006). There is some evidence, however, to suggest that pedagogical practices aimed at improving attitudes towards statistics, in fact promote attitudes to the particular class or teacher where the learning occurs. D'Andrea and Waters (2002) found that attitude improvements in their study were directed towards the statistics course and not towards the field of statistics.

The social climate of the learning environment also plays an important role in developing interest. In a mathematics education context, Bikner-Ahsbahs (2004) proposed that a type of interest, termed *situated collective interest*, can emerge in a group situation where one by one students become involved in an activity and come to value the activity. Through observations of children she was able to provide some evidence to support this theory. Also in a mathematics context, Kunter et al. (2007) demonstrated that students' interest is influenced by their evaluation of their teacher's classroom management strategies. In particular, interest is predicted by students' perceptions of the extent to which teachers clearly outline class rules and the extent to which teachers monitor their students' progress.

In the statistics education context, Cobb and Hodge (2002) reported that the social climate of the classroom contributes to the value that students place on statistics. Moreover, Mvududu (2003) found that aspects of a constructivist classroom, in particular personal relevance and student negotiation, are associated with positive attitudes towards the field of statistics.

Individual factors that promote interest in statistics. As shown in Table 3.1, several mathematics education studies demonstrated an association between a student's prior knowledge and his or her level of interest. Similar conclusions were reached in the statistics education context where several studies demonstrated an association between both prior mathematics and statistics achievement, and levels of attitudes towards statistics. The direction of this relationship has also been explored. In the mathematics education context, Köller et al. (2001) identified interest in early adolescence as a predictor of later interest but not achievement. In addition, they reported that although interest in Grade 7 does not predict achievement in Grade 10, interest in Grade 10 does predict achievement in Grade 12. The strength of the association between prior knowledge and interest is known to be influenced by the structure of the knowledge domain in question. Lawless and Kulikowich (2006) reported a stronger association for statistics than for psychology and argued that the former was a more structured knowledge domain.

Several studies also demonstrated a link between students' conceptions of their competency and their level of interest. Lopez et at. (1997) provided evidence to suggest that students' self-efficacy beliefs predict their interest in mathematics. Marsh et al. (2005) and Trautwein et al. (2006) both demonstrated the link between students' academic self-concept and interest in mathematics, with Trautwein et al. (2006) asserting that self-concept is a strong predictor of interest, which almost entirely mediates the influence of achievement and tracking, which is the grouping of students of similar ability. Moreover, they argued that this relationship is influenced by the frame of reference used by students to judge their competency: High achieving students in a group of even higher achieving students are likely to report low levels of interest in mathematics whereas low achieving students in a group of even lower achieving students are likely to report high levels of interest. In the statistics education context, competency-based beliefs are known to be associated with attitudes towards statistics (Finney & Schraw, 2003; Sorge & Schau, 2002). The nature of this relationship was explored by Tempelaar (2006) who observed a strong linear association between the cognitive competence and affect subscales of the "Survey of Attitudes Towards Statistics (SATS)" (Schau, Stevens, Dauphinee, & Del Vecchio, 1995). This result suggests that a strong relationship exists between competency-based beliefs and positive affect in the statistics education context: Students enjoy doing those tasks that they believe can be undertaken successfully.

Differences in the operationalisation of interest

In the mathematics education context, differences were evident regarding the operationalisation of the interest construct. The German studies (Köller et al., 2001; Kunter et al., 2007; Marsh et al., 2005; Trautwein et al., 2006) regarded interest as having both a value and an emotion component, with the former including the importance of the task and the latter the enjoyment of the task. The concept of importance, however, may assess the utility of the task, which is an extrinsic motivator. As discussed in Section 3.1, such importance does not reflect interest, although it may emerge into interest. Other studies operationalised interest through asking students to indicate their level of interest in a given task (Lawless & Kulikowich, 2006; Lopez et al., 1997; Sciutto, 1996). Student assessments of importance and interest may be different.

Discussion of review

Self-determination Theory (Deci & Ryan, 1985) provides a unifying framework for interest or attitudinal studies, such as those reported in this section. Deci (1992) argued that a person experiences interest when he or she encounters novel activities in a context that allows for the satisfaction of his or her basic psychological needs, that is, competence, autonomy and social-relatedness. In a middle school context, a student's need for autonomy and social-relatedness can be met if aspects of the classroom environment are conducive. The content analysis reported in this section identified the social climate as a factor that positively influences students' attitudes. Mvududu (2003), for example, reported a statistically significant association between student negotiation and positive attitudes towards statistics ($r = .25$). A student's need for competence in statistical literacy, however, can be met if he or she possesses the necessary individual factors, that is, a sufficient knowledge of statistical literacy and positive competency-based beliefs regarding his or her ability to acquire statistical literacy. Prior knowledge and competency-based beliefs are identified individual factors that contribute to interest and/or positive attitudes.

Overarching the meeting of basic psychological needs is the requirement that in order for students to experience interest they must encounter novel activities. Pedagogical strategies were identified that contributed positively to both interest and attitudes. The extent to which these strategies utilized novel activities, however, is unclear. In his study of interest development, Mitchell (1997) utilized learning activities that were meaningful to students and which encouraged their involvement. Arguably true involvement comes from collative sources that include novelty. In the statistics education context, Allredge et al. (2006), D'Andrea and Waters (2002), and Leong (2006) provided students with familiar contexts and reported positive changes in attitudes. Given that novelty is a requirement for interest, yet familiarity appears to promote positive attitudes, the use or otherwise of novel activities is perhaps the point at which interest development as compared to attitude development differ.

The review has established a significant gap in the literature. Of the studies cited in the review, several examined interest but in a mathematics education context, and a large proportion examined positive affect, but in a tertiary statistics context. Only one study (Mitchell, 1997) examined the concept of interest in statistics in a senior secondary school context. Despite the dearth of research, there is a significant body of research available in both the secondary mathematics and tertiary statistics contexts. The content review of the material suggests that individual factors including prior knowledge and competency based beliefs contribute to interest development. Further, aspects relating to the learning environment also contribute to this development. In relation to the learning environment, it is argued that factors relating to the classroom teacher, who is the principal architect, feature most prominently in students' interest development.

The conclusions of the research review are based primarily on related contexts. Whereas it is acknowledged that research findings from both the secondary mathematics and tertiary statistics contexts are relevant to this study, the degree of relevance is in question. Certainly middle school children in Australia are introduced to most statistical concepts in their mathematics classes (National Curriculum Board, 2009). Yet there is a trend in Australian mathematics syllabi, from a computational formula-driven approach to the learning of statistics, to a more practical data-oriented approach (Watson, 2006). It is in a climate, where children are able to play with data, that possibilities exist for a divergence of mathematics and statistics related research. Similarly, the way students' interest in statistics develops is likely to depend on their age. Adolescents are prone to greater variation in emotions than adults and as a result are likely to become more excited with interesting activities, but increasingly bored with mundane activities. Such differences in emotional stability imply that adolescents may be more susceptible to changing and increasing their interest than adults.

3.4 A model of interest development

The review of the empirical literature, reported in the previous section, confirms the earlier discussion that the development of middle school students' interest in statistical literacy should be influenced by factors related to both the individual and the situation. In relation to the individual, the literature review confirms the MDL's prediction that prior knowledge should be closely associated with interest. The review also highlights the importance of students' self-competency beliefs on interest development. In relation to situational factors, the literature review has highlighted the importance of good pedagogical practice and social interaction on interest development. Such situational factors are ultimately related to the classroom teacher. The discussion in this section builds upon the earlier sections in the chapter and proposes a model of interest development that is specific to the middle school statistical literacy context.

In a middle school context, it is hypothesised that students' interest in statistical literacy will be influenced by their self-efficacy beliefs and their prior knowledge. These relationships, in turn, will be mediated by a number of teacher and individual factors. The proposed inter-relationships between these factors are described below.

Self-efficacy beliefs

Students' beliefs regarding their competency can be assessed through a construct termed self-efficacy, which is defined as "beliefs in one's capabilities to organise and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). This construct is future orientated (Bong & Skaalvik, 2003) and is typically assessed through items that ask students to indicate their level of confidence in achieving specific rather than general tasks. Of all the psychosocial factors, self-efficacy is considered to be one of the best predictors of achievement in an educational context (Robbins et al., 2004). Consequently it is likely that self-efficacy will provide more insight into middle school students' interest development than other measures of students' self-competency beliefs.

Silvia (2003) argued that the relationship between self-efficacy beliefs and interest is complex. On the basis that interest emerges from collative sources, he argued that students who are uncertain about their ability to complete a task will be more interested in that task. Silvia reasoned that this uncertainty will be present for intermediate levels of self-efficacy: Students low in self-efficacy are certain that they cannot do the task and those high in self-efficacy are certain that they can do the task. In a series of experiments Silvia demonstrated a quadratic link between self-efficacy and interest. Such a quadratic relationship may have been evident in a study by Gehlbach et al. (2008) who reported that increased levels of interest are associated with decreased levels of self-efficacy. Students were less self-efficacious with respect to novel and complex tasks, yet these tasks were of most interest. It is argued, therefore, that students' self-efficacy beliefs will be associated with their interest but not linearly.

Prior achievement

The MDL predicts that the development of interest in statistical literacy should coincide with the development of knowledge of related statistical concepts. Measures of students' prior achievement, in as much as they are valid measures of knowledge, should directly predict levels of interest in statistical literacy. Such measures, however, will also be used by students as they form their self-efficacy beliefs. Thus prior achievement, in as much as it is an accurate reflection of a student's ability, should also directly influence his or her self-efficacy beliefs.

Teacher factors

As discussed in the research review, a middle school student's need for autonomy and social-relatedness is very much constrained by the classroom environment, of which the teacher is the primary architect. If the needs of a student can be met from this environment, then the student may indicate *connectedness* with the particular mathematics or statistics classroom.

Connectedness to the classroom can be regarded as a subset of school connectedness, a concept that has been extensively researched (Townsend & McWhirter, 2005; Whitlock, 2006). McNeely and Falci (2004) argued that there are two major dimensions of school connectedness: teacher support and social belonging. They found that teacher support, rather than social belonging, influences pro-social behaviors: A perceived lack of teacher support correlates with delinquency behaviours. It is expected, therefore, that teacher support will have the greatest influence on achievement-based behaviours. Klem and Connell (2004) proposed three components of teacher support: involvement with the students, provision of autonomy for the students, and the provision of structure. In relation to the last, Kunter et al. (2007) reported that students' perception of rule clarity predicts later levels of subject interest. It is hypothesised, therefore, that teacher factors, of which perceived teacher support is primary, will directly influence students' interest. As social persuasions are known to influence students' self-efficacy beliefs (Usher & Pajeres, 2006), it is argued that teacher support will also influence students' interest indirectly through their self-efficacy beliefs.

Individual factors

Interest development occurs as the result of a number of factors specific to the individual. Some, such as their achievement in statistical literacy, and their self-efficacy in statistical literacy, have been discussed already. There are a number of other individual factors, however, that should influence interest development, although in some cases this influence may be indirect through students' prior achievement. These factors include their current personal interests, their goals, their age, their gender and the frame of reference they use to make self-assessments of interest.

The influence of personal interests and students' goals on interest has been discussed at length in Section 3.1. There is evidence to suggest that age is also a factor that influences reported interest levels of middle school students, although the direction of this influence is not altogether clear. As was reported in Section 2.3, there tends to be a decline in interest as students progress through the middle school. Yet, the MDL predicts an increase in students' interest in statistical literacy as they gain knowledge in the domain; presumably such knowledge increases as students progress through the middle school.

Gender is also likely to influence middle school students interest in statistical literacy, although it is unclear whether boys or girls will report higher average levels of interest. In their meta-analysis of interest based studies, Schiefele et al. (1992) reported that the association between achievement and interest is stronger for males than females. Such a difference could be attributed to broad gender differences in students' personal interests, which have been explored in a number of contexts. In a science education context, for example, Jenkins and Pell (2006) reported that girls are more likely to be interested in topics that deal with the self and the natural world, whereas boys are more interested in topics that deal with destructive technologies. In relation to career interest (Holland, 1985), Lubinski, Benbow, and Morelock (2000) reported that girls are significantly more likely to favour social and aesthetic careers, while boys are likely to favour economic and political careers. In a statistical literacy context, if teachers present the statistics in a context that generally conflicts with these gender differences, then it is possible that this will result in gender differences in students' levels of interest. In a learning context, gender differences in reported interest can also result from known gender stereotypes that are associated with the subject. Several authors report a gender stereotype associated with mathematics that associates high performance with males (Kiefer & Sekaquaptewa, 2007; Ryan & Ryan, 2005; Smith, Sansone, & White, 2007). In a study of undergraduate females, Smith et al. (2007) reported that women who are anxious about their performance in mathematics are more susceptible to the influence of gender stereotypes associated with the subject and as a result adopt performance avoidance goals and report lower levels of interest. It is unclear; however, whether mathematics related gender stereotypes apply in the statistical literacy context.

The frame of reference that students use to make self-assessments of interest may also vary from person to person and influence their ratings of interest. In order to explain somewhat contradictory results in mathematics and verbal achievement – strong within-domain associations between self-concept and achievement but zero or negative cross-domain associations – Marsh (1986) proposed that students make self-assessments of their achievement using two frames of reference. They either compare their achievement with others in their class, an external frame of reference, or they compare their achievement with their own achievement in other circumstances, an internal frame of reference. More recently, Goetz, Frenzel, Hall and Pekrun (2008) investigated the influence of the frame of reference on the relationship between enjoyment and achievement. They reported similar results to those of Marsh and colleagues: strong within domain associations between enjoyment and achievement but negative cross-domain associations. These results suggest that the frame of reference might be applicable to students' ratings of interest.

The preceding discussion identified key factors that it is hypothesised influence middle school students' interest in statistical literacy. These factors and their proposed inter-relationships are shown in Figure 3.2, where straight arrows represent linear effects and curved arrows represent quadratic effects.

3.5 Research questions

Given the study aims that were outlined in Section 1.2 and the discussion in this chapter, the research questions for the study can be listed as:

1. How valid is it to base a measure of middle school students' interest in

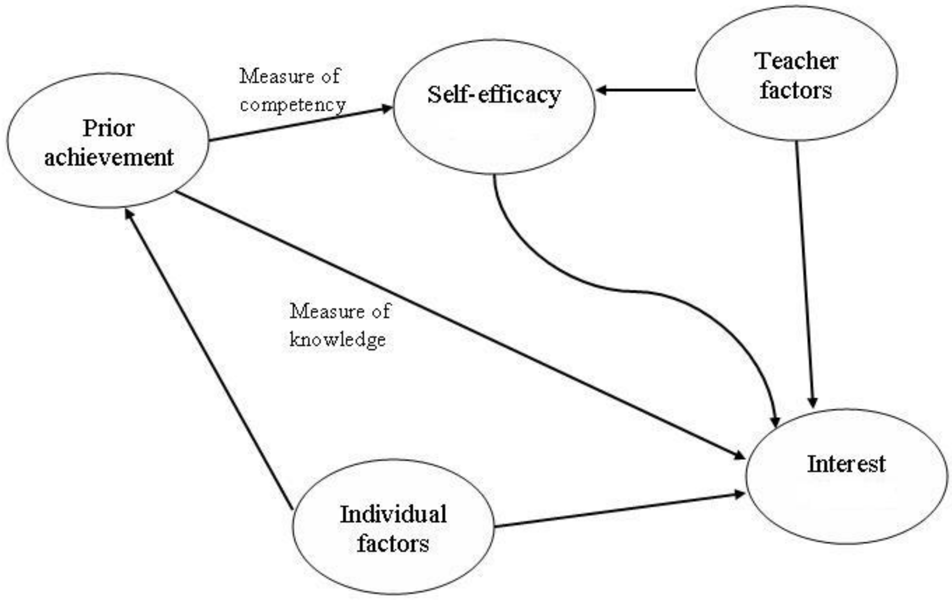


Figure 3.2. Hypothesised antecedents of interest in statistical literacy.

statistical literacy on their responses to a series of interest self-descriptions?

2. How do factors unique to an individual, such as their age, prior achievement, gender, and self-competency beliefs, contribute to their interest in statistical literacy?

    (a) To what extent do students' frames of reference influence their interest in statistical literacy?

    (b) To what extent do students differentiate between mathematics and statistics when making an interest assessment?

3. To what extent does students' interest in statistical literacy influence their subsequent achievement in statistical literacy?

3.6 Chapter summary

The discussion in the chapter commenced with a review of the nature of interest and theories associated with its development. These theories, in turn, predict that both individual and situational factors contribute to interest development. The Model of Domain Learning was presented as the major interest-based model of learning. Although this model appears to have been untested on adolescent learners, it predicts that their knowledge of statistical literacy should be closely associated with their interest in the domain. In particular, the model predicts that adolescent students are likely to rely on situational interest rather than individual interest as a source of motivation. Indeed they are likely to report low levels of individual interest. Due to the unique aspects of adolescent development, a review of the literature was conducted in order to identify additional factors that might influence the interest development of middle school students. The review confirmed that predictors of interest can broadly be classified as being related to either the individual or the situation. Moreover, the review also identified the absence of a suitable instrument for assessing such interest and the need to distinguish in such an instrument between students' perceptions of importance and interest.

Based on the results of the research review and also the theories of interest development that were presented in earlier sections, a proposed model of interest development was then presented. The model acknowledged the importance of individual factors, such as prior achievement and self-efficacy beliefs, on interest. It also recognized the importance of factors, such as teacher support, on interest. The model, along with the MDL, serve as construct models, used to assess validity issues relating to the proposed interest measure.

The discussion in the chapter revealed a significant gap in the literature as it relates to middle school students' interest in statistical literacy. Theoretical models such as the MDL appear not to have been tested in a secondary school context and the research review only identified one study that is closely related to this context. The proposed interest measure reported in this dissertation is therefore a timely addition to quantitative research in the area of middle school students' education. The methodology for developing the instrument is outlined in the next chapter.

Chapter 4

Study methodology

The methodology in the study is quantitative in that it seeks to assign a number to a given student's interest and then to use the number to make interpretations about his or her interest. The assignment of numbers to psychological constructs such as interest is problematic and the techniques for achieving this are complex. Rasch models can be used for developing valid measures in such situations and it is these models that form the foundation for the analysis of data in the study.

The discussion in this chapter describes the methodology used in the study. It commences with an outline of the methodology, which includes a statement detailing the researcher's ontological and epistemological stance. The discussion then reports on the subjects involved in the study, and in particular how they were selected. It describes the instruments that were used in the study and outlines the data that were collected. Following this, the discussion describes the methods of data analysis used to answer each of the research questions. Most analysis methods used in the study are situated in a Rasch measurement paradigm, consequently the discussion also provides a background to Rasch models and their use in scale development. The discussion concludes with a description of the data-analytic procedures used in the study. Details regarding the development of the study's instruments are reported in the following chapter, whereas results of the study relating to the specific research questions are reported in Chapter 6.

4.1 Outline of methodology.

Prior to providing an overview of the methodology it is relevant to discuss the ontological and epistemological stance taken by the researcher. Karl Popper

(1902-1994) asserted that there were three worlds of “things”: the objective world of material things, the subjective world of minds, and, the world of ideas, art and science. He regarded the last as products of the human mind that may exist independently of any knowing subject: A man-made yet autonomous world (Magee, 1973). In this sense, the interest that a middle school student has for statistical literacy is assumed to exist independently of the researcher. Yet it is acknowledged that social situations, such as the classroom, are highly complex. Post-positivism is a term used to describe a research paradigm with an ontological belief that “objective social facts do exist independently of and external to human beings, but these facts are subject to uncertainty and probability” (Pickard, 2007, p. 7). Such a belief leads inevitably to a methodology that is primarily quantitative, one relying on the identification of variables, and the use of experimentation or observation to test hypotheses. It also underlies the need to use stochastic Rasch models, rather than deterministic models, to explain relationships between measured variables. In this study the reality of middle school students’ interest for statistical literacy is approximated using a few salient variables, which it is assumed can be measured.

Given the broad ontological and epistemological stance, described above, the methodology associated with the quantitative data collected in the study is governed by more technical considerations. In particular, there is the need to consider the nature of data before applying statistical models to them. This broad approach has influenced the methodology, with a deliberate decision to use Rasch models to validate the internal structure of the proposed measures. Unlike factor analytic techniques, Rasch models acknowledge the ordinality of data generated from Likert-type scales. In addition to this, the subsequent analysis of students’ data employs models that as closely as possible reflect the nature of these data, including, for example, their inherent hierarchical structure.

Broadly, then, the methodology in the study commenced with the development of a theoretical model that delineated the statistical literacy interest construct. Based on the model, specific items were developed and subsequently tested on a large representative sample of Australian middle school students aged between 11 and 17 years. Through a cyclic process of testing and development, a final measure of interest was constructed. The measure was then used in a series of statistical models in order to identify factors that influence students' interest in statistical literacy and to establish the influence of this interest on their cognitive development. As is discussed, the results of the modelling provided external validity evidence for the proposed measure. During the process of scale development, a secondary measure of students' self-efficacy in statistical literacy was also developed. The key stages of the methodology are summarised below.

1. Planning, which included gaining ethical clearance to undertake the study.
2. The development of the Statistical Literacy Interest Measure (SLIM) and the students' Self-efficacy for Statistical Literacy (SESL) scale. The development of these measures in turn involved a number of steps:
   (a) The specification of appropriate theoretical models.
   (b) Writing of items that reflect these models.
   (c) Expert review of the items.
   (d) Initial quantitative testing of the items.
   (e) The collection of external measures to assess the validity of the proposed measures.
   (f) Modification of the items and subsequent testing.
3. The use of the two proposed measures to analyse middle school students' interest and to answer the research questions.

4.2 Study participants

Sample design

The study sought to obtain a cross-sectional sample of Australian middle school students. Although some Australian schools have dedicated middle schools, the students in the study deemed to be middle school students were those in Years 7, 8 or 9 of school, although some older and younger students were also included. An Australian Year 7 student is typically 12 years of age and in his or her eighth year of school if he or she has completed a preparatory year.

The sample is a selected sample and schools were not chosen randomly. A convenience sample of schools from the Australian state of Queensland was obtained. In addition to this all schools participating in the StatSmart project were invited to participate. The StatSmart project aimed to investigate the influence of teachers' professional development in statistical literacy on their students' outcomes. It involved 17 schools from three Australian states: Victoria, South Australia, and Tasmania. Although all schools participating in StatSmart were invited to participate in this study, specific schools were targeted so that as closely as possible the resulting sample might reflect the major demographic features of the population of Australian middle school students, assumed to consist of approximately equal proportions of each gender and equal proportions of students in each of Years 7, 8, and 9.

In addition to gender and year level, the proportion of students in government and non-government schools was considered. The Australian Bureau of Statistics (2008) reported that 61% of all Australian secondary students attended government schools, the remainder attended independent or Catholic schools. Approximately 41% of boys and 45% of girls who attended non-government schools also attended a single-sex school (Australian Bureau of Statistics., 1997).

The use of such a selected sample of schools, although not ideal, may not be of major consequence to the generalizations that are to be made from the study's results. In a recent meta-analysis of 13 major English educational studies, Hutchison (2009) analysed the influence of clustering on study outcomes. More specifically he examined the clustering caused from collecting data from students within schools. For each study he calculated the "coefficient of intra-class correlation (ρ)" (Kish, 1965), which is defined as the ratio of the variance between clusters, in this instance between schools, to the total variance. A value of $\rho = 0$, theoretically indicates the random assignment of students to schools, in that all variation in the sample is attributed to within school effects and no variation exists between schools. A value of $\rho = 1$, on the other hand, theoretically indicates that each student within a given school is identical for the given attribute, in that all variation in the sample is attributed to between school effects and no variation exists within given schools. Hutchison (2009) reported that for secondary school students the mean value of ρ for attitudinal items was .04, which indicates that most variation in attitudinal items is a result of within-school effects and not between-school effects. Hutchison also reported that the mean value of ρ for attainment items is .28, which indicates that stronger between-school effects occur for attainment. These results indicate that there is much less variation between English schools for student attitudes than for their attainment. It is likely that a similar result would hold in Australia for student interest, in that only a small proportion of the overall variance in interest scores should be attributed to school effects. The random selection of schools, therefore, is not of as much importance when dealing with attitudinal measures as when dealing with attainment measures.

Selection of students

A number of schools from four Australian states were invited to participate and a copy of the principal invitation is shown in Appendix D. Those who agreed to participate nominated suitable classes of students who were asked, via their class teacher, to participate. A copy of the student invitation is shown in Appendix D. Students who agreed and whose parents agreed then completed the questionnaire. Consequently a total of 1384 students from 16 Australian schools were invited to participate in the study and 791 returned complete surveys across the project, which is a response rate of 57%.

Data collection occurred in three stages over a 12 month period. In the first stage pilot testing was undertaken using a sample of Queensland middle school students. The second stage, which occurred six months later, involved a sample of middle school students from schools participating in the StatSmart project. The third and final data collection stage involved students from both StatSmart and Non-StatSmart schools. A breakdown of schools and students in each stage is shown in Table 4.1.

Table 4.1

Details of students and schools in each stage of the study

	Students			Schools	
Stage	Number N	Mean age (yrs)	Males (%)	Government	Independent
Pilot	221	13.3	35	3	3
Second	145	13.9	54	2	3
Final	425	13.6	47	2	7
Overall	791	13.6	46	5	11

More independent schools were willing to participate in the study than government schools, consequently 38.7% of all participating students attended government schools. Thirty-one percent of the 485 students who attended independent schools did so in a single-sex setting.

Overall the ages of students ranged from 11.2 to 16.8 years, with a mean of 13.6 years. Their year level at school ranged from Year 6 through to Year 10. The number of students in each year level is shown in Table 4.2.

Table 4.2

Number of students in each year level

Year 6	Year 7	Year 8	Year 9	Year 10
24	250	260	212	45

A total of 483 students participating in this study attended StatSmart schools. Of these students, 145 participated at the end of the first year of the study, while the remaining 338 did so during the first half of the following year.

4.3 Instruments used and data collected

Statistical Literacy Interest Measure (SLIM)

The major instrument used in the study is the Statistical Literacy Interest Measure (SLIM). It contains 16 items from a larger interest inventory of 30 items, shown as items R1 through to R36 in Appendix A. The development of the items in the interest inventory and the subsequent development of SLIM are described in Chapter 5, whereas final results are reported in Chapter 6.

Self-efficacy for statistical literacy (SESL) scale

Given the expected close association between interest and self-efficacy, the second major instrument used in this study is the Self-efficacy for Statistical Literacy (SESL) scale, shown in Appendix A as items S41b through to S50c. As with SLIM, the initial development of this instrument is described in Chapter 5 and final results are reported in Chapter 6.

Demographic and other data

All students were asked to provide some demographic data. These included their age, their gender and their year level at school. Students in the second and final stages were also asked to provide their names so that achievement data could be linked to their interest data.

Students in the final stage were asked four questions regarding the frame of reference they used when making interest assessments. These questions, shown as items IE42 to IE45 in Appendix A, were answered as self-descriptions with the existing five-point Likert scale. The first self-description, worded "compared to others in my class I am good at maths," sought to assess the extent to which students used an external frame of reference. The second self-description, worded "out of all my subjects I usually get my best marks in maths," sought to assess the extent to which students used an internal frame of reference. The third self-description, worded "I find statistics more interesting than other work we do in maths," sought to assess the extent to which students compared their interest in statistics with their interest in other areas of mathematics. The last self-description, worded "the statistics that I do in maths classes is more interesting than the statistics that I do in other subjects," sought to assess the extent to which students compared their interest in the statistics encountered in mathematics classes, with the statistics encountered in other classes.

Achievement data

Teachers of students in the second and final stages of the study were asked to provide a rating of their students' mathematics achievement. Teacher ratings of student achievement are known to be strongly predictive of actual student achievement (Egan & Archer, 1985) and display high levels of validity (Hoge & Coladarci, 1989). The teachers in this study were asked to rate each student on a five point scale from A, the best category of achievement, to E, the worst category of achievement. This A to E assessment category is used throughout Australia, having been mandated by the Australian Government (Department of Education, Science and Technology, 2005). Of the 570 students participating in these two stages, achievement data were available for 452. The distribution of their grades is shown in Table 4.3.

Table 4.3

Distribution of mathematics grades (Maths-grade)

Category	Frequency	Percent
A grade	116	25.7
B grade	190	42.0
C grade	107	23.7
D grade	29	6.4
E grade	10	2.2
Total	452	100.0

In order to control the influence of classroom factors on achievement, a relative mathematics grade was also considered. More specifically, the student's grade relative to the median grade of his or her class was determined. As an example, a student with a maths grade of B in a high-ability class where the median grade was A, was assigned a below median grade. These adjustments resulted in a three category structure that is shown in Table 4.4. Although this variable enabled classroom factors to be controlled, the resulting three category structure resulted in an unavoidable loss of statistical power (Manor, Matthews, & Power, 2000).

Table 4.4

Distribution of relative mathematics grades (RelMaths-grade)

Category	Frequency	Percent
Below median grade	120	26.5
Median grade	227	50.2
Above median grade	105	23.2
Total	452	100.0

A measure of students' statistical literacy knowledge (SLK) was also available from some of those students in this study who attended StatSmart schools. Students in these schools who were actually involved in the StatSmart project completed a series of tests that assessed their knowledge of statistical literacy. More specifically, upon entering the project students completed a pre-test, approximately six months later they undertook a post-test, and finally 12 months later they completed a longitudinal test. The items used in these tests and the method used for scoring these items, are detailed in Callingham and Watson (2005). Further details regarding the methodology used in the StatSmart project are described in Callingham and Watson (2007).

Of the 483 students in this study attending StatSmart schools, 188 did not complete a StatSmart test. Such students were recruited by the teacher from classes that they had not nominated for participation in the StatSmart project. Teacher motives for including or not including classes of students in the study are unknown. It is unlikely, however, that the high proportion of missing data in this instance would adversely influence the study's results. As a result of these missing data, SLK scores were only available for 295 students. Of these students, 161 completed their StatSmart tests at the end of the first year of the study with the remainder completing theirs at the beginning of the next year. Seventy-one of the students who completed their StatSmart tests at the end of the first year of the study completed their interest assessment approximately six months later in the first half of the second year of the study. During this intervening period, however, summer holidays occurred making it unlikely that their interest in statistical literacy would have changed significantly.

Teacher influences

As discussed, students who completed a StatSmart test did one of three tests: a pre-test, a post-test, or a longitudinal test. The type of test students did, therefore, is a variable that represents a measure of how long students were in a StatSmart school. In many cases it also represents a measure of how long they were in a class taught by a StatSmart teacher, in that students in the class of a StatSmart teacher did a pre-test near the beginning of the school year and a post-test near the end of the year. Students who did the longitudinal test, which was administered one year later, may not have been with the same teacher, but were in the same school. Given that it was the teachers who were directly involved in the intervention, this variable represents a measure of the influence of the teacher and/or school over and above other individual factors. Of the 295 students in this study who completed StatSmart tests, 49% did the pre-test, 32% did the post-test and the remainder did the longitudinal test.

Variables used during modelling

In order to answer the research questions, a number of variables were created that reflect the data described above. A summary of these is shown in Table 4.5.

Table 4.5

Summary of instruments and associated variables

Instrument	Assessment method	Variable
SLIM	Rasch-scaled student responses	Interest
SESL	Rasch-scaled student responses	Self-efficacy
StatSmart tests	Rasch-scaled student responses	SLK
Achievement data	Teacher obtained estimate from A to E	Maths-grade
Achievement data	Teacher estimate relative to class median grade	RelMaths-grade
StatSmart tests	Item recording the type of test.	Teacher

4.4 Analysis of data

Given that the study is quantitative in nature, this section commences with an overview of the quantitative analysis. It then expands upon the overview, commencing with a theoretical background to the Rasch models used in the study and then reporting details of the analysis as it relates to each of the specific research questions, outlined in Section 3.5.

An overview of the quantitative analysis is shown in Figure 4.1, which shows the progression of analysis, from raw data sources – shown as rectangles on the figure – through to scaled person abilities – also shown as rectangles. During this progression, statistical and/or measurement models – shown as ovals on the figure – were used for a variety of purposes that include scale construction and the modelling of data. The top half of the diagram depicts the scale construction phase, reported in Chapter 5 and the beginning of Chapter 6. The bottom half of the diagram depicts the modelling phase, which is reported in the later sections of Chapter 6.

The analysis of data was in the main sequential, in that it commenced with students' responses to the items in the study questionnaire. The two primary scales – the Statistical Literacy Interest Measure (SLIM) and the Self-Efficacy for Statistical Literacy scale (SESL) – were then constructed from these data using the Rasch Rating Scale model (Andrich, 1978). The process used for the development of these two scales, however, was iterative, in that items displaying misfit were removed from the analysis and the data from remaining items re-analysed. During these iterations dimensionality was assessed and outliers reviewed.

After the two scales were constructed, students' responses to all scales, including the external measures that are described in Chapter 5, and the StatSmart tests, were analysed using the appropriate Rasch model. These models were then used to obtain student ability estimates on each of the scales.

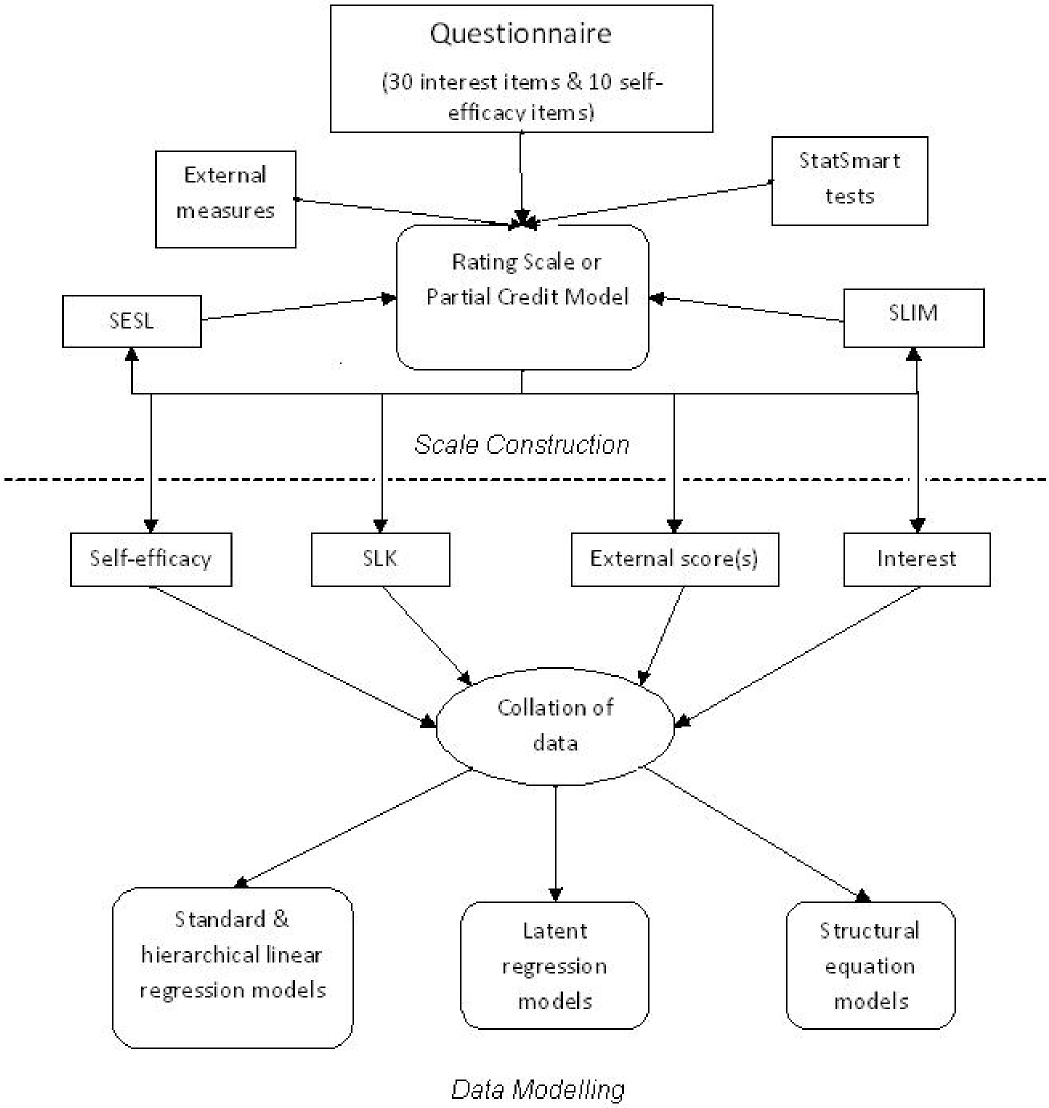


Figure 4.1. Overview of quantitative analysis undertaken in this study

The lower half of the diagram depicts the subsequent collation and modelling of data that occurred in order to answer the research questions.

The Rasch measurement model

The Rasch measurement model, which was first developed by the Danish mathematician Georg Rasch (1901-1980), provides an appropriate method for analysing ordinal data. It is a theoretical stochastic model that is arguably a practical realisation of additive conjoint measurement (Perline, Wright, & Wainer, 1979). In other words, the theoretical Rasch model has the potential to transform the ordinal data obtained from Likert type scales into data that, according to axiomatic measurement theory (Michell, 1990), are true interval data.

In its most basic form, the Rasch model is applied to dichotomously scored test results. It assumes the existence of a unidimensional latent ability trait, measured by a variable β. Further, the difficulty of test items δ, are assumed to be marks or positions upon this variable. The model assumes that the probability of the n^{th} student correctly answering the i^{th} item (P_{ni}) is related to the difference between his or her ability β_n and the difficulty of the particular item δ_i. The greater this difference, the more likely it is that a student will answer the item correctly. More specifically, the Rasch model assumes a logistic relationship between this probability and the difference $\beta_n - \delta_i$, that is:

$$\log_e \left(\frac{P_{ni}}{1 - P_{ni}} \right) = \beta_n - \delta_i. \tag{4.1}$$

The basic Rasch model has been used extensively in the construction of tests of achievement, and the quality control of these tests (Carmichael & St. Hill, 2006; Keeves & Alagumalai, 1997).

The dichotomous Rasch model has been expanded to cater for ordinal data (Andrich, 1978) and thus can be applied to the data produced from Likert scales. The Rating Scale Model (RSM) predicts the probability of the n^{th} student selecting the k^{th} category of the i^{th} item. Such a probability, in turn, is based on the probability that the student will select the k^{th} category in preference to the $(k-1)^{th}$ category ($P_{nik|k-1}$). This extension of the dichotomous Rasch model is given by:

$$P_{nik|k-1} = \frac{\exp(\beta_n - \delta_i - \tau_k)}{1 + \exp(\beta_n - \delta_i - \tau_k)} \tag{4.2}$$

where τ_k is the category threshold parameter. This parameter is the point on the interest continuum at which there is an equal likelihood of the student selecting either the k^{th} or the $(k-1)^{th}$ categories.

The Partial Credit Model (PCM) is also an extension of the Rasch model that is designed to cater for ordinal data. Unlike the RSM that assumes thresholds are fixed for each item, the PCM allows different thresholds for different items. Consequently the PCM can cater for tests with items that have different numbers of possible response categories. The formulation of the PCM is as above, except that the thresholds (τ_{ki}) are subscripted by both threshold number (k) and item number (i).

Given a close correspondence between the data and the constraints of the theoretical model, a student's score (as measured by his or her total response to the items in the test), the difficulty of the items (as measured by the total student response to each item) and the model's expected probability that he or she will answer a given category, will form a conjoint system, which in turn implies that the three variables can be measured on an interval scale. The challenge, though, is to ensure that the measurement instrument used for the particular group of students produces expected frequencies that closely match the requirements of the theoretical model. It is the role of the practitioner to modify the data in order to fit the constraints of the theoretical model, rather than the standard practice of modifying the model to fit the data. In practice this means finding items that are suitable for the subjects in question.

Model parameters for both models can be estimated by maximising a distribution conditioned on the sufficient statistics. Precise formulations of these models, together with details on parameter estimates are described in both Andrich (1978) and Anderson (1997). In this study, the software package *Winsteps* (Linacre, 2006) was used to obtain item difficulty and person ability estimates. Both estimates are reported in logits, which are the natural logarithm of the odds ratio.

Fit statistics. The issue of model fit is important in Rasch analysis, as it provides evidence for the structural validity of the measure. Fit statistics are based on the difference between expected model values and observed values and are used to assess the proximity of an empirical data-set to its theoretical conjoint equivalent. Poorly fitting items or categories are analysed and modified in order to obtain data as close as possible to a conjoint system.

Rasch modelling programs commonly produce two fit statistics for each item: the *outfit* (v_i) and, the *infit* (u_i). Since the expectation of each is 1.0, items whose fit statistics differ considerably from this value can be regarded as being inconsistent with the model. In line with a recommendation by Keeves and Alagumalai (1999), items whose infit lie in the range $0.77 < u_i < 1.30$ can be regarded as have satisfactory fit. As the outfit statistic is known to be influenced by erratic student responses (Bond & Fox, 2007), it should display a greater variance than the infit. Consequently a larger acceptance interval of $0.60 < v_i < 1.40$ is appropriate for this statistic (Bond & Fox, 2007). Both u and v can be transformed into approximate standardized normal statistics, denoted Z_u and Z_v respectively (Smith, 1991, p. 545), which can be compared against critical values in the usual manner. In general, values of Z_u and Z_v that exceed 3.0 indicate misfit. All four statistics are reported in this study.

Items whose fit statistics lie above the acceptance interval display *underfit* and are characterized by high degrees of noise. Those whose fit statistics lie below the acceptance interval display *overfit* and are characterized by responses that are too predictable. Overfit is not likely to have any practical consequences for measurement situations in the social sciences (Bond & Fox, 2007). As a result an emphasis is placed on underfit in this study as reported by the infit statistic.

The fit statistics are used to detect random error in the model. Systematic error may also be present in the model if different groups of students respond to given items in different ways, termed *differential item functioning* (DIF), or if other latent dimensions are evident. DIF can be detected if item estimates obtained from the responses of one group are significantly different from those estimates obtained from the other group(s). It is recommended that when a number of items are tested for DIF simultaneously, a Bonferroni adjustment be used in order to minimise the likelihood of incorrectly detecting items (Linacre, 2006a).

Unidimensionality. The presence of secondary latent dimensions can also produce systematic error. In order to assess the existence of such secondary dimensions, Linacre (1998) recommended that a principal component analysis (PCA) be conducted of the standardised residuals, that is, of the unexplained variance remaining after the major latent trait is removed. The presence or otherwise of multidimensionality can then be tested through application of a multidimensional Rasch model (Adams & Wu, 1997). Such a model accommodates the presence of subsets of items in a given test, each one assessing a different unidimensional latent trait. The test of multidimensionality involves a comparison of the deviance of competing models, in this instance a unidimensional model encompassing all items with multidimensional models involving subsets of items. In this study such testing was conducted using the Rasch software program *Conquest* (Wu, Adams, Wilson, & Haldane, 1998).

Category statistics. In addition to the consideration of item fit statistics, properties of the category threshold parameter estimates (τ_k or τ_{ki}) also need to be considered. Primarily these thresholds should be ordered, so that the estimate of the threshold between category 1 and 2 (τ_2), is less than the estimate of the threshold between category 2 and 3 (τ_3). In addition to this basic property, Linacre (1999) recommended that the distance between thresholds in a five category scale should exceed 1 logit and that there should be at least 10 counts in each category.

Reliability of measure. Apart from considerations of validity, a measure must also demonstrate its reliability or its accuracy. More specifically, reliability is defined as the ratio of true score variance to observed variance (Haertel, 1997), with the former usually partitioned into observed and error variance components. In a Rasch analysis, the error variance is estimated from the standard errors associated with each person's ability measure (Smith, 2001) and the resulting reliability estimate is termed the *person separation reliability* (R_p).

Analysis related to Research Question 1

> How valid is it to base a measure of middle school students' interest in statistical literacy on their responses to a series of interest self-descriptions?

To answer this research question, a measure of interest was constructed and evidence was collected to support the validity of interpretations made from this measure. The development of this measure, in turn, involved a number of procedures that are reported in Chapter 5. The primary analysis centered on the application of the Rasch Rating Scale Model to student responses, as described in Bond and Fox (2007).

Analysis related to Research Question 2

> How do factors unique to an individual, such as their age, prior achievement, gender, and self-competency beliefs, contribute to their interest in statistical literacy?

In order to answer this research question, interest and self-efficacy person abilities, hereafter termed *scores*, were assigned to each student on the basis of the Rasch analysis of their responses to both SLIM and SESL. In addition to these, a statistical literacy knowledge (SLK) score was also assigned to each student. The SLK score was calculated on the basis of students' responses to the StatSmart tests described in Section 4.3. The Partial Credit Model was applied to the responses of 2081 students who completed the StatSmart tests at the end of the first year of this study or early in the second year. This analysis was then used to create SLK scores for all students including those 295 that also participated in this study.

The measures were then used in a series of linear regression models that sought to test the paths in the hypothesised model described in Section 3.4. The regression models are based on a number of standard assumptions that include the random selection of subjects, independence between their responses and that variables are measured without error. Randomness was not possible in this study, however dependence between students in the same class and/or school can be overcome if hierarchical linear models (Raudenbush & Bryk, 2002) are used instead. In this study, the hierarchical nature of the data was analysed with mixed effects models using the software package R (R Development Core Team, 2009), as described in Faraway (2006). Measurement error in predictor variables can be attenuated using reliability considerations (Aiken & West, 1991). It is recommended that latent regression models be used to overcome measurement error in the response variable (Adams & Wu, 1997). The software package *Conquest* (Wu et al. 1998) was used in this study to apply these models.

In addition to linear models, a structural equation model (Byrne, 2001) was used to investigate the inter-relationships among variables. Such models are not developed within a Rasch measurement paradigm and were used in this study for comparative purposes. They consist of a structural component, in this case one that reflects the hypothesised model described in Section 3.4, and a measurement component, one that describes the relationship between observed student responses to items and the latent variables used in the structural model. In this study, model path coefficients and fit statistics were calculated using the software package AMOS (Arbuckle, 2008) and two model-fit statistics were reported: the comparative fit index (CFI), because it is considered more suitable for smaller samples (Bentler, 1990), and the root mean square error of approximation (RMSEA), because it is regarded as the "most informative criteria in covariance structure modelling" (Byrne, 2001, p. 84). Model fit is regarded as satisfactory if $CFI \geq 0.95$ and $RMSEA \leq 0.06$ (Hu & Bentler, 1999).

Structural equation models assume that the ordinal data generated from the Likert scales reflect distinct points on an underlying continuous variable. Such an assumption may be tenuous. Nevertheless, Byrne (2001) argued that the estimation process appears to be quite robust to this violation provided that the ordinal data have at least four categories and the distribution of the ordinal data for the specific items is symmetric. When the distribution of some items are skewed positively and others negatively, however, estimated path coefficients are likely to become distorted (Bollen & Barb, 1981).

Students' frame of reference. In order to assess the influence of students' frame of reference on their interest assessment, students' responses to items IE42 and IE43 were compared with their measure of interest. Item IE42, "compared to others in my class I am good at maths," assessed students' use of an external frame of reference, whereas item IE43, "out of all my subjects I usually get my best marks in maths" assessed their internal frame of reference. An analysis of variance was used to determine whether students' interest assessment was influenced by the extent to which they favored one or both of these frames of reference. In addition to this, the interaction of the two frames of reference on students interest was assessed graphically.

Students' ability to differentiate between mathematics and statistics. In order to assess the extent to which students differentiate between mathematics and statistics, students' responses to items IE44 and IE45 were analysed. Item IE44, "I find statistics more interesting than other work we do in maths," assessed students ability to differentiate statistics from the rest of the mathematics curriculum, whereas item IE45, "the statistics that I do in maths classes is more interesting than the statistics that I do in other subjects," assessed their ability to differentiate the statistics encountered in maths from those encountered in the wider curriculum. The analysis of students' responses to these two items was primarily descriptive.

Analysis related to Research Question 3

> To what extent does students' interest in statistical literacy influence their subsequent achievement in statistical literacy?

Using a similar methodology as that described for Research Question 2, several linear regression models were initially developed with SLK as the response variable. These models were then used to generate a path model that would accurately reflect the data. This path model, in turn, was then analysed and tested using the structural equation modelling process described earlier.

4.5 Data analytic procedures used in the study

The preceding sections have described the subjects, instruments and analysis methods used in the study. In this section procedures specifically related to the analysis of data are detailed. In particular, the following discussion addresses the pooling of data-sets that was necessary during the study, and the treatment of outliers and missing values.

Treatment of data sets

The pilot testing of items in both instruments was based on a sample of students from Queensland. Subsequent testing of items was to have been based on a mix of students attending schools in Queensland and schools in the StatSmart project. As a result of low response rates from non-StatSmart schools, however, a large proportion (80%) of this latter sample, reported in Table 4.1 as collectively the second and final stages, attended StatSmart schools. Given the effectiveness of the StatSmart intervention, it is possible that such students may have different response patterns to non-StatSmart students. Indeed, as is reported in Chapter 6, this was found to be the case. In order to obtain a more representative sample, therefore, the responses from students in all three stages of the study were pooled to form a large sample, on which final testing of the items was performed. The pooling of these data was possible because 24 of the 30 interest items and 6 of the 10 self-efficacy items remained unchanged throughout the study, which is ample for common item linking using Rasch models (Wright & Stone, 1999). The pooling of data also ensured that the sample size was sufficient to provide a high level of stability to item calibrations (Linacre, 1994). The subsequent modelling of data was also based on this pooled sample.

Treatment of missing data

The software package Winsteps (Linacre, 2006b), used during the study, employs an estimation method that ignores missing item responses. Instead, the program estimates person abilities on observed marginal counts. This is a strength of the particular estimation method, but it does mean that ability estimates for some students can be based on very little statistical information. Accordingly, it was decided to remove the responses of students who completed less than one half of the scale items. Seventeen students, for example, failed to complete the second page of the study questionnaire and in doing so only responded to five of the 16 items in SLIM. Their responses were removed from the analysis. Similarly, four students completed five or fewer items in SESL and their responses were also removed.

The case matching that occurred among the various data sets used in the data modelling stage often resulted in missing data. For example, if data modelling involved a variable for which only half of the subjects had observations, then only that half of the data-set was used. In random samples, it is essential to ascertain whether such data are missing at random or in a systematic way. The sample in the study, however, was not random and the limitation has been duly noted. The detection of bias in the missing data was therefore considered to be unnecessary, although subsequent interpretations noted its possibility. The smaller data sets that emerged from case-matching also impacted upon the representativeness of the sample. Again, this limitation in the analysis was unavoidable but duly noted.

Treatment of outliers

During the application of statistical models the existence of outliers or influential data points can adversely influence the model's estimates. The detection of outliers occurred at two stages during the analysis: during scale construction, shown in the top half of Figure 4.1, and then later during data modelling. In regard to scale construction, the Rasch model routinely reports person-fit statistics, calculated in the same way as the item-fit statistics discussed earlier. The responses of students with abnormally high or low item fit statistics, those with standardised values exceeding 3.5, were considered outliers and reviewed. Unlike traditional statistical models where outliers reside in the tail of population distributions, the atypical responses produced by Rasch outliers often mean they reside near the middle of the ability distribution

(Wright, 2000) where they are unlikely to have significant leverage. During the analysis of outliers in this stage, their removal in most cases had very little impact upon item statistics. In addition, the analysis of specific person-item responses for these outliers was inconclusive, in that it was difficult to judge whether unusual item responses were invalid, in error, or due to the inherent idiosyncratic nature of personal interest choices. For these reasons, it was decided to retain the responses of students identified as Rasch outliers, especially given the low-stakes nature of the interest assessment and the apparent lack of influence these data had on item statistics.

The detection of outliers also occurred during data modelling, shown in the bottom half of Figure 4.1, with standardized plots and residuals used for this detection. More specifically, data points with standardized residuals exceeding 3.5 were deemed to be outliers, such a cut-off ensured that only the most extreme outliers were identified. These outliers were then assessed for the degree of influence they had on regression coefficients. This influence, in turn, was judged through an inspection of residual plots and in some instances calculation of "Cook's Distance D_i" (Cook, 1977), with values of $D_i > 1.0$ indicating strong influence. In this way data from students with, for example, extremely low interest scores but extremely high self-efficacy scores, were removed from the modelling process and coefficient estimates were based on the remaining data.

4.6 Chapter summary

As was detailed in the chapter, the emphasis in the study was on the use of quantitative techniques to answer the research questions. Accordingly a major portion of the chapter has described the data analytical methods used to answer these questions. Of these, the Rasch model features because it appropriately models the ordinal data produced from Likert scales and produces measures that reflect the view adopted in the study that social facts are "subject to uncertainty and probability" (Pickard, 2007, p. 7). A number of other models were also introduced, each with a view to accommodating the unique features of the data. The structural models, however, ignore the inherent ordinality of the data and in a sense have emerged from a methodological paradigm that is diametrically opposed to that of the Rasch model. Their use in this study, therefore, seems and is methodologically inconsistent, but pragmatism must prevail because suitable methods aligned with the Rasch paradigm are currently not available. In summary then, the data-analytic methods described in the chapter and subsequently used in the study, were chosen with a view to modelling the data in way that accurately reflects their nature. Care was therefore taken to ensure that model assumptions were checked and where-ever possible limitations of models are reported.

The discussion in the chapter has provided a detailed account of the methodology used for the study and a rationale for its use. It has described the process for selection of subjects and the data collected. The discussion in the next chapter describes the initial development of the two proposed instruments, namely the Statistical Literacy Interest Measure and the Self-Efficacy for Statistical Literacy scale.

Chapter 5

Instrument development and pilot study

The discussion in this chapter reports the development of two instruments used in the study, the Statistical Literacy Interest Measure (SLIM), and the Self-Efficacy for Statistical Literacy (SESL) scale. It commences with a theoretical review of issues relating to the validation of psychometric measures and then develops operational models of both interest and self-efficacy in statistical literacy. Following this, the discussion describes the development of a bank of items written to reflect these operational models. It then reports on the initial testing of these items that is based on a pilot study conducted in Queensland, Australia. As as result of this testing, the discussion proposes interval measures of interest and self-efficacy in statistical literacy that appear to conform to the requirements of the Rasch measurement model. Finally, the discussion addresses the validity issues introduced in the beginning of the chapter and presents preliminary evidence to suggest that interpretations based on the two proposed measures are valid.

5.1 Theoretical background

The following discussion addresses issues relating to the validation of psychometric scales, it then describes the development of theoretical models of interest and self-efficacy that are subsequently used as the basis of item construction.

Scale validation

The process of scale validation is investigative in nature and is primarily one of obtaining evidence to support the intended use of the scale (Wolfe & Smith, 2007a). It is envisaged that SLIM and SESL will both be used for evaluative purposes. They will facilitate the affective, rather than cognitive, evaluation of educational interventions and also enable a continued exploration of the statistical literacy hierarchy.

The development of both instruments was completed in such a way as to obtain evidence for subsequent interpretations that may be made from these measures. Messick (1995) suggested that there are six forms of evidence that are needed to support the validity construct:

1. Content evidence includes arguments that relate to the relevance, representativeness and technical quality of the items. The relevance and representativeness of items can be judged by expert review and rely, in part, on the identification of a "Universe of Generalisation" (Kane, 2006), a theoretical model describing the proposed trait(s). In a Rasch measurement paradigm, evidence to support the technical quality of items is provided in the fit statistics (Wolfe & Smith, 2007b).

2. Substantive evidence refers to the extent to which underlying theories predict the observed outcomes. Wolfe and Smith (2007a) argued that substantive evidence should be based on at least three underlying theoretical models. The first, termed the *internal model*, describes the dimensions and components of the construct and how they interact. It is the Universe of Generalisation, described above. The second theoretical model, termed the *external model*, describes how the construct interacts with external but related constructs. In this study, the relationship between constructs related to interest is presented in Section 3.4. The last theoretical model, termed the *developmental model*, describes how the construct changes over time. The development of interest with age is described in Section 2.3, whereas its development with knowledge, predicted by the Model of Domain Learning, is described in Section 3.2.

3. Structural evidence refers to the extent to which the internal structure of

the measure reflects the theoretical structure of the construct. The use of a Rasch measurement model, as is proposed in this study, implies that the underlying construct is unidimensional.

4. Evidence as to the construct's generalisability, refers to the extent to which the findings from this sample of items and students, can be applied to the construct in other samples of students. A simple test of the generalisability of the measure is to examine the invariance of item difficulty estimates between two samples of students (Smith, 2001).

5. External evidence refers to the extent to which the scores obtained from the measure correlate with other previously validated constructs. Given that the development of both proposed measures was done because no others exist in this particular context, the provision of external evidence was achieved through an exploration of students' interest and self-efficacy in mathematics.

6. Consequential evidence concerns the future impact that any proposed instrument may have on students who complete the instrument. Given that both instruments are designed to be used for evaluative purposes, it is important that specific items do not differentiate between sub-groups of students (Smith, 2001).

The first stage in the development of an instrument is the specification of an operational or internal model (Kane, 2006). The model is used as the basis for item development and later as a theoretical benchmark against which content validity is assessed. The discussion in the remainder of the section outlines the specification of operational models of middle school students' interest and self-efficacy in statistical literacy.

An operational model of interest in statistical literacy

Based on motivation theory (Schunk, 1996), for many students in a middle school context, their interest in and their knowledge of statistical literacy are dynamic and interactive, in that their content knowledge influences their interest, and their interest influences their content knowledge. Because of this assumed interaction, the discussion in this subsection seeks to clarify and define the nature of the statistical literacy interest construct. It is suggested that there are three main elements associated with students' interest: reflective interest, curiosity interest, and importance interest. Along with these, two content components are also proposed. The outcome at the end of this section is a taxonomy grid constructed using the three interest elements along the horizontal axis and the two content components along the vertical axis, as the starting point to develop an operational model of students' interest in statistical literacy.

The interest assessed using self-report survey questions is regarded as an estimate of the students' individual interest in a specific topic (Schiefele et al., 1992). As such, students' responses to interest surveys typically reflect the value that they place on the context or activity described in the survey items. This value is typically influenced by their past experience, current interests, knowledge, and goals. It is also influenced by their level of emotional attachment to the topic.

The first element of interest, termed *reflective interest*, is assessed through items with the common stem "I'm interested in". The stem targets both the specific situations that students might encounter, such as "working out the probabilities for dice," and also a student's desire to re-engage in statistics, such as "getting a job involving statistics." It is assumed that students who endorse the latter have those predispositions to re-engage with statistics that are associated with high levels of individual interest. The Model of Domain

Learning (Alexander, 2003), however, predicts that the novice learners typically encountered in a school setting are more likely to be motivated by the situation and that such learners will exhibit typically low to moderate levels of individual interest. Such students, therefore, should find it easier to endorse items that assess interest in a situation than those that assess re-engagement.

It is also possible for students to anticipate and to reflect upon their interest towards or valuing of content knowledge, which they have yet to experience. For this reason a second element is included in the interest model: A desire to find out about a specific interest object. This element, termed *curiosity interest*, is assessed through items that ask students the extent to which they would "like to know about" certain facts that are related to statistical literacy. This interest element can be regarded as a form of epistemic curiosity (Litman, 2008). Students who would like to find out about statistical literacy do so because they have some, but incomplete, knowledge about the subject or the associated contexts. Because of this, some students may find it easier to endorse items that assess curiosity interest compared with endorsing those items that assess their reflective interest in specific content situations.

Many students in the middle school years may be motivated to engage with statistical literacy because it is seen by them as a necessary part of their school and post-school life goals. Their valuing of statistical literacy may be regarded as primarily extrinsic. Nevertheless, Boekarts and Boscolo (2002) argued that such students can experience interest. For this reason a third element, termed *importance interest*, is proposed. This element is assessed through the common stem "It's important to me personally." Ryan and Deci (2000a) argued that behaviour motivated from perceived importance reflects a lower level of autonomy than behaviour from interest. In this study, it is hypothesised that lower levels of autonomy are manifest in lower levels of the valuing that is associated with interest. It is argued that students who can only see the importance of statistical literacy will have less interest associated value for it, than those who can also acknowledge an interest in specific situations and indeed indicate a willingness to re-engage.

The use of three elements of interest implies a degree of multidimensionality of the construct. In this regard it is considered to be similar to the contemporary perspective regarding students' self-concept, which is seen as being both multidimensional and having an inter-linking hierarchy, in that the different strands come together to form a general or overall construct. This is a notion that Hattie (2009) called the "rope" model, where researchers can either investigate the individual strand(s) or the inter-linked strands, the "rope" of the construct. Following this line of thought several authors regard interest as having two dimensions, importance and emotion, with the former assessed through the item stem "it's important to me personally" and the latter through use of the terms interest or enjoyment. Empirically, however, these dimensions appear to be poorly distinguishable (Köller et al. 2001; Tsai, Kunter, Lüdtke, Trautwein, & Ryan, 2008). Similarly, epistemic curiosity is regarded as synonymous with interest (Kashdan & Silvia, 2009), hence indirectly contributing to the notion that the different strands of the construct called interest come together as one overall general dimensional construct. Although studies have used all or some of these three elements of interest, none have suggested a taxonomy grid model to construct an overall assessment instrument. In regard to this taxonomy grid model, shown in Figure 5.1, the three elements of interest are: importance, curiosity and reflective and these are constructed along the horizontal axis.

In addition to the three elements of interest, it is argued that in a school situation, students' self-reported interest will have two content components (Hoffman, 2002). The first relates to the actual subject matter and the second to the contexts and activities encountered when they learn this subject matter. In regard to the taxonomy grid model, these two content components are constructed along the vertical axis.

Horizontal axis – Interest elements: Importance Curiosity Reflective
Vertical axis – Content components:
1. Statistical literacy subject matter: • Sampling • Graphs • Averages • Chance • Beginning inference • Statistics in general 2. Activities and contexts associated with the learning of statistical literacy, for example: • The use of technology • Classroom and school contexts that may involve the use of data relating to the students themselves. • Wider contexts, including sports and social issues that may or may not be presented in the media.

Figure 5.1. Model to describe students' interest in statistical literacy

The requisite knowledge for a statistically literate person, the subject matter in this instance, is situated in the chance and data strand of all Australian mathematics curricula. The subject matter may for convenience be presented in topics that are identified by Watson (2006) as: sampling or data collection, graphs, averages, chance, beginning inference, and variation. The last topic, although paramount in statistics is difficult to assess, for as Watson (2006, p. 219) herself acknowledges "many curriculum documents do not even mention the word except in connection with the introduction of the standard deviation." Accordingly topics in the current model of interest are restricted to the first five of Watson's topics, reflecting an earlier classification proposed by Holmes (1986). In addition to these five topics a "statistics in general" topic has also been included. Such a topic allows for the inclusion of more general items, those that might span a number of topics, for example: an interest in using statistics to prove a point or win an argument.

A student's interest in the learning of statistical literacy will be influenced by the contexts in which the material is presented and the activities that they encounter. A review of the literature associated with the teaching of statistics suggests that contexts including sports (Lock, 2006), social issues (Bidgood, 2006), and the students themselves (Lee & Famoye, 2006) will enhance student interest. It is argued that a student's interest in statistical literacy will also be influenced by media contexts: It is in the media that students often encounter messages that contain statistical elements (Watson, 1997).

The activities that students encounter in the learning of statistical concepts will also influence self-reports of their interest. As was discussed in Section 3.3, the degree of novelty associated with such activities should influence students' levels of interest. Novelty can be created through the use of technology (Bakker, Derry, & Konold, 2006; Finzer, 2006; Lane, 2006; Mitchell, 1993). In addition to novelty, Mitchell (1993) argued that situational interest would develop into individual interest if the activities encountered were meaningful. It is argued that in a statistics context, data exploratory activities that enable students to answer meaningful questions will enhance their interest.

An operational model of self-efficacy in statistical literacy

The following discussion seeks to define the hierarchy and content coverage for the self-efficacy in statistical literacy construct. In regard to its hierarchy, Bandura (1997) argued that the most powerful source of students' self-efficacy beliefs were their mastery experiences. These, in turn, should be influenced by the cognitive complexity of the task in question. In the statistical literacy context, factors influencing the complexity of tasks have been identified by Watson and Callingham (2003), who developed a statistical literacy hierarchy. It is argued that the hierarchical structure of items in SESL should reflect this hierarchy as described in Section 2.2.

In regard to the content coverage of the construct, it is expected that students' self-efficacy in statistical literacy will be influenced by the topics, in Figure 5.1, that comprise statistical literacy. Context also plays a particularly key role in the development of statistical literacy. Watson (2006) argued that students at higher levels of the statistical literacy hierarchy are more able to interact critically with the contexts in which tasks are situated. Students' self-efficacy towards statistical literacy, therefore, should be influenced by the context in which the tasks are situated. Contexts, however, are chosen by teachers to suit the specific needs of their students and can vary widely. As a result, this study has focussed on more general contexts, in particular those that are school-, and media-related.

5.2 Construction of items

Construction of interest items

The taxonomy grid model shown in Figure 5.1 became the theoretical starting point for the generation of a bank of items to populate the grid and form the basis of SLIM. In particular a bank of 40 self-descriptions was developed to reflect the model. A sample of these items is shown in Table 5.1, which also details the interest element and content component that each item is thought to assess.

Construction of self-efficacy items

The SESL scale was developed in order to understand the interest construct. Given its secondary role and the need to minimise respondent burden, the number of items in the measure was restricted to ten. The development of SESL mirrored that of the interest items. A number of items were written to reflect the topics and contexts associated with statistical literacy and a sample is shown in Table 5.2.

Table 5.1

Sample of interest items

Interest element	Content component: Topic	Content component: Context	Item
Importance	Graphs	Media	It's important to me personally that I can understand graphs that appear on the internet or in newspapers.
Curiosity	Chance	Social issues	I would like to know how scientists calculate the chance of rain.
Reflective	Averages	Sport	I'm interested in using averages to compare sports teams or players.
Reflective	General	None	I'm interested in learning more about statistics.

Table 5.2

Items to assess self-efficacy

Topic	Item
Average	I am confident that I am able to find when a newspaper article has used the wrong type of average.
Sampling	I am confident that I am able to explain how to select a fair sample of students for a school survey.

Expert review of items

The items from both measures were initially reviewed by a panel of experts in the statistical literacy and measurement domains. All were based in the Faculty of Education at the University of Tasmania. The panel was asked to provide feedback regarding the appropriateness of items and also the layout and readability of the survey. This feedback was provided verbally and/or in writing.

After expert review, the items were then reviewed by a group of 45 practicing teachers of middle school students who were involved in the StatSmart project. The teachers were asked to complete the survey as a typical student might do and to note perceived difficulties with any language. As a result of this second review, the language used with some items was altered. For example items assessing probability were re-written to include the term "chance" as it was felt that students were more familiar with the latter word.

Based on the results of this review, 30 of the original 40 interest items were deemed suitable for trialling, as were ten self-efficacy items. The 40 self-descriptions were then compiled into a questionnaire that used a five-point Likert scale, ranging from 1 (*statement doesn't describe me at all*) to 5 (*statement describes me well*). All statements were expressed in a positive way as evidence suggests that the practice of mixing negatively and positively worded statements reduces reliability (Netemeyer, Bearden, & Sharma, 2003).

5.3 Trialling of items

After the initial construction of items, the study questionnaire was prepared and given to a sample of students for testing. The following discussion describes the student sample in this pilot study and provides more details on the data obtained from these students. It then describes the process by which items were further developed during the study, with such development being guided by both informal teacher feedback and the testing of data against the requirements of the Rasch model. The result of the initial testing process was a sample of 30 interest items and ten self-efficacy items from which valid interval measures of interest and self-efficacy were obtained.

Student sample

As reported in Chapter 4, 221 students from six schools participated in the pilot study. The schools included: a large metropolitan government high school, two rural government high schools, an independent girl's high school, and two independent co-educational middle schools. Most students in the pilot attended secondary schools (78%), with 85 enrolled in Year 8, the first year of high school in this state, and 88 enrolled in Year 9. Of the students enrolled in Year 8, however, 17 attended a dedicated middle school, as did all of the Year 7 students, although it is unknown whether these students were taught by specialist mathematics teachers or generalist middle school teachers.

Additional data collected

In addition to the self-descriptions developed for both scales and demographic data, students in the pilot study were also asked to complete two previously validated scales, which were used to provide evidence of external validity.

A sample of ten items from the Mathematics Interest Inventory (MII) (Stevens & Olivaréz, 2005) was used to obtain a measure of students' interest in mathematics. The MII contains 27 items and was developed and validated on a sample of 724 students in the United States of America, whose ages ranged from 9 to 18 years. Stevens and Olivaréz (2005) reported a three-factor structure to the MII of which the largest, consisting of ten items, assessed the degree to which students report a positive attachment to mathematics. The second factor assessed students' negative attachment to mathematics, whereas the third reflected the amount of time they spend on mathematics. Given the need to minimise respondent burden, only the 10 items of the first factor were used in this study. These items are shown in Appendix A as items M1 to M10.

In addition to the MII, students also completed nine items that were adapted from the Self-efficacy for Learning and Performance subscale of the Motivated Strategies for Learning Questionnaire (MSLQ) (Printrich & De Groot, 1990). This version of the MSLQ was written specifically for junior high school students for any subject. The self-efficacy subscale is usually given to students during the actual subject being assessed and contains items such as "compared with others in *this* class I expect to do well." Since some students in this study completed the questionnaire in classes other than mathematics, for example their form class, it was necessary to specify in each item of the MSLQ the class as mathematics. Therefore the previous item was worded “compared with others in *my maths* class I expect to do well.” The items used in this study are shown in Appendix A as items M11 to M19.

Initial testing of interest items

Collection of data during the pilot study occurred over a period of ten weeks. During this period items were continually reviewed, both on the basis of teacher feedback and initial testing. Item testing involved the application of the Rasch Rating Scale model to students’ responses and an examination of fit and difficulty statistics. As a result of this review some items were modified or removed. For example, feedback from participating teachers revealed that students, particular those in Year 7, were unable to answer items that assessed basic inference. An item that asked students for their level of interest in using data from a survey to find out about a large population was removed. In addition to this, some items with very specific contexts tended to elicit erratic responses from students. An item originally designed to assess students’ interest in sports-related averages was worded, “I’m interested in batting averages in cricket or goal averages in netball.” Several students who gave typically low responses for all other items gave a high response for this item, presumably because of their interest in cricket or netball, rather than statistics. This item was written in a more general form as: “I’m interested in using averages to compare sports teams or players.” Such a wording still assessed sports related averages, but in a more general context.

Testing during this stage also revealed a lack of spread in the relative difficultly of items. There was a lack of items that reflected apparent upper levels of interest. To rectify this situation additional items with a general context were included. To assess higher levels of interest, for example, the item "I get so involved when I work with data that I sometimes lose all sense of time" was included. This item was designed to assess the extent to which students may experience "flow" (Csikszentmihalyi, 2002) when they work with data. Although the experience of flow typifies a state of very high situational interest, Csikszentmihalyi (2002, p. 41) argued that after the experience the self becomes more in union with the ideas beyond the self. This union, in turn, reflects John Dewey's notion of "true interest" (Dewey, 1910, p. 91). A student's endorsement of such an item, therefore, should reflect high levels of individual interest. Similarly, the item "I like to work on statistics problems in my spare time" was also included to assess the re-engagement typical of very interested students.

During this initial testing period, three items were modified and three were replaced. Twenty-four of the original items remained unchanged. The 30 items used as the basis for the interest measure are shown in Appendix A and are prefixed with an R, C or I according to whether they are considered to assess reflective, curiosity or importance interest respectively. The classification of these items, with respect to the taxonomic grid model, is shown in Table 5.3. This table indicates that the coverage of items over the identified elements in the taxonomic grid model is adequate.

Initial testing of self-efficacy items

As with the interest items, the items in SESL were tested and further developed during the pilot study. Feedback from teachers suggested an item that assessed confidence to "use data from a sample to answer questions about the whole population" was considered to be inappropriate for younger students, as was a similar interest item discussed earlier. It was replaced with an item (S50b) that assessed the confidence "that I can use data to investigate questions that I

Table 5.3

Cross-classification of inventory items by interest and content element

Content element	Interest element: Importance	Curiosity	Reflective
		Topic	
Sampling		C20	R1, R2, R10
Graphs	I27, I28, I29	C22	R9
Averages	I23		R6b, R7
Chance	I24	C16, C21	R11
Inference	I26	C17, C19	
General	I25, I30b	C38	R3, R4, R13, R14, R15, R31, R36
		Context	
Technology			R12b
School/class	I27, I29, I30b	C21	R10, R11
Media	I23, I26, I28	C20	R1, R9
Social issues	I24	C16, C17, C19	R2, R7
Sports		C22	R6b

might have." Similarly, an item that assessed the confidence to "explain what the word random means" was considered to be inappropriate for many students and was replaced with an item (S47b) that assessed the confidence to "explain when conclusions that are based on surveys might be wrong."

Testing during this stage also reported model underfit for an item designed to assess the confidence in calculating an average. The item, worded "I am confident that I am able to calculate an average result using a calculator or computer if necessary," reported severe underfit ($u_i = 1.51$) and was replaced with an item (S41b) that assessed the confidence to "to solve problems that use averages."

In addition to the three item changes described, one item that assessed confidence in calculating probabilities associated with dice and coins was considered to duplicate another. Given the then limited range of item difficulties, it was replaced with an item (S48b) that assessed the confidence "to arrange data correctly into a table."

As a result of the testing, four items were altered. Six items, however, remained unchanged. The final set of items for SESL is shown in Appendix A as items S41 to S50, although as shown, items S48b and S50b were modified later in the study. The classification of these items by topic is shown in Table 5.4. As is seen from this table, the items sample each of the identified topics of statistical literacy. There is an over-emphasis of items assessing data presentation, although this topic does form a major part of chance and data in the middle school years.

Table 5.4

Classification of self-efficacy items by topic

Topic	Item code
Sampling	S49
Graphs	S44, S45, S46, S48b
Averages	S41b, S42
Chance	S43
Inference	S47
General	S50b

5.4 Development of measures

After the initial development of the items, students' responses were analysed using the Rasch Rating Scale model, with an emphasis on ascertaining the degree of fit between these responses and the requirements of the model. An iterative approach was used, in that items displaying severe misfit were removed from the analysis and student responses to the remaining items were analysed. The process continued until convergence occurred, in that student responses to the final sample of items satisfied the requirements of the model. The results of the process are reported in this section and the significance of the results is discussed in the next section.

The Statistical Literacy Interest Measure (SLIM)

Using the iterative process described, 22 of the 30 interest items formed a measure that explained 62% of the variance in student responses and reported a person separation reliability of $R_p = .88$. These items and relevant statistics are shown in Table 5.5, which reports the item code, an item description, the number of valid student responses (N), the difficulty or interestingness of the item (δ_i), and the infit statistic (u_i). Other relevant statistics are reported in Table B.1 of Appendix B. The estimated category thresholds (τ_k) were: -0.95, -0.29, 0.21 and 1.04. The ordering of these thresholds suggests that the five category structure used was satisfactory, although the distance between them is somewhat less than the recommended value of 1 logit (Linacre, 1999).

As is seen from Tables 5.5 and B.1, item fit statistics appear to be within acceptable limits. The exception is item C16, which although reporting satisfactory infit, reports significantly high outfit ($v_i = 1.47$). The outfit statistic, however, is highly susceptible to unusual student responses, and indeed the removal of just one student's response from this analysis reduced the outfit statistic for the item to $v_i = 1.14$. For this reason the item was retained. The content coverage of items appears to be adequate over the three interest elements despite the removal of several items assessing reflective interest.

Table 5.5

SLIM selected statistics based on pilot study

Item	Description	N	δ_i	u_i
R31	Lose all sense of time when working with data.	78	0.94	0.95
C38	All there is to know about statistics.	81	0.71	0.79
R15	Getting a job that involves statistics.	220	0.61	0.97
C19	How politicians make decisions that are based on data.	220	0.49	0.94
R2	Surveys about how people feel.	220	0.32	1.12
R14	Learning more about statistics.	221	0.30	0.84
R9	Reading graphs in the media	220	0.24	1.22
R11	Working out probabilities for dice, coins and spinners.	221	0.24	0.96
R3	Working on problems involving data and statistics.	220	0.16	0.98
C17	How a survey can be used to predict who will win the next election.	221	0.13	0.95
R12b	Using computer programs to help me investigate data	81	0.07	1.29
I23	Can understand news reports that use averages.	221	-0.01	0.87
I25	Understand the words that are used in statistics.	221	-0.07	0.82
C20	Whether a survey in the media about students was correct	221	-0.18	1.00
C16	How scientists calculate the chance of rain.	221	-0.19	1.19
C21	Whether a game I was playing was fair.	221	-0.19	1.17
I28	Can understand graphs that appear on the internet or in newspapers.	221	-0.50	0.90
I30b	Can use data to investigate questions that I might have.	167	-0.51	1.10
I27	Use the correct graph when displaying my data.	221	-0.63	1.06
I24	Know how to calculate the chance of being injured from risky behavior.	221	-0.64	1.00
I29	Can arrange data into tables.	221	-0.69	1.03
I26	Can believe scientific claims that are based on data.	221	-0.81	1.18

Development of Self-Efficacy for Statistical Literacy (SESL) scale

All ten self-efficacy items formed a measure that explained 68% of the variance in student responses and reported a person separation reliability of $R_p = .78$. These items and relevant statistics are shown in Table 5.6, which reports the item code, an item description, the number of valid student responses (N), the difficulty of the item (δ_i), and the infit statistic (u_i). Additional relevant statistics are reported in Table B.2 of Appendix B. The estimated category thresholds (τ_k) were: -1.21, -0.48, 0.26 and 1.44. The ordering of these thresholds suggests that the five category structure used was satisfactory, although the distance between them is somewhat less than the recommended value of 1 logit (Linacre, 1999).

As is seen from Table 5.6, item S48b, which assessed confidence to arrange data into a table, reported overfit. In addition to this, item S50b, which assessed a confidence to use data to investigate questions, reported high levels of infit, suggesting some underfit for this item. Given these two problems and the need for a further item assessing chance, these two items were replaced prior to the main study. Item S48b was rewritten "I am confident that I can look up the correct number from a table of numbers" and is shown as item S48c in Appendix A. Item S50b was replaced by "I am confident I can work out the most likely outcome from a game involving chance" and is shown as item S50c in Appendix A.

5.5 Preliminary validity evidence

Using the six forms of validity evidence, presented in Section 5.1, the following discussion reports preliminary evidence for the validity of the two measures developed during the pilot study. Given that further data were collected and are reported later in the dissertation, the following discussion is deliberately concise.

Table 5.6

SESL selected statistics based on pilot study

Item	Description	N	δ_i	u_i
S42	Find when a newspaper has used the wrong average.	221	0.80	0.96
S47b	Explain when conclusions based on surveys are wrong.	81	0.64	0.77
S43	Explain to a friend how probability is calculated.	221	0.30	0.99
S45	Explain the meaning of a graph in a newspaper.	220	0.21	0.93
S46	Find a mistake in someone else's graph.	220	0.03	1.04
S49	Explain how to select a fair sample for a school survey.	221	0.03	0.95
S50b	Use data to investigate questions	167	-0.27	1.25
S41b	Solve problems that use averages	80	-0.44	1.09
S44	Show data correctly on a bar chart.	221	-0.51	1.18
S48b	Arrange data correctly into a table	81	-0.79	0.63

Preliminary validity evidence for SLIM

Content evidence. The relevance of items constituting SLIM was assessed by the panel of experts. As is seen from Table 5.3, the original 30 items adequately sampled the operational model of interest. In creating SLIM, however, eight items were removed. Although all content topics are still represented across the remaining items, there are no items specifically assessing reflective interest in averages. Similarly, both items with sports contexts (R6b and C22) elicited student responses that were inconsistent with the requirements of the measurement model. Thus no remaining items assess interest in sports related contexts. Nevertheless, the items in SLIM still sample most elements of the taxonomic grid. The item fit statistics reported in Table 5.5 are satisfactory, supporting the technical quality of the items.

Substantive evidence. Broadly the hierarchical arrangement of items, as shown in Table 5.5, reflects the taxonomic arrangement of interest elements in the internal, or operational model, described earlier. It was expected that only the most interested students would endorse self-descriptions that assessed re-engagement with statistics, such as "wanting to know all about statistics" (item C38). Similarly it was expected that students who acknowledge the experience of flow when doing statistics (item R31) are also likely to have high levels of interest. It was also expected that the valuing associated with importance would reflect lower levels of interest than that associated with reflective or curiosity interest. As is shown in the table, all importance interest items are lower in the hierarchy than reflective interest items. In regard to age development, evidence cited in Section 2.3 suggests that as students progress through adolescence their interest in learning will generally decline. There was no evidence of such decline, with no reported association between students' ages and their SLIM scores. In regard to the external model, discussed in Section 3.4, it was expected that students' interest and self-efficacy in statistical literacy would be associated. This was the case, with a reported positive association between students' SLIM and SESL scores ($r = .59$, $p = 0.00$).

Structural evidence. The major assumption of the Rasch model is the existence of a unidimensional underlying trait. As recommended by Linacre (1999), a principal component analysis (PCA) of the residuals was undertaken and a plot of these loadings against the item difficulties is shown in Figure 5.2. Ignoring the absence of items in the top right hand quadrant of the plot, their scatter suggests an absence of structure in the residuals, which itself lends support for the presence of a single unidimensional factor. Similarly the fact that the principal component explains 62% of the variance also supports the unidimensionality of SLIM.

Generalisability. Issues relating to the generalisability of the instrument are discussed later in the dissertation after the inclusion of additional data.

External evidence. As most students in the sample were assessed during their mathematics classes and as most of the concepts underlying statistical

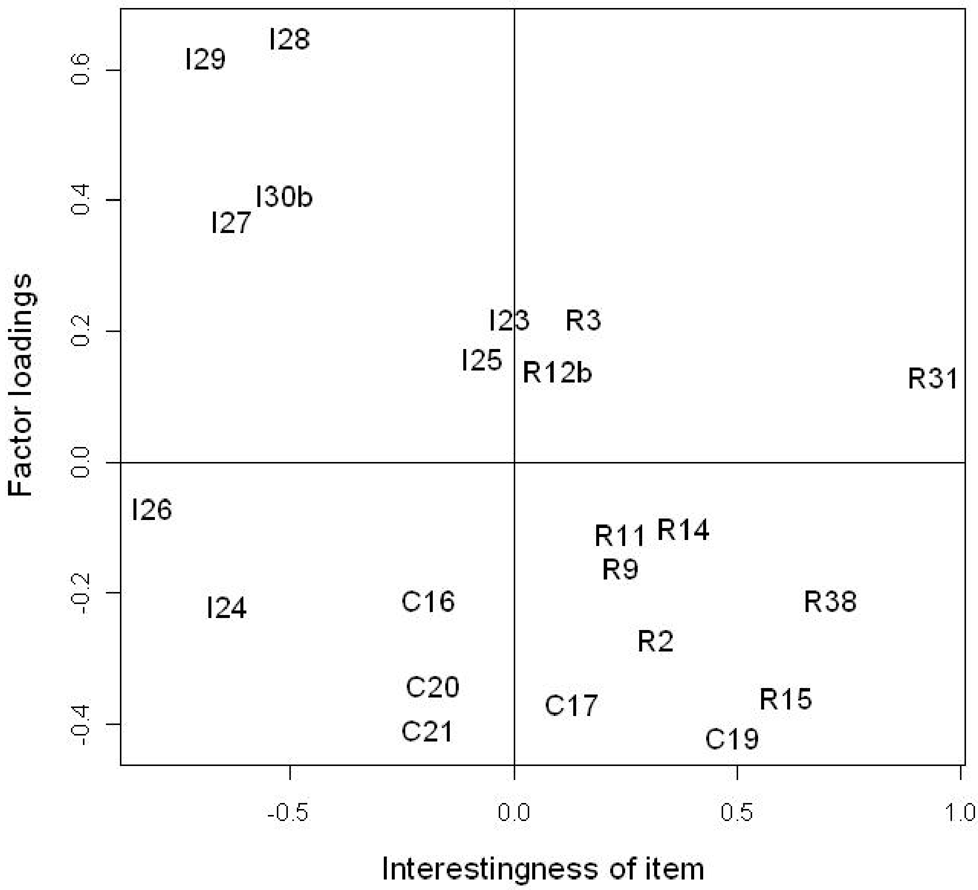


Figure 5.2. Factor loadings of residuals against item difficulties for SLIM

literacy are introduced in the mathematics syllabus, it was expected that their interest in statistical literacy should be positively associated with their interest in mathematics. The responses of students to the Mathematics Interest Inventory (MII) were analysed using the Rating Scale Model. These formed a measure that explained 82.7% of the variance and reported a person separation reliability of $R_p = .89$. Each of these 221 students were then assigned an estimated mathematics interest score. The strength of the linear association between these students' SLIM and MII scores was moderate ($r = .54$, $p = .00$).

Consequential evidence. In this instance items in SLIM were analysed for evidence of DIF by gender. Item difficulties were estimated on the basis of male responses and then on the basis of female responses. There were four items where the estimated difficulties differed significantly at the 5% level. Boys found more interest in working on problems involving data and statistics (item R3) and using data to investigate questions (item I30b). Girls, on the other hand found more interest in whether a survey about students was correct (item C20) and how to calculate the chance of injury from risky behavior (item I24). All item difficulties are plotted on Figure 5.3, which shows that on a whole-of-test basis there is little differentiation by gender.

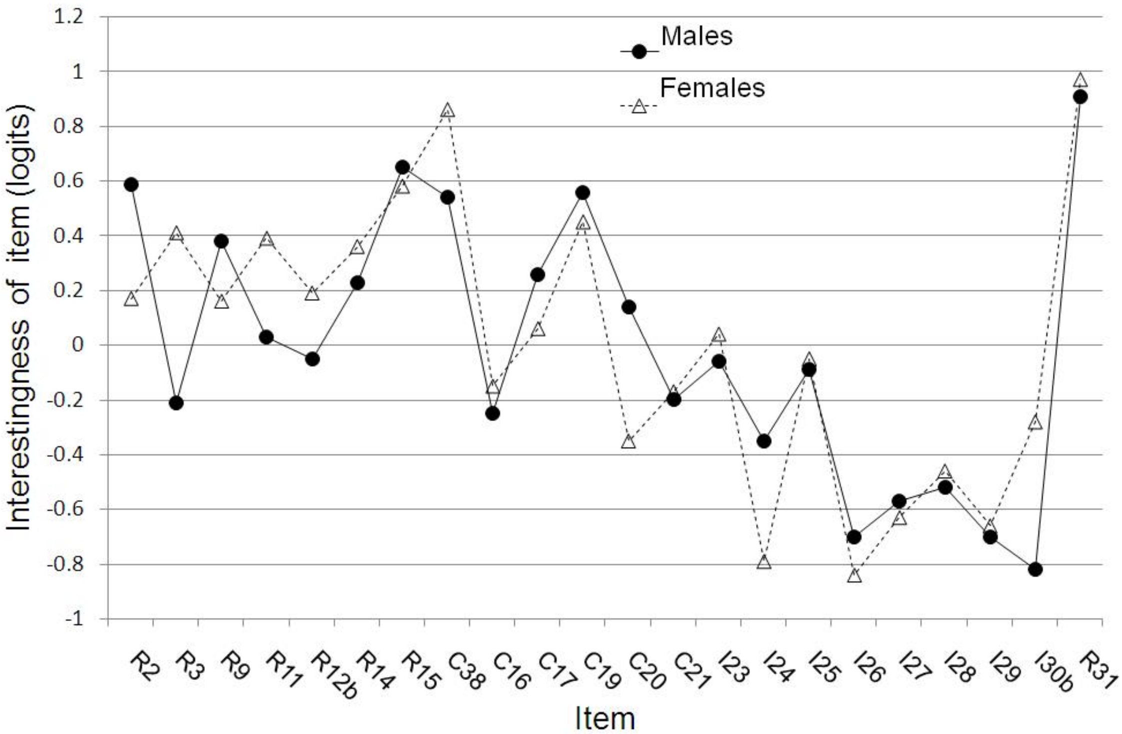


Figure 5.3. Interestingness of items based on male and female responses

Preliminary validity evidence for SESL

Content evidence. The relevance of SESL items was judged by the panel of experts. In regards the representativeness, Table 5.4 indicates that the items of SESL sample all the topics of statistical literacy, although there is a large proportion assessing data presentation. The items' fit-statistics reported in

Table 5.6 are satisfactory, supporting their technical quality.

Substantive evidence. Internally, the hierarchical structure of the SESL should reflect the statistical literacy hierarchy as described by Watson and Callingham (2003). The most difficult item reported in Table 5.6 is confidence to "find when a newspaper has used the wrong average" (item S42). Such a task requires students to engage critically with a media context and would thus fall in the upper levels of the statistical literacy hierarchy. Similarly, the second most difficult item is confidence to "explain when conclusions based on surveys are wrong" (item S47b). Although no context is provided, this task should also require students to engage critically with a statistical message and should consequently fall in the upper levels of the associated statistical literacy hierarchy. At the other end of the scale, arranging data into tables (item S48b) and showing data correctly on a bar chart (item S44) reflect an ability to master basic statistical concepts and skills. Almost all students in this age group should have encountered bar graphs and tables and this is reflected in their confidence towards these items. The items in the middle section of the self-efficacy scale reflect early levels of statistical literacy and require students to interpret, sometimes critically, statistical messages. Developmentally it is expected that students should become more confident towards statistical literacy as they progress through the middle school, gaining more exposure to statistical concepts. In this sample, however, there was no evidence of any association between students' ages and their self-efficacy scores. As reported, the expected positive association between self-efficacy and interest was evident for this sample.

Structural evidence. A PCA of the residuals was undertaken and a plot of the loadings against the item difficulties is shown in Figure 5.4. The random placement of these points, together with the large proportion of variance explained by the principal component support the unidimensionality of the construct.

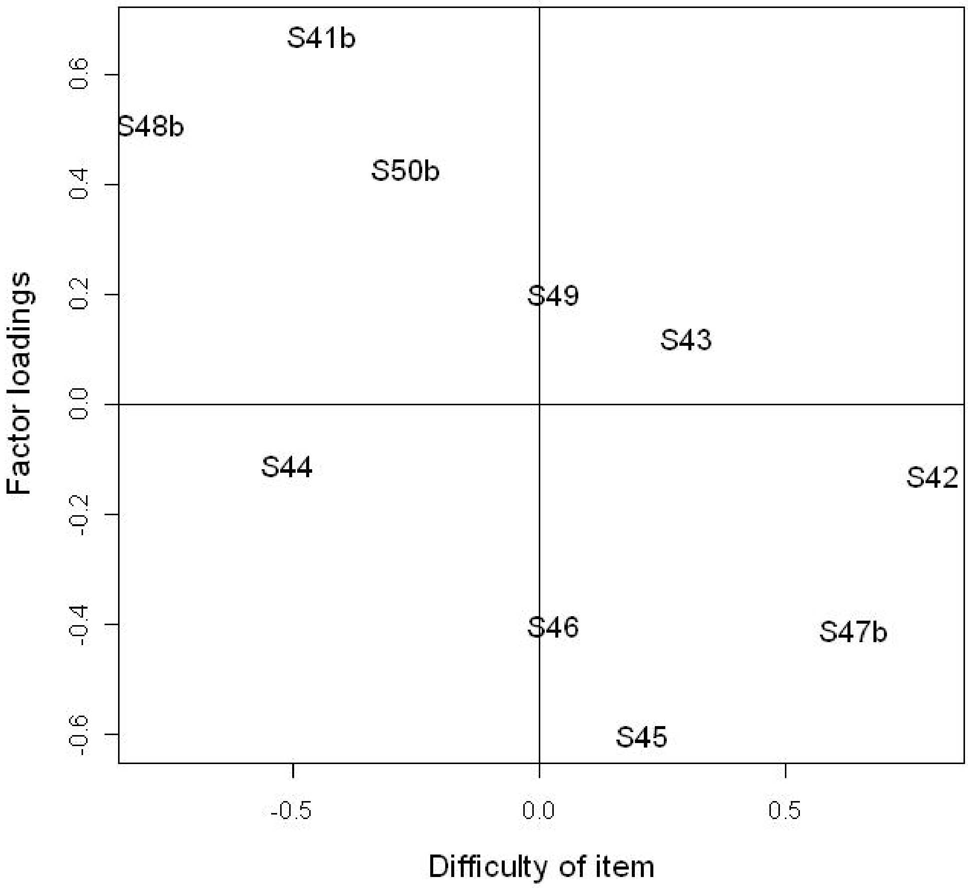


Figure 5.4. Factor loadings of residuals against item difficulties for SESL

Generalisability. Issues relating to the generalisability of the instrument are discussed later in the dissertation after the inclusion of additional data.

External evidence. As with students' interest, it was expected that their self-efficacy in statistical literacy would be positively associated with their self-efficacy in mathematics. The responses of students to the Motivated Strategies for Learning Questionnaire (MSLQ) were analysed using the Rating Scale Model. These were found to form a measure that explained 78.7% of the variance and reported a person separation reliability of $R_p = .92$. All students were thus assigned an estimated self-efficacy in mathematics score. The strength of the linear association between mathematics and statistical literacy self-efficacy scores was moderate ($r = .56$, $p = .00$).

Consequential evidence. In this instance items in SESL were analysed for evidence of DIF by gender. No items displayed evidence of DIF and as shown on Figure 5.5, the test functions the same for both genders.

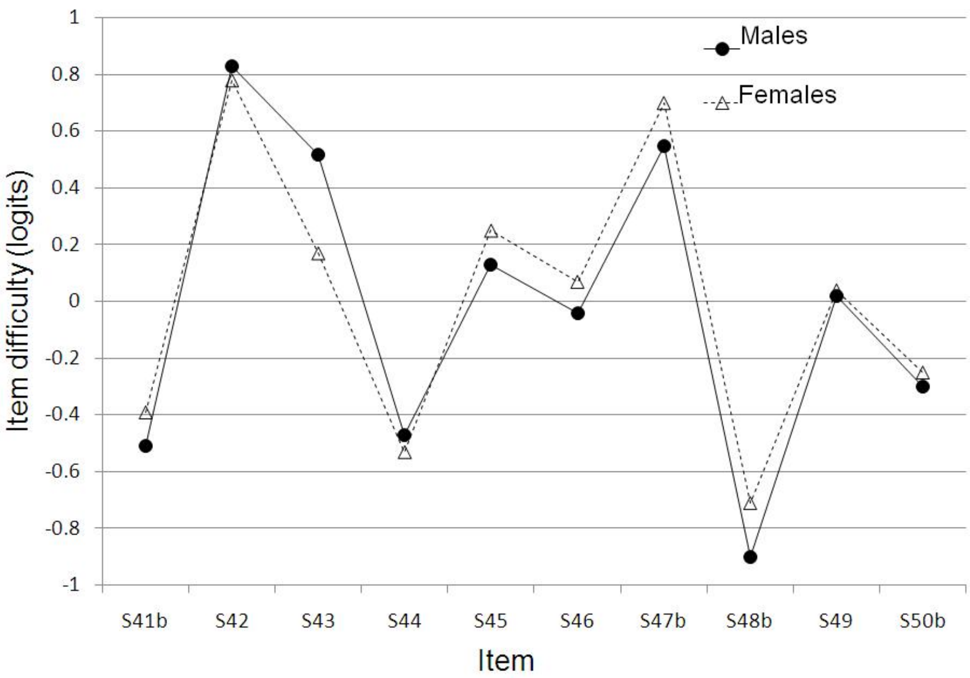


Figure 5.5. Difficulty of items based on male and female responses

5.6 Chapter summary

In the chapter the procedures used to develop the proposed measures of interest and self-efficacy in statistical literacy were reported. The chapter commenced with a theoretical review that outlined the types of validity evidence required to support interpretations made from these instruments. The subsequent procedures used to develop the two instruments were then based on the review.

As a result of the need to establish content validity for both instruments, a major part of the chapter was devoted to the establishment of theoretical models that described the internal structure of the two traits under consideration. Banks of items were constructed on the basis of these models and subsequently assessed by an expert review. The resulting items were then analysed using the Rasch Rating Scale model and interval measures of interest and self-efficacy were proposed.

In the last section of the chapter, the six forms of validity evidence outlined in the theoretical review were addressed for each of the two proposed measures. The evidence presented, although preliminary, supports the validity of interpretations that are to be made from these instruments. The establishment of validity, however, is an argument that requires a research program rather than a single empirical study (Kane, 2006). For this reason, issues relating to the validity of the two measures are addressed again in the next chapter and are based on additional data collected from a larger sample of students.

Chapter 6

Study results

The discussion in this chapter reports the results of the study. The presentation of these results reflects the quantitative analysis overview represented in Figure 4.1 of Chapter 4. It commences with results relating to the construction of the interest and self-efficacy scales and in particular validity evidence for these two scales. The discussion then addresses each of the study research questions, which in turn relate to the data modelling stage of the analysis overview.

6.1 The Statistical Literacy Interest Measure

The 22 items developed during the pilot were tested on the 570 students in the second and final stages of the study. Four of the items reported significantly high underfit and these are shown in Table 6.1. All items assessed reflective interest and arguably students in this sample had more exposure to the contexts associated with each item than students in the pilot, with the contexts eliciting more extreme interest responses. Students' responses to item R12b that assessed an interest in "using computer programs to help me investigate problems involving data" may have been more influenced by the computer context than the investigation of data. Alternatively, the students in the pilot study may not have had sufficient experience with the context and thus answered generically. In any case and as reported in Section 4.5, it was decided to conduct the subsequent analysis on the pooled data from all students in the study so that the sample would be more representative of the Australian middle school population.

Using the iterative approach described in Section 5.4 and based on the pooled sample, 16 items were found to form a parsimonious measure of interest. This measure explained 66% of the variation in student responses and reported

Table 6.1

Interest items displaying misfit

Item ID	Description	N	u_i	Z_u	v_i	Z_v
R2	Surveys about how people feel	554	1.43	6.7	1.92	9.9
R9	Reading graphs in the media	553	1.33	5.3	1.42	5.6
R12b	Using computers to investigate data	552	1.28	4.7	1.34	5.1
R31	Experiencing flow	541	1.37	5.3	1.99	9.9

a person separation reliability of $R_p = .88$. The specific items, number of valid responses (N), item difficulty estimates (δ_i), and infit statistics (u_i), are shown in Table 6.2, where they are ordered by difficulty. Other relevant item statistics are reported in Table B.3 of Appendix B. The estimated category thresholds (τ_k) were: -1.44, -0.46, 0.41 and 1.48. These are ordered and reasonably well separated, suggesting that the five category structure used in the instrument is satisfactory (Linacre, 1999). Additional category statistics are reported in Tables B.4 and B.5 of Appendix B.

One of the benefits of using the Rasch measurement model is that both the interest level of students and the interestingness of items can be placed on the one scale. Figure 6.1 shows this information. The first column of the figure shows the logit scale, whereas the second shows the interest level of students, which ranges from approximately -4.0 logits up to 2.6 logits. The third column of the figure shows the four thresholds for each item, one less than the number of Likert categories. The threshold denoted R15.3, for example, is the point on the scale where there is an equal probability of students giving a response of 2 or 3 to item R15. Also shown on this figure, are the locations of the mean student score (M) on the logit scale and also the location of one standard deviation (S) and two standard deviations (T) on this scale. Similar markings are shown on the item side of the scale.

Table 6.2

Items constituting the Statistical Literacy Interest Measure

ID	Item	N	δ_i	SE(δ_i)	u_i
R15	Getting a job that involves statistics.	766	0.76	0.04	1.14
C38	All there is to know about statistics.	633	0.53	0.04	1.05
C19	How politicians make decisions that are based on data.	771	0.43	0.04	0.99
R14	Learning more about statistics.	772	0.42	0.04	0.87
R3	Working on problems involving data and statistics.	772	0.39	0.04	1.03
C17	How a survey can be used to predict who will win the next election.	770	0.09	0.04	1.13
C16	How scientists calculate the chance of rain.	772	0.00	0.04	1.19
C20	Whether a survey reported on the radio or TV about students was correct.	774	-0.05	0.04	1.11
I23	Can understand news reports that use averages.	773	-0.05	0.04	0.89
I25	Understand the words that are used in statistics.	765	-0.07	0.04	0.79
I24	Know how to calculate the chance of being injured from risky behavior.	773	-0.25	0.04	1.15
I26	Can believe scientific claims that are based on data.	769	-0.33	0.04	1.05
I30b	Can use data to investigate questions that I might have.	714	-0.35	0.04	0.92
I28	Can understand graphs that appear on the internet or in newspapers.	772	-0.46	0.04	0.86
I27	Use the correct graph when displaying my data.	767	-0.51	0.04	0.94
I29	Can arrange data into tables.	771	-0.54	0.04	0.97

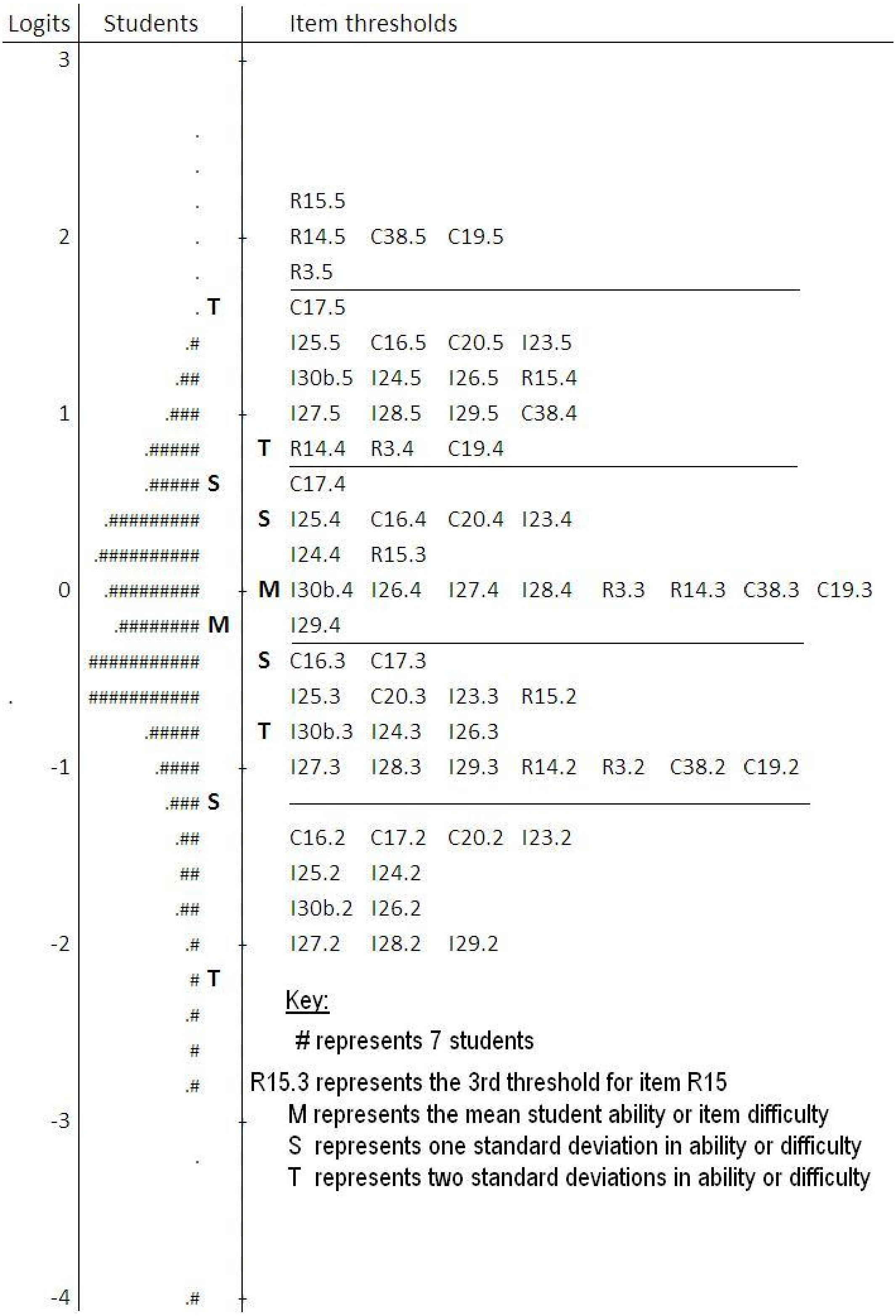


Figure 6.1. Wright map for SLIM

Figure 6.1 shows a series of horizontal lines that represent the location of natural breaks in the ordering of thresholds. In this instance the largest break in item difficulty occurred between items R3 and C17, where the difference in difficulties was 0.3 logits. Given that the mean standard error of person interest scores was also 0.3 logits, smaller partitions were not considered in the overall hierarchy. In this way the hierarchy of thresholds can be partitioned into five broad bands. The lower of these extends downwards from approximately -1.2 logits and includes students with very low levels of interest for statistical literacy. The second lowest band ranges from approximately -1.2 logits through to -0.3 logits and includes students with below average to average levels of interest. The third lowest band ranges from approximately -0.3 logits through to 0.6 logits and includes students with average to above average levels of interest. The second highest band ranges from approximately 0.6 logits through to 1.7 logits and includes students with high levels of interest. The highest band extends upwards from 1.7 logits and includes those few students with very high levels of interest.

Content evidence

The initial paneling process and subsequent refinement of items, described in Chapter 5, contributed to their relevance. As is seen from Table 6.3, the items comprising SLIM are representative, in that they sample all interest elements and most learning contexts associated with statistical literacy. The three reflective interest items, however, assess only general contexts and this feature is discussed further in the next chapter. The satisfactory fit of items in SLIM is evidence for the their technical quality (Wolfe & Smith, 2007b).

The items comprising SLIM appear to span the interest scale adequately. The location of items on the interest scale, as shown in Figure 6.1, does suggest the need for further item development in the lower reaches of the scale.

Table 6.3

Cross-classification of SLIM items by interest and content element

Content element	Interest element: Importance	Curiosity	Reflective
		Topic	
Sampling		C20	
Graphs	I27, I28, I29		
Averages	I23		
Chance	I24	C16,	
Inference	I26	C17, C19	
General	I25, I30b	C38	R3, R14, R15
		Context	
Technology			
School/class	I27, I29, I30b		
Media	I23, I26, I28	C20	
Social issues	I24	C16, C17, C19	
Sports			

Substantive evidence

The discussion in this section reports how student responses to SLIM align with the internal, or operational model that was outlined in Chapter 5. It also reports how changes in student responses to SLIM compare with expected developmental changes in interest during middle school. The discussion in Section 6.4 then reports how student responses to SLIM align with the external model proposed in Section 3.4.

In relation to the internal model the analysis in this instance focuses on agreement between the observed and expected hierarchy of item difficulties. The ordering of items, shown in Table 6.2, shows a hierarchical structure to the estimated difficulties of the items within SLIM. As expected, students found it easier to endorse items assessing importance interest than those assessing reflective interest. Of the importance interest items, students found it easiest to endorse the importance of correctly displaying their data. For example, being able to "arrange data into tables" (item I29) and using "the correct graph when displaying my data" (item I28) were the two easiest items. Such items are likely to assess students' valuing of task mastery, and accordingly represent low levels of interest. Items that assessed the importance of using statistical literacy in wider contexts, such as knowing "how to calculate the chance of being injured from risky behavior" (item I24) and being able to "understand news reports that use averages" (item I23) were more difficult for students to endorse. At the other end of the hierarchy, it was expected that students would find it most difficult to endorse a desire to re-engage with statistical literacy, as such views represent very high levels of interest. The most difficult item was an interest in "getting a job that involves statistics" (item R15) and the second most difficult item a desire to know "all there is to know about statistics" (item C38). It was also expected that students would find it easier to endorse an interest in the situation, such as "working on problems involving data and statistics" (item R3). As expected, most curiosity interest items were of less interest than the reflective interest items. The exceptions were a desire to know "all there is about statistics" (C38), which actually assesses re-engagement, and a desire to know "how politicians make decisions that are base on data" (item C19), suggesting that students of this age have little desire to engage in political contexts. Broadly, therefore, the observed hierarchy of item difficulties aligned with the theoretical hierarchy of interest elements.

Theories of adolescent development, discussed in Section 2.3, predict that students' interest will decline as they progress through their middle school education. Controlling for self-efficacy, the partial correlation between age and interest was negative ($r = -.10$, $p = .01$). This is as expected and provides further substantive validity evidence for SLIM.

The Model of Domain Learning, discussed in Section 3.2, predicts that increased levels of interest in statistical literacy will accompany increased levels

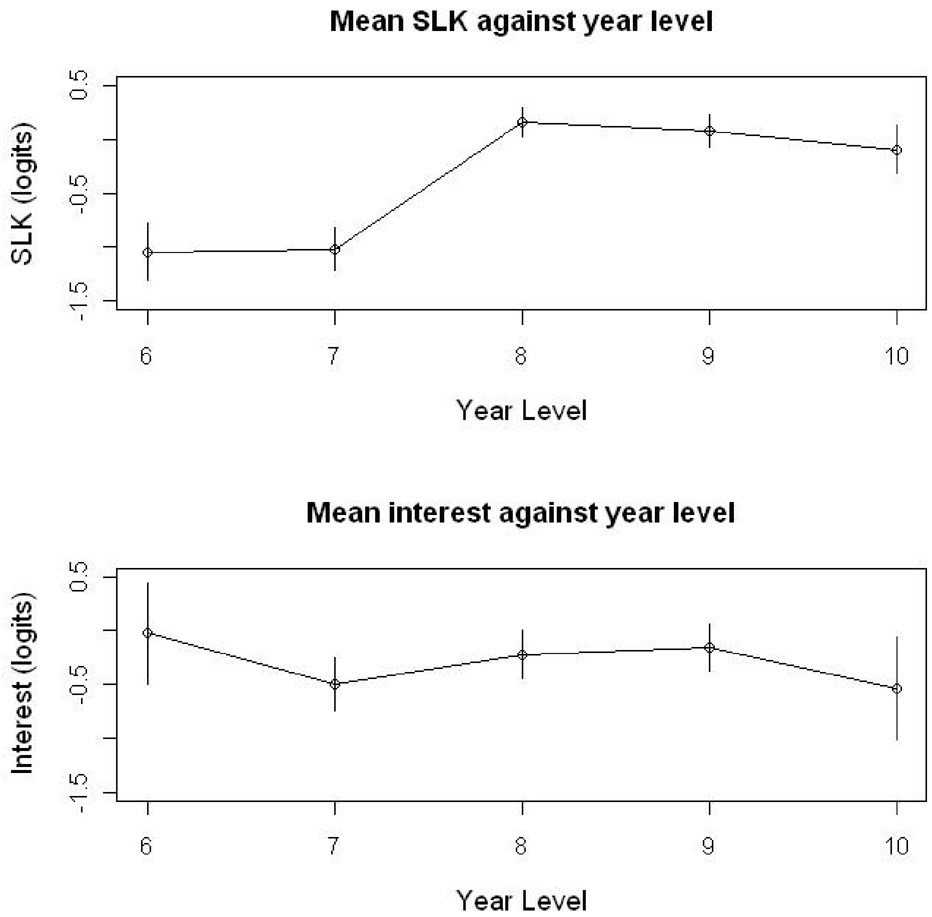


Figure 6.2. Comparison of SLK and interest by year level

of knowledge in the domain. The top plot in Figure 6.2, displays mean levels of statistical literacy knowledge scores by year level together with 95% confidence intervals for 295 students in Years 6 through to 10 for whom both SLK and Interest scores were available. As is seen from this plot, levels of SLK appear to increase significantly between Years 7 and 8, but from there remain relatively constant. The bottom plot in Figure 6.2 shows the mean levels of Interest by year level for the same sample. As is seen from this latter plot, the expected increase in Interest that should have accompanied the increase in SLK between Years 7 and 8 did not occur, although there was a slight non-significant increase in Interest up until Year 9.

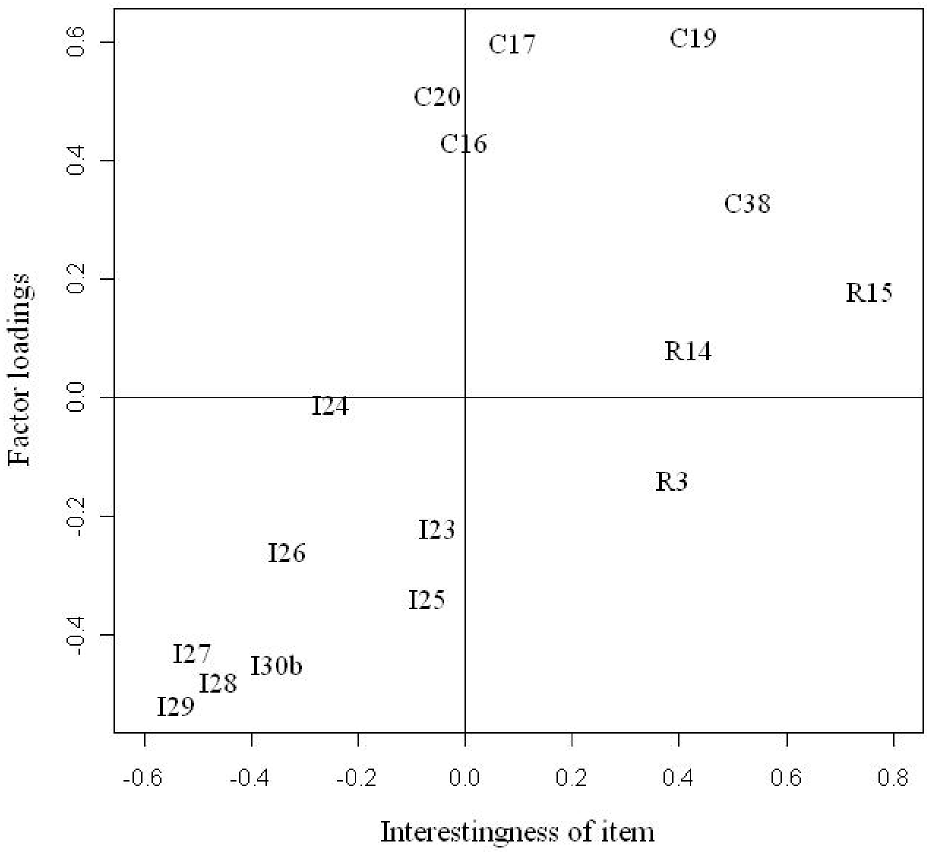


Figure 6.3. Factor loadings of residuals against item difficulties for SLIM items

Structural evidence

The major assumption of the Rasch model is that the latent trait is unidimensional. A plot of the loadings against difficulties, as shown in Figure 6.3, suggests some structure in the residuals with all of the importance items grouped together in the lower left-hand quadrant. In addition to this, the eigenvalues of the first dimension and its contrasts are reported as $\lambda = \{2.5, 1.8, 1.3, 1.2, 1.1\}$. The first two of these exceed the recommended minimum of 1.4 (Smith & Miao, 1994), suggesting the presence of multiple dimensions.

Given this apparent structure in the residuals, it was decided to test the data for evidence of multidimensionality. An exploratory factor analysis, details of which are reported in Table B.6 of Appendix B, suggested the presence of three factors aligning with the elements of interest and therefore the three item stems. In order to test for unidimensionality, a multidimensional Rasch model was applied to the 16 items, with the reflective items assigned to the first dimension, the curiosity items to the second dimension, and the importance items to the third dimension. In comparison to a unidimensional model, the application of the three dimensional model improved model fit. Based on a comparison of deviance test (Wu & Adams, 2006) this improvement was statistically significant ($\chi^2_5 = 578$, $p = 0.00$). Thus the evidence suggests the presence of three dimensions, although these are highly correlated with all correlations exceeding .75.

The apparent multidimensionality may be more related to the structure of the questionnaire than the actual interest construct. Curtis and Boman (2007) argued that the use of the same stem for several items can induce local independence and thus apparent multidimensionality. Further testing of the measure needs to occur using the same items but arranged in a different order. In any case, the high correlations between the three dimensions lend support for a single higher order factor (Thompson, 2004), one that arguably assesses a broad valuing of statistical literacy.

Evidence of generalisability

A simple test of the generalisability of the measure is to examine the invariance of item difficulty estimates between two samples of students (Smith, 2001). In this instance differential item functioning (DIF) of items was assessed by gender, year level at school, and attendance at a StatSmart school.

DIF by gender. Figure 6.4 shows the item difficulty estimates for males and females. Statistically significant differences at the 5% level, after application of the Bonferroni adjustment, are marked on the graph. Males found it easier to endorse an interest in "working on problems involving data and statistics" (item R3), whereas females found it easier to endorse finding out "how a survey can be used to predict who will win the next election" (item C17) and "whether a survey reported on the radio or TV about students was correct" (item C20). Further details of this analysis can be found in Table B.7 of Appendix B.

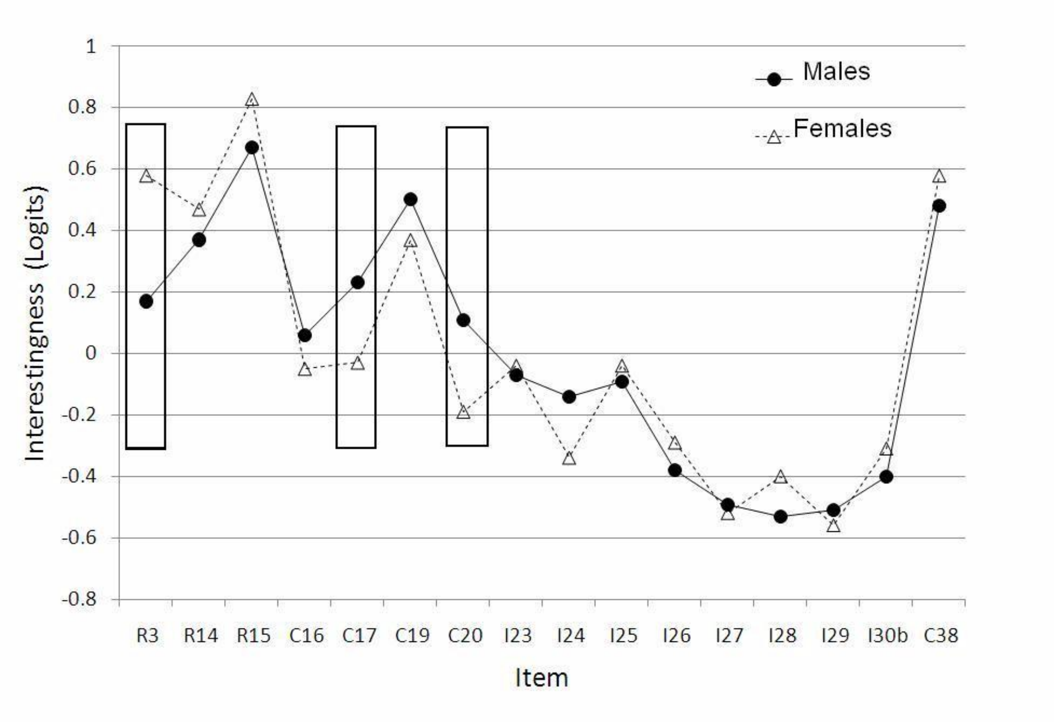


Figure 6.4. SLIM DIF by gender

DIF by year level at school. Figure 6.5 shows the item difficulty estimates for students in Years 7, 8, 9 and 10. Only one class of 23 students was in Year 6 and its results were omitted from this analysis. Statistically significant differences at the 5% level, after application of the Bonferroni adjustment, are marked on the graph. Significant differences by year level were evident in relation to two items. Year 7 students found it easier to endorse an "interest in getting a job that involves statistics" (item R15) than older students. On the other hand, they found it harder to endorse the importance of believing

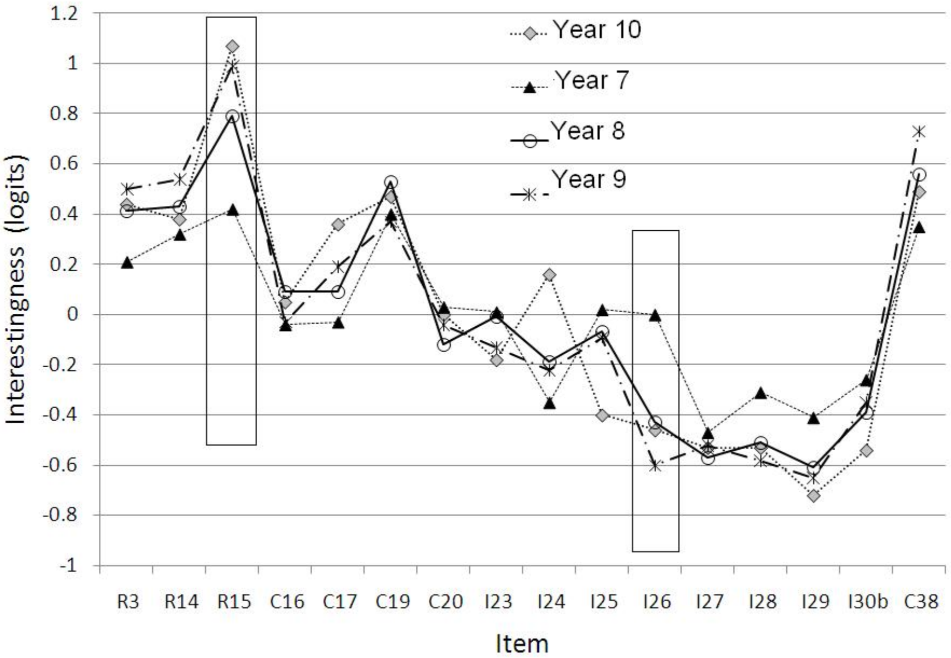


Figure 6.5. SLIM DIF by year level at school

"scientific claims that are based on data" (item I26) than the other year levels. Further details can be found in Table B.8 of Appendix B.

DIF by attendance at StatSmart school. Figure 6.6 shows the item difficulty estimates for students attending StatSmart and Non-StatSmart schools. Statistically significant differences at the 5% level, after application of the Bonferroni adjustment, are marked on the graph. Students attending Non-StatSmart schools, found it harder to endorse the importance of understanding "news reports that use averages" (item I23). Students attending StatSmart schools, on the other hand, found it harder to endorse the importance of knowing "how to calculate the chance of being injured from risky behavior" (item I24) and believing "scientific claims that are based on data" (item I26). Further details of this analysis are reported in Table B.9 of Appendix B.

Summary. While several items displayed evidence of DIF for different subgroups of students, Figures 6.4, 6.5, and 6.6 show that on a whole-of-test basis, the instrument appeared to perform in the same way for most students.

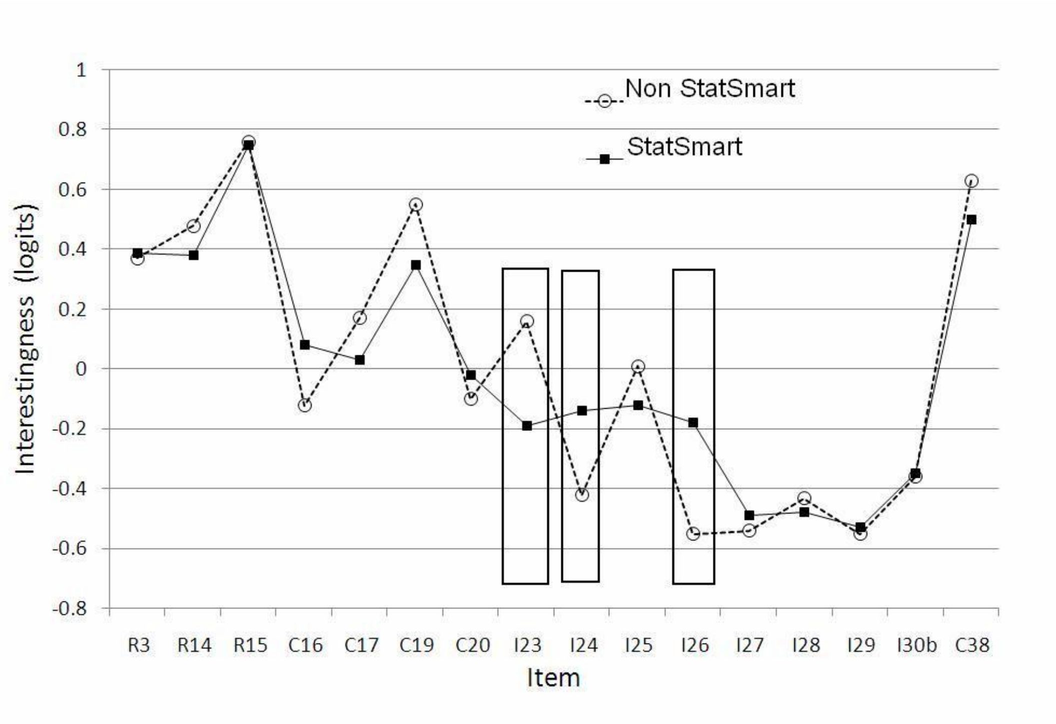


Figure 6.6. SLIM DIF by attendance at StatSmart schools

External evidence

The results of the pilot study indicated that students' interest in statistical literacy is positively associated with their interest in mathematics. Although this version of SLIM has six less items than that used in the pilot, the correlation between the two measures for that sample of students is still positive (r = .57, p = .00).

Consequential evidence

The results reported earlier on the analysis of item DIF suggest that a small number of items in SLIM differentiated between groups of students on the basis of gender, year level at school, and/or their experience with statistics. On a whole-of-test basis, however, the impact of this is likely minimal, in that although some subgroups tended to favor one item others tended to favor alternative items. In any case it is possible to adjust interest scores, post-hoc, to cater for such DIF (Bond & Fox, 2007).

6.2 The Self-Efficacy for Statistical Literacy scale

The ten items developed for SESL during the pilot were re-analysed on the basis of responses from students in the pooled sample. They collectively formed an interval measure of self-efficacy in statistical literacy that explained 71% of the variance in student responses and reported a person separation reliability of $R_p = .84$. All items displayed satisfactory fit, although confidence "to work out the most likely outcome from a game involving chance" (item S50c) reported evidence of underfit with standardised infit and outfit values exceeding 3.0. Given the need for more items assessing chance and the fact that both mean square values were within accepted limits, it was decided to retain this item.

The specific items for SESL, number of valid responses (N), item difficulty estimates (δ_i), and infit statistics (u_i), are shown in Table 6.4, where they are ordered by difficulty. Other relevant item statistics are reported in Table B.10 of Appendix B. The estimated category thresholds (τ_k) were: -1.71, -0.60, 0.47 and 1.83. These are ordered and well separated, suggesting that the five category structure used in the instrument is satisfactory (Linacre, 1999). Additional category statistics are reported in Table B.11 of Appendix B.

Content evidence

As reported in Chapter 5, the initial panelling process and subsequent refinement of SESL items contributed to their relevance. In addition to this the ten items sample each of the identified topics of statistical literacy. The Wright map, shown in Figure 6.7, indicates that the items of SESL adequately span the self-efficacy scale. The reported fit statistics are all within the accepted range, thus providing evidence for the technical quality of the items.

Table 6.4

Items and selected statistics for SESL

ID	Item (Confidence to solve:)	N	δ_i	u_i
S42	Find when a newspaper has used the wrong average.	783	0.75	0.93
S47b	Explain when conclusions based on surveys are wrong.	645	0.50	0.80
S43	Explain to a friend how probability is calculated.	785	0.14	1.01
S45	Explain the meaning of a graph in a newspaper.	781	0.07	0.91
S46	Find a mistake in someone else's graph.	783	0.06	0.98
S48c	Look up the correct number from a table of numbers.	419	0.05	0.96
S49	Explain how to select a fair sample for a school survey.	783	-0.07	1.03
S50c	Work out the most likely outcome from a game involving chance.	423	-0.41	1.27
S41b	Solve problems that use averages.	646	-0.48	1.11
S44	Show data correctly on a bar chart.	785	-0.61	1.10

Substantive evidence

The substantive evidence presented in this section relates primarily to the internal or operational model that was described in Section 5.1. In addition to this, evidence regarding the relationship between self-efficacy and external constructs is presented, as is evidence regarding developmental aspects of self-efficacy.

The hierarchical structure of SESL, reported in Table 6.4, has in the main remained the same as that reported in the pilot study. Consequently the position of items still reflects the statistical literacy hierarchy, as identified by Callingham and Watson (2005). The inclusion of the two additional items, however, warrants further discussion. Confidence to "work out the most likely outcome from a game involving chance" (item S50c) reflects an ability to master basic tasks associated with statistical literacy and does not involve

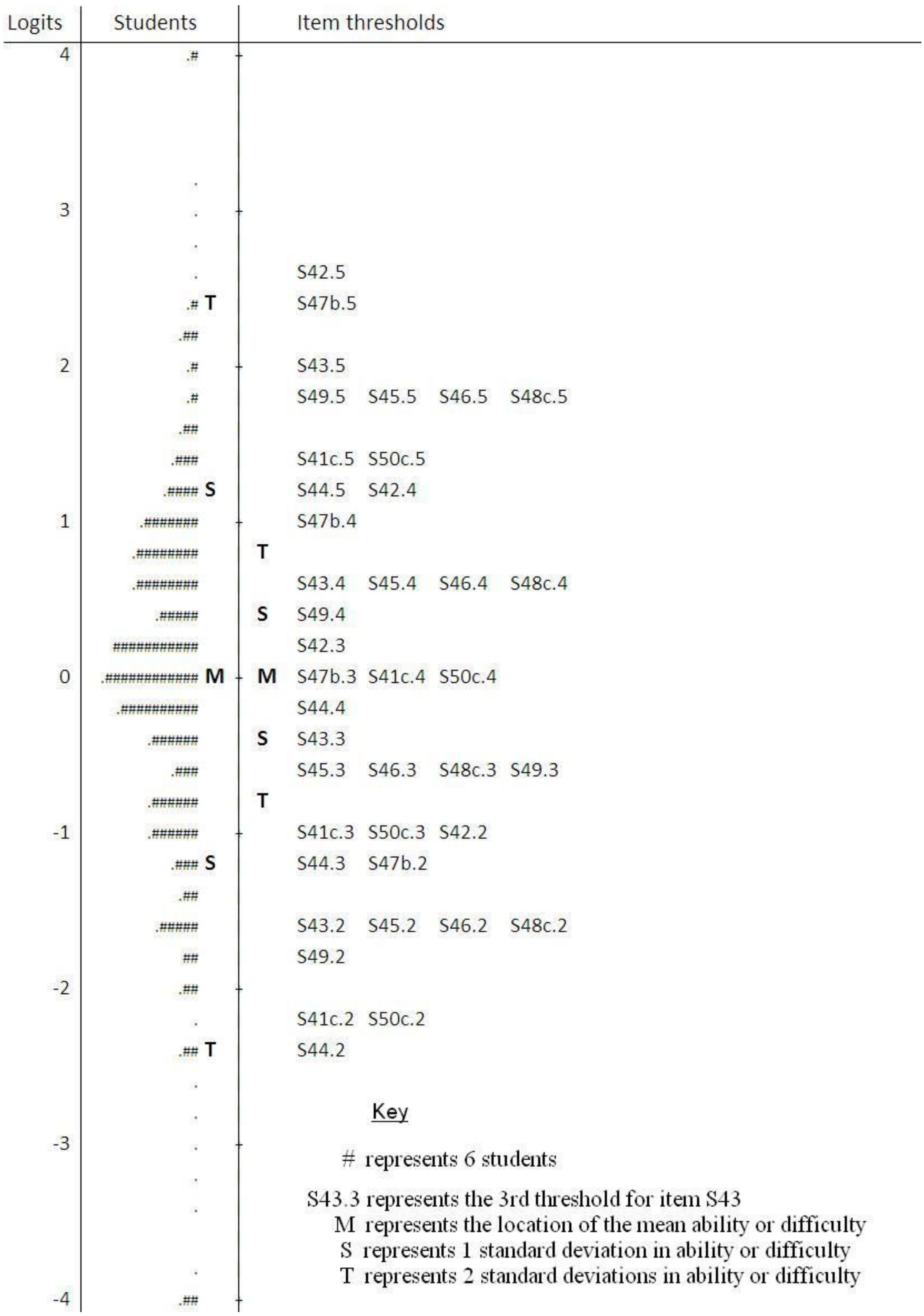


Figure 6.7. Wright map for SESL

explaining or critically analysing. Its placement near the bottom of the hierarchy is as expected. It is surprising, however, that "looking up the correct number from a table of numbers" (item S48c) was not regarded as the easiest item. Arguably this item is not specific enough and should ideally provide more details about the table.

Externally, self-efficacy is known to be strongly associated with achievement. Based on the 452 students for whom mathematics achievement was reported, an analysis of variance (ANOVA) was undertaken of the variable self-efficacy using RelMaths-grade as the factor. This indicated a significant association between the two variables ($F = 9.48$, $p = .00$). The mean self-efficacy score for students with a mathematics grade below the class median was significantly lower than the mean self-efficacy score of students with a mathematics grade above the class median.

Developmentally, it is expected that as students progress through the middle school they should encounter more and more statistical concepts, thus gaining self-efficacy in statistical literacy as they age. There was a significant, albeit weak, correlation between students' self-efficacy in statistical literacy and their age in years ($r = .11$, $p = .00$). The relative weakness of this association may indicate that students' self-efficacy beliefs are relatively stable during this period. In her longitudinal study, Watt (2005) reported that students' expectancies of success were quite stable during their middle school education, even showing a slight decline. Marcoulides et al. (2008) argued that any changes in a student's academic motivational state are more likely to occur during late childhood than adolescence.

Structural evidence

A factor analysis of the residuals was undertaken and a plot of these loadings against SESL difficulties is shown in Figure 6.8. The random positioning of the items on this plot suggests the presence of no structure in the residuals and confirms the unidimensionality assumption. Similarly, the reported eigenvalues of the first dimension and its contrasts are $\lambda = \{1.6, 1.4, 1.2, 1.2, 1.1\}$ which are close to or below the recommended minimum of 1.4 (Smith & Miao, 1994). More recently, however, Raiche (2005) questioned this recommended minimum and reported that the first eigenvalue often exceeds this value in random data and in many cases so does the second. The fact that the principal component explains 71% of the variance supports the presence of a single dimension.

The internal consistency of the measure, as estimated using Cronbach's alpha, is .91. This suggests that the items correlate closely and assess the same dimension, again confirming the unidimensional nature of the construct.

Evidence of generalisability

In order to assess the generalisability of SESL, differential item functioning was assessed by gender, year level at school, and attendance at a StatSmart school. After the application of the Bonferroni adjustment, no item in SESL displayed significant evidence of DIF, at the 5% level, for any of the three tests. Further details of these analyses are reported in Tables B.12, B.13 and B.14 of Appendix B.

External evidence

The results of the pilot demonstrated that students' self-efficacy in statistical literacy is associated with their self-efficacy in mathematics. The minor changes that were made to the instrument are unlikely to alter this finding.

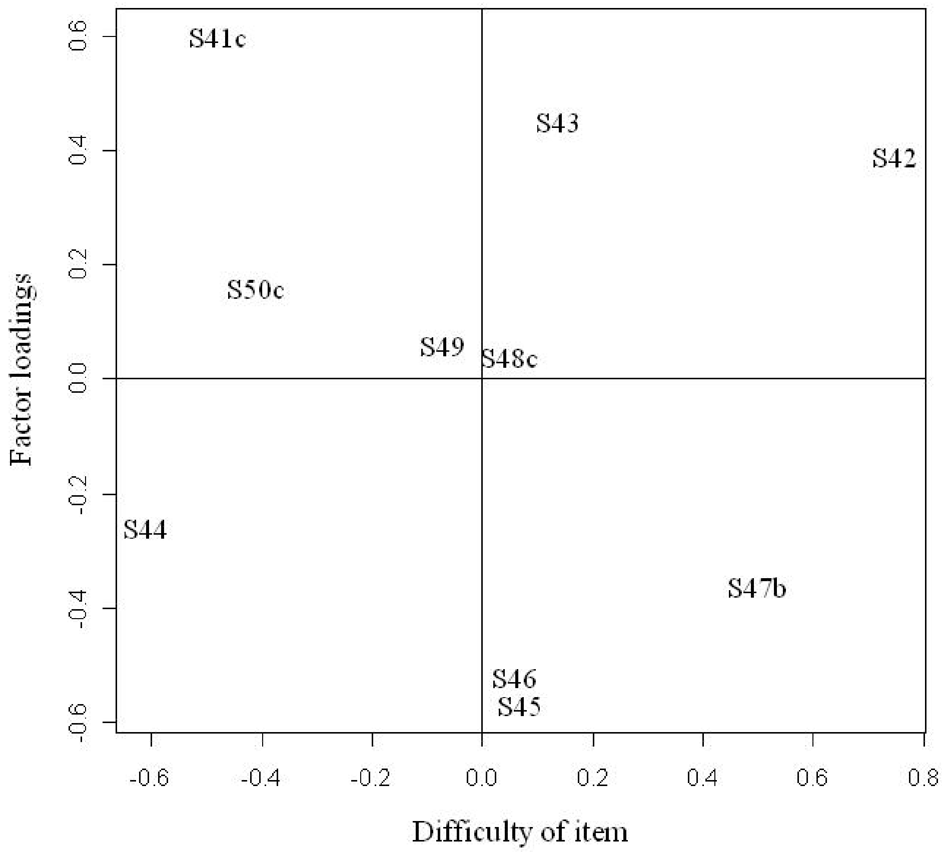


Figure 6.8. Factor loadings of residuals against item difficulties for SESL

Consequential evidence

The results of the DIF analysis, reported above, suggest that SESL is unlikely to differentiate between subgroups of the Australian middle school population. Again, the evidence suggests that the scale provides a valid measure of middle school students' self-efficacy in statistical literacy.

6.3 Results related to Research Question 1

> How valid is it to base a measure of middle school students' interest in statistical literacy on their responses to a series of interest self-descriptions?

The validity evidence presented earlier suggests that the 16 self-descriptions comprising SLIM form a valid interval measure of middle school students' interest in statistical literacy. The instrument explained approximately two thirds of the variance in students' responses and in the main conformed to the requirements of the Rasch measurement model. There was some indication of multidimensionality, although it is unclear whether this reflects the inherent complexity of the construct or is merely a statistical artefact. Further testing of the instrument is therefore required.

6.4 Results related to Research Question 2

> How do factors unique to an individual, such as their age, prior achievement, gender, and self-competency beliefs, contribute to their interest in statistical literacy?

This section commences with an exploration of relevant bivariate relationships between students' individual factors. It then explores how these individual factors interact using a series of regression models. This is done in order to test the hypothesised model shown earlier in Figure 3.2.

Bivariate relationships

This section reports the existence, or otherwise, of bivariate relationships between interest and a number of other factors relevant to the individual, including: gender, age, self-efficacy, prior achievement and knowledge. The findings reported in this section are based on the pooled sample and the nomenclature used for key variables is as described in Table 4.5. Unless stated otherwise, all reported statistically significant relationships are significant at the 5% level.

Gender. Overall, girls found statistical literacy slightly less interesting than boys with a reported 0.18 logits difference in mean levels of interest

($t = 2.24$, effect-size $d = 0.16$). Given that the average standard error associated with each person's interest score was approximately 0.3 logits, however, this difference may not be of any practical significance. Boys appeared to be slightly more self-efficacious towards statistical literacy than girls, with a reported 0.22 logits difference in mean levels of self-efficacy ($t = 1.99$, $d = 0.15$), but again this difference may also be of no practical significance. There was no apparent gender difference in mean levels of SLK.

Age. As reported, there was evidence of a weak negative association between students' ages and levels of interest, but only when controlling for self-efficacy. Similarly, there was evidence of a weak positive association between students' ages and self-efficacy. A positive linear association between students' ages and SLK was also evident ($r = .43$, $n = 295$).

Self-efficacy. A moderate linear association was evident between the variables Self-efficacy and Interest ($r = .62$, $n = 775$), which is similar in magnitude to the average value of .59 reported by Rottinghaus, Larson, and Borgen (2003) in their meta-analysis of 60 interest/self-efficacy studies. In addition to this, there was an association between Interest and the square of Self-efficacy ($r = -.28$, $n = 775$). This latter result supports Silvia's (2003) contention that students' self-efficacy will influence their interest quadratically, in that students are likely to have less interest in tasks if they are certain that they can complete them or if they are certain that they cannot complete them.

Mathematics achievement. Students' achievement in mathematics was positively associated with their interest in statistical literacy. An ANOVA of interest against Maths-grade, as described in Table 4.3, found that students with lower maths grades reported lower levels of interest than those with higher mathematics grades ($F = 9.94$). There was no significant association between the variables RelMaths-grade and Interest, which appears to contradict the findings of Trautwein et al. (2006) who found that students' achievement relative to their immediate peers was a predictor of their interest. Prior achievement in mathematics also influenced students' self-efficacy beliefs. An ANOVA of self-efficacy against Maths-grade found that students with lower mathematics grades reported lower levels of self-efficacy than students with higher maths grades ($F = 17.62$). Similarly there was a significant association between the variables RelMaths-grade and Self-efficacy ($F = 9.48$), in that students with mathematics grades higher than the class median were more likely to have higher levels of self-efficacy in statistical literacy.

Statistical literacy knowledge. There was some evidence of a weak association between Interest and SLK ($r = .11$, $n = 295$) although this was only significant at the 10% level. This is much lower than the average value of .31, reported by Schiefele et al. (1992) in their meta-analysis of 121 studies that examined the interest achievement relationship. The strength of this association may have been influenced by the temporal proximity of the two tests in that some students completed SLIM up to six months after they had completed the StatSmart tests. For students who completed both tests at the end of the first year of this study there was a higher association between Interest and SLK ($r = .27$, $n = 70$). The strength of the Interest/SLK association may have also been influenced by gender in that there was virtually no reported association for boys and a weak association for girls ($r = .20$, $n = 148$). This latter result contradicts the finding of Schiefele et al. (1992) who reported that the interest achievement association is stronger for boys than for girls.

Examining inter-relationships with linear models

Initially a simple linear regression model was applied to the data with the variable Interest as the response. As a means of catering for possible dependence between students in the one class and/or school, a mixed effects model was then applied to the data. Because neither of these linear models allows for the inherent measurement error in the response variable and for comparative purposes, a latent regression model was also used. The software available in the study, however, did not extend to hierarchical latent regression models. The modelling process was used to develop a path model that was then compared with the theoretical model shown in Figure 3.2.

Simple linear regression model with interest as the response. For this sample of students the only significant predictors of Interest were the variables Self-efficacy and Age. In addition to this, the square of Self-efficacy was also found to be a significant predictor of Interest. Given that measures of SLK and prior mathematics achievement were not significant predictors of the variable Interest, the model was re-applied to a larger set of 768 students for whom interest, age and self-efficacy scores were available. This is after the data from four influential outliers, two of whom were male, were removed. The model is shown as Equation 6.1, which displays the standard errors of each coefficient underneath it in brackets and the residual error as the term ε_{ij}. It explained 46% of the variance in student interest scores and diagnostic plots, shown in Figure C.1 of Appendix C, suggest that standard assumptions regarding the normality of residuals and homogeneity of residual variance have been met for the model.

$$\begin{aligned}\text{Interest} \;=\; & \underset{(0.36)}{0.61} - \underset{(0.03)}{0.06}\ \text{Age} + \underset{(0.02)}{0.45}\ \text{Self-efficacy} \\ & - \underset{(0.01)}{0.05}\ \text{Self-efficacy}^2 + \varepsilon_{ij} \end{aligned} \tag{6.1}$$

Mixed effects model with interest as the response. The above model was tested for both state and school random effects. The effects of class grouping was not tested, however, because class membership details were not available for 105 of these students. More specifically, the model was initially modified to include in the intercept term both random, state and school effects. Only school effects, however, contributed significantly to model fit. Following this, the model was then modified to include random school effects in both the intercept term and all coefficients. The inclusion of a random effect in the age coefficient, however, did not significantly improve model fit. The resulting mixed effects model, shown in Equation 6.2, reports better fit than the original linear regression model. The associated reduction in deviance was 15.37 on 6 degrees of freedom, which is statistically significant at the 5% level.

$$\begin{aligned} \text{Interest} &= \underset{(0.40)}{(0.61 + b_{0i})} - \underset{(0.03)}{0.06}\ \text{Age} + \underset{(0.03)}{(0.46 + b_{1i})}\text{Self-efficacy} \\ &\quad - \underset{(0.01)}{(0.05 + b_{2i})}\text{Self-efficacy}^2 + \varepsilon_{ij} \end{aligned} \tag{6.2}$$

The three random variables b_{0i}, b_{1i}, and b_{2i} model the variation due to the grouping of students by school. These are all assumed to be normally distributed with a mean of zero and with standard deviations of 0.14, 0.09 and 0.02 respectively. The random variable ε_{ij}, on the other hand, models the individual variation. Given that its standard deviation is reported as 0.82, the variation due to the grouping of students is small in comparison to the individual variation.

Latent regression model with interest as the response. In order to overcome the measurement error in the response variable, a latent regression model was also applied to the data. This model, shown in Equation 6.3, explained 47% of the variance in interest. As can be seen from Equations 6.1 and 6.3, not adjusting for measurement error in the response variable tended to inflate the magnitude of the model's coefficients.

$$\text{Interest} = \underset{(0.04)}{0.55} - \underset{(0.02)}{0.05}\ \text{Age} + \underset{(0.02)}{0.40}\ \text{Self-efficacy} - \underset{(0.01)}{0.04}\ \text{Self-efficacy}^2 \tag{6.3}$$

Simple linear regression model with self-efficacy as the response. Given the failure of mathematics achievement measures to predict interest, it was decided to investigate possible predictors of students' self-efficacy. In order to maximize statistical power, a linear regression model was applied to the responses of 427 students for whom prior mathematics achievement, self-efficacy and interest scores were available. This is after the data from nine influential outliers were removed, seven of whom were male. The resulting model, shown as Equation 6.4, explained 43% of the variance in self-efficacy. Diagnostic plots, shown in Figure C.2 of Appendix C, suggest that standard assumptions regarding the normality of residuals and homogeneity of residual variance have been met for this model. In this model, mathematics achievement is presented in terms of the variable RelMaths-grade, as defined in Tables 4.4 and 4.5.

$$\begin{aligned}\text{Self-efficacy} &= \underset{(0.55)}{-\,2.57} + \underset{(0.04)}{0.18}\ \text{Age} + \underset{(0.04)}{0.83}\ \text{Interest} + \underset{(0.11)}{0.30}\ \text{Median grade} \\ &\quad + \underset{(0.14)}{0.41}\ \text{Above median grade} + \varepsilon_{ij} \qquad (6.4)\end{aligned}$$

The development of a path model

The results of the linear models reported in this section lend support for the hypothesised quadratic relationship between self-efficacy and interest. They also support the hypothesised influence of prior achievement on self-efficacy. They do not, however, support the presence of a direct link between prior achievement and interest. In regards to the hypothesised influence of individual factors, age had a negative influence on interest but a positive influence on self-efficacy. In the presence of other factors, the influence of gender was negligible. The influence of the teacher and/or school, in the form of how long students participated in the StatSmart project, was not a significant predictor of either interest or self-efficacy. These findings suggest that the path model, shown in Figure 6.9, is a more accurate representation of the data than the hypothesised model, shown as Figure 3.2 in Section 3.4.

The path model shown in Figure 6.9 was tested using AMOS and estimated path coefficients and covariances are also shown on this figure, whereas the full model, including both structural and measurement components, is shown as Figure 6.10. When a direct path from mathematics achievement to

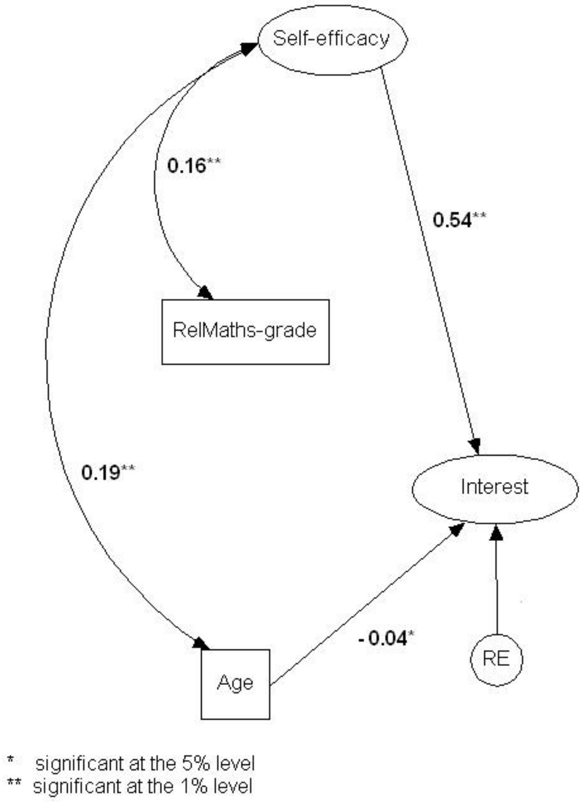


Figure 6.9. Path model showing antecedents of students' interest in statistical literacy

interest was included in the model the resulting path coefficient was 0.01 and not statistically significant. It is not possible to test the hypothesised quadratic link between self-efficacy and interest using a structural model and software limitations prevented the application of multilevel path models to these data.

Reported model fit statistics for the structural model, shown in Figure 6.10, were within acceptable limits ($CFI = 0.971$, $RMSEA = 0.055$), providing support for the path model that was developed through the application of linear models to the data.

Students' frame of reference

Students' use of an external or internal frame of reference (FoR) was analysed through their responses to two items, shown as items IE42 and IE43 in

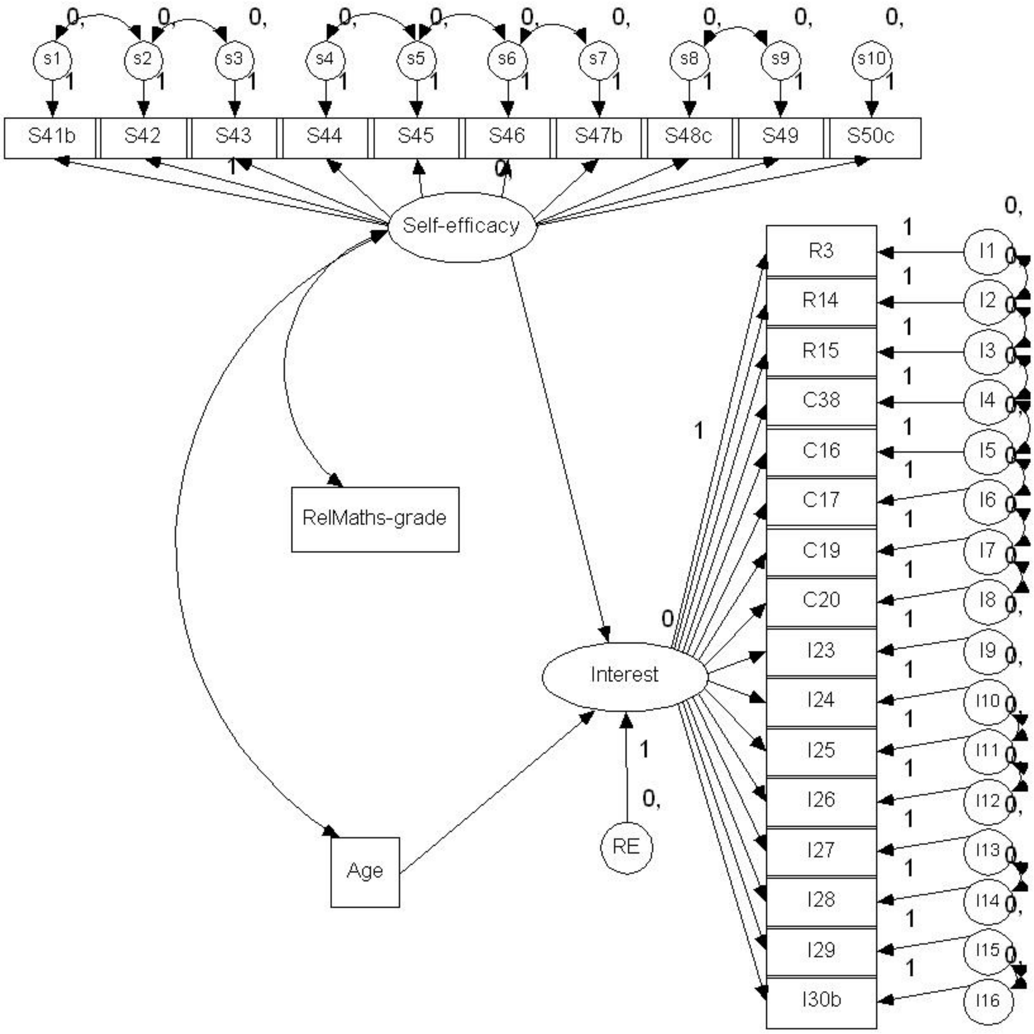


Figure 6.10. Measurement and structural components of interest model

Appendix A. "Compared to others in my class I am good at maths" (item IE42) assessed students' use of an external FoR, whereas "out of all my subjects I usually get my best marks in maths" (item IE43) assessed their internal FoR. In order to carry out meaningful cross-tabulations the existing five categories on the Likert scale were collapsed into three categories, grouping the higher two together and the lower two together. A cross-tabulation of both items with the new category structure is shown in Table 6.5.

Students' responses to these items were associated with external measures of their achievement. Table 6.6 shows a cross-tabulation of RelMaths-grade against external and internal FoR. Most students who attained below median

Table 6.5

Internal against external frame of reference assessment

		Internal reference (IE43)			
		Negative	Neutral	Positive	Total
	Negative	128	22	6	156
External	Neutral	38	35	30	103
reference	Positive	15	34	98	147
(IE42)	Total	181	91	134	406

grades assessed themselves as negative on the external FoR and similarly most students who attained above median grades assessed themselves as positive on the external FoR. A chi-square test of association between RelMaths-grade and external FoR was significant ($\chi^2 = 48.30$, $p = .00$). In relation to the internal FoR, most students who attained below median grades assessed themselves negatively and most students who attained above median grades assessed themselves positively. A chi-square test of association between RelMaths-grade and internal FoR was also statistically significant ($\chi^2 = 39.00$, $p = .00$).

Table 6.6

RelMaths-grade against external and internal FoR

		RelMaths-grade			
		Below median	Median	Above median	Total
	Negative	62	68	21	151
External	Neutral	30	53	16	99
reference	Positive	12	76	49	137
	Total	104	197	86	387
	Negative	70	86	24	174
Internal	Neutral	19	46	21	86
reference	Positive	13	71	41	125
	Total	102	197	86	385

In regard to interest, students who had positive or neutral assessments on either or both frames of reference tended to score higher on SLIM than those who had negative assessments. An ANOVA was performed for the variable interest using the three category external FoR as the factor, students who felt that they were more competent than their peers tended to have higher interest than students who felt they were less competent ($F = 38.60$, $p = .00$). Similarly, based on their internal FoR, students who felt that mathematics was their best subject also tended to report higher levels of interest ($F = 22.29$, $p = .00$). As is shown in Table 6.5, however, not all students with a positive assessment on one frame of reference had a positive assessment on the other. The interaction was examined graphically and is shown in Figure 6.11, which displays mean interest scores for each of the nine groups reported in Table 6.5 as well as 95% confidence intervals for statistically distinct groups. This shows that for most students there was no association between their interest in statistical literacy and their responses to either FoR question. Group means were close to zero for seven of the nine groups. The exception were those students who provided negative assessments on the external FoR, of whom 67% were female. For this group of students the internal FoR also appeared to have an influence on their assessments of interest, in that a change from negative to positive on the internal FoR produced a statistically significant gain in interest. Statistically significant differences in mean interest levels also occurred between students who had a negative assessment on the external FoR and those with neutral or positive assessments.

Students' ability to differentiate between mathematics and statistics

In order to explore students' ability to differentiate between mathematics and statistics, they were asked to compare their interest in statistics relative to other aspects of mathematics (item IE44) and relative to other subjects (item

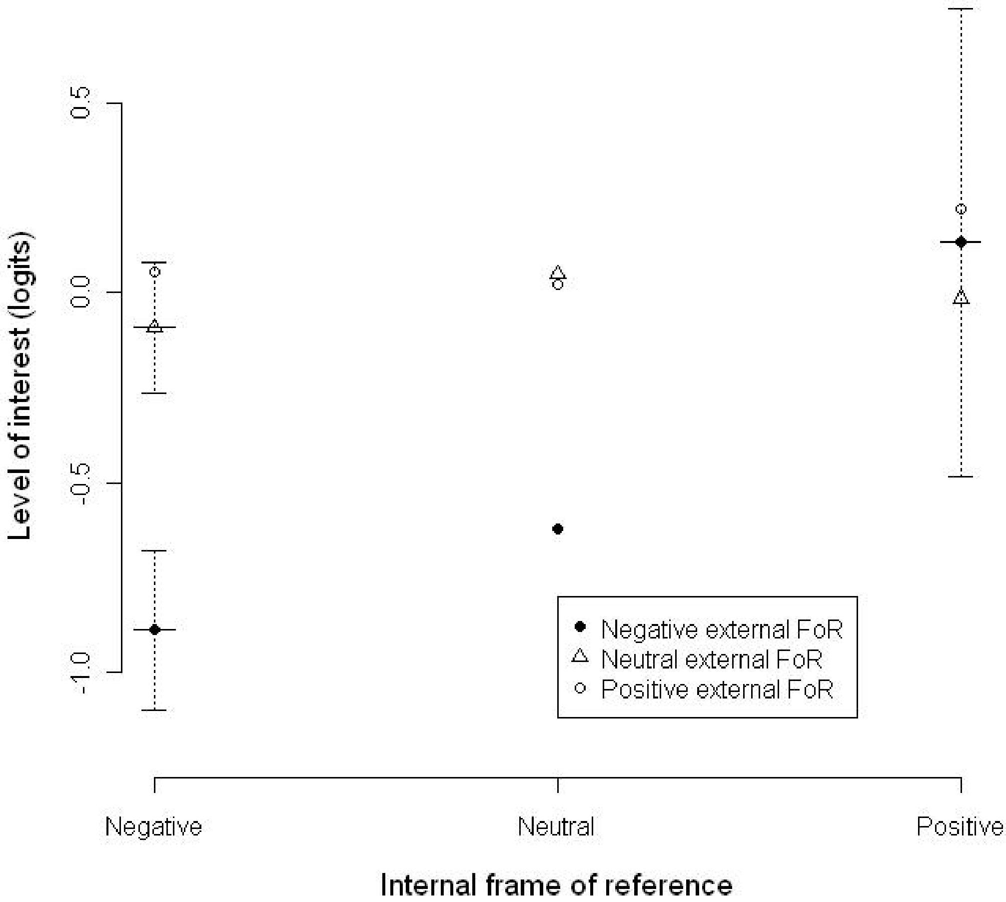


Figure 6.11. Interest by level of internal and external frame of reference as based on students' responses to items IE42 and IE43

IE45). For both items, the existing five categories were collapsed into three categories, as described in the previous subsection. A count of student responses to both items with this new category structure is reported in Table 6.7. As is seen from this table, only 13% of students considered that the statistics encountered in mathematics classes was of more interest than the other work done in mathematics, 63% of students responded negatively to the item and it is assumed considered the other work done in mathematics was of equal or more interest. Similarly, only 17% of students considered the statistics encountered in mathematics classes was of more interest than the statistics

Table 6.7

Interest in statistics compared with maths and other subjects

	Disagree	Neutral	Agree	Total
Item IE44: Statistics is more interesting than other work done in maths	259 (63%)	97 (24%)	54 (13%)	410
Item IE45: Statistics done in maths is more interesting than statistics done in other subjects	232 (57%)	108 (26%)	69 (17%)	409

encountered in other subjects.

Students' responses to items IE44 and IE45 appeared to be influenced by their competency beliefs about mathematics. A cross-tabulation of students' responses to item IE44 against their responses to item IE42 is shown in Table 6.8. Students who responded negatively to item IE44 and considered that statistics was no more or less interesting than the other work done in mathematics were also more likely to respond negatively to item IE42 and see themselves as less competent than their peers at mathematics ($\chi^2 = 17.0$, $p = .00$). Similarly, a cross-tabulation of students' responses to item IE45 against their responses to item IE43 is shown in Table 6.9. Students who considered that the statistics encountered in other subjects was more interesting than that encountered in mathematics were also more likely to respond negatively to item IE43 and regard mathematics as one of their more difficult subjects ($\chi^2 = 70.5$, $p = .00$). Student responses were also influenced by gender, with boys much more likely to respond positively to both items than girls. For example, males made up 64% of students who answered positively to item IE45, yet only made up 39% of all negative respondents.

Table 6.8

Item IE44 against external FoR

		Item IE44			
		Disagree	Neutral	Agree	Total
	Negative	120	24	15	159
External	Neutral	57	29	16	102
reference	Positive	81	44	22	147
(IE42)	Total	258	97	53	408

Table 6.9

Item IE45 against internal FoR

		Item IE45			
		Disagree	Neutral	Agree	Total
	Negative	137	28	15	180
Internal	Neutral	49	32	10	91
reference	Positive	43	47	44	134
(IE43)	Total	229	107	69	405

Summary

The models presented in this section objectively paint a picture that clearly demonstrates the integral part that middle school students' competency beliefs play in the development of their interest. The results demonstrate that these students' self-efficacy beliefs were influenced by their prior mathematics achievement and in turn strongly influenced their interest. This relationship appeared to be quadratic, in that there was an association between the square of self-efficacy and interest. In addition to this, the strength of the relationship appeared to be influenced by the school. In forming competency beliefs, students appeared to use both an external and internal frame of reference, although the internal FoR appeared to have a greater influence on interest when students had negative assessments on the former. Of the other individual factors available in this study, only students' ages appeared to have an influence on their interest, both directly, and indirectly through their self-efficacy. Students' prior achievement in mathematics did not predict their interest in statistical literacy, except through their self-efficacy. This lack of a direct link between achievement and interest may be the result of differences between the two domains, where the prior achievement measured was in mathematics and interest was in statistical literacy. Teacher and/or school factors available in this study did not contribute to students' interest, except for the evidence in the mixed effects models that the school, as a grouping factor, mediated the relationship between self-efficacy and interest.

6.5 Results related to Research Question 3

> To what extent does students' interest in statistical literacy influence their subsequent achievement in statistical literacy?

In order to explore the influence of interest on achievement, a series of models were initially applied to the data in a similar way to that used to answer Research Question 2. Based on the results of these models, a path model was then developed and subsequently tested using AMOS.

The use of linear regression models

A simple linear regression model was applied to the data of 204 students for whom interest, achievement and SLK scores were known. Gender was not a significant predictor of SLK and with the variable Self-efficacy included in the model, Interest ceased to predict SLK. In addition to this, the influence of Interest as a predictor of SLK only became significant when the variable Age was included in the model. The final model, shown as Equation 6.5, explained 43% of the variance in SLK scores. It shows that in the presence of the variables Age, RelMaths-grade and Teacher, Interest is a significant predictor of

SLK. As discussed in Section 4.3, the variable Teacher describes the type of StatSmart test students did, being either pre-test, post-test, or longitudinal test. The variable therefore represents a measure of teacher and/or school influences on the students. Diagnostic plots shown in Figure C.3 of Appendix C, suggest that standard assumptions regarding the normality of residuals and homogeneity of residual variance have been met.

$$\begin{aligned} \text{SLK} \;=\; & -\underset{(0.54)}{4.93} + \underset{(0.04)}{0.29}\text{ Age} + \underset{(0.04)}{0.11}\text{ Interest} + \underset{(0.11)}{0.31}\text{ Median grade} \\ & + \underset{(0.13)}{0.50}\text{ Above median grade} + \underset{(0.13)}{0.78}\text{ Post-test} \\ & + \underset{(0.12)}{0.64}\text{ Longitudinal-test} + \varepsilon_{ij} \end{aligned} \tag{6.5}$$

As is seen from this model, with all other factors constant, students completing the post-test, on average scored 0.78 logits higher than those completing the pre-test. Consequently the factors associated with the teacher and/or school appeared to have a greater influence on students' achievement than individual factors such as interest and age. Indeed Interest appeared to play a relatively minor role in predicting SLK.

The linear model reported above was also tested for teacher, school and state random effects. In this instance, only the inclusion of random teacher effects in the intercept term contributed significantly to model fit. When this term was included in the model, however, Interest ceased to become a significant predictor of SLK, again suggesting that in the presence of teacher factors individual interest plays a minor role in students' achievement. This model, shown as Equation 6.6, reported a standard deviation associated with the grouping factor of 0.47, which is similar in magnitude to the standard deviation associated with the residual error, reported as 0.53. The similarity between the two is in stark contrast to the findings reported in the earlier interest model, shown as Equation 6.2, where the standard deviation associated with the grouping factor, in that case school, was much smaller than that associated with the residual error. This particular result confirms the findings of Hutchison (2009), who reported that school or teacher effects appear to have a much greater influence on students' achievement than on their interest.

$$\begin{aligned}\text{SLK} \;=\;& \underset{(0.89)}{(-4.24 + b_{0i})} + \underset{(0.06)}{0.25}\ \text{Age} + \underset{(0.09)}{0.33}\ \text{Median grade} \\ &+ \underset{(0.10)}{0.55}\ \text{Above median grade} \\ &+ \underset{(0.15)}{0.55}\ \text{Post-test} + \underset{(0.15)}{0.53}\ \text{Long-test} + \varepsilon_{ij} \qquad (6.6)\end{aligned}$$

Structural equation model

Given the results of the regression equations, the path model, shown in Figure 6.12, represents the relationship between Interest, SLK and the other predictor variables. It was tested using AMOS and estimated path coefficients are shown on this figure, whereas the full model is shown in Figure 6.13.

Although the reported model fit statistics are somewhat less than satisfactory ($CFI = 0.88$, $RMSEA = 0.07$), the structural model provides some support for the path model derived from the use of linear models. A multilevel path model may have explained these data better, but software limitations prevented their application. Further research could address this limitation.

Summary

The results presented in the section suggest that interest has a weak, possibly non-significant, influence on student's achievement. Such a result is not uncommon in the literature, with Marsh et al. (2005, p. 411) reporting a "consistent pattern of near-zero, non-significant effects between interest and achievement."

The path model, shown in Figure 6.12, reflects the Expectancy-Value (EV) model of learning (Eccles & Wigfield, 2002), in as much as self-efficacy was a measure of students' expectations of success and interest a measure of

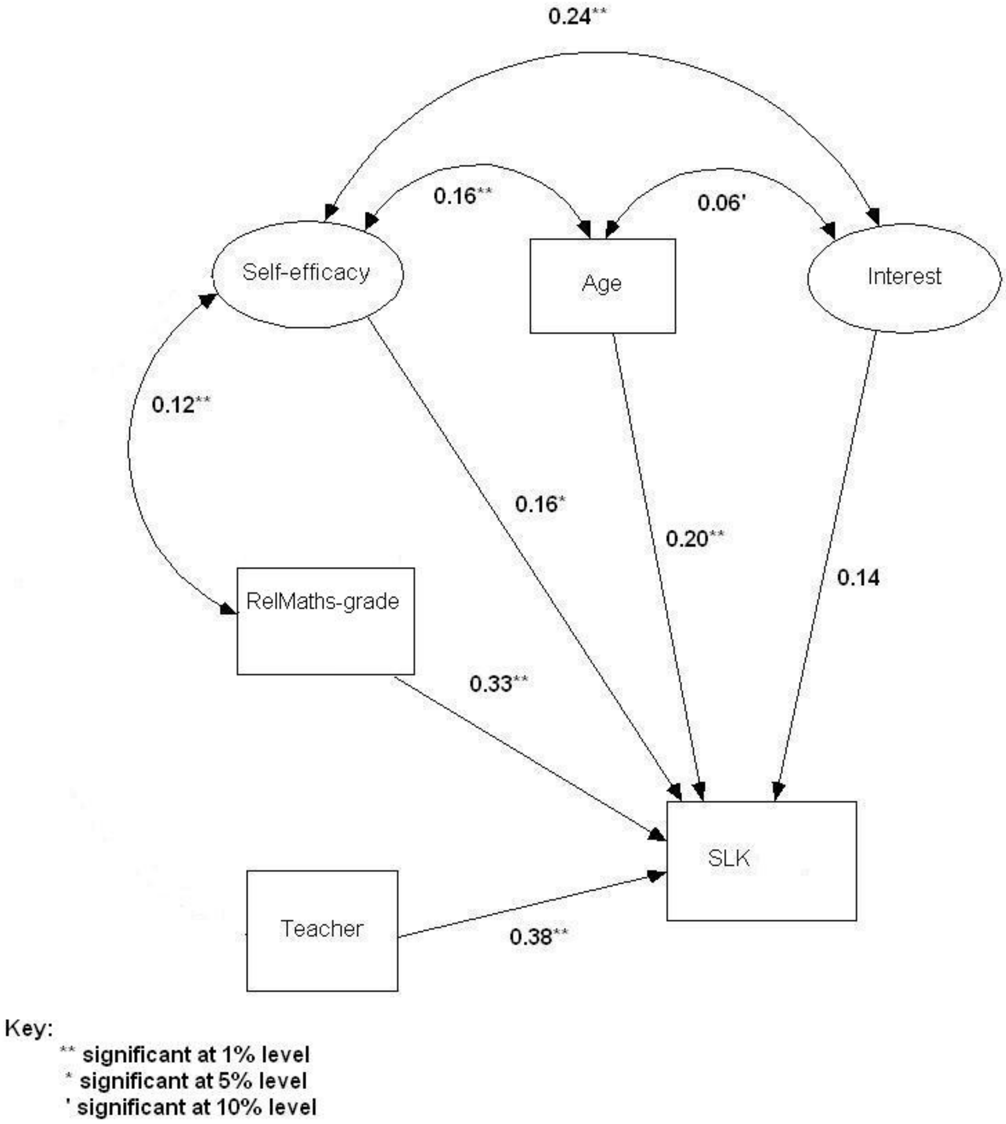


Figure 6.12. Path model summarising findings from linear models

their task-value. In the EV model, expectation of success is considered to be a stronger predictor of achievement than task-value (Wigfield, Tonks, & Eccles, 2004). The results reported earlier, and in particular the absence of an interest/achievement association for boys, suggest that this view may need to consider the influence of gender, in that for this sample of students, task-value appeared to be more influential for girls than for boys in the prediction of their achievement.

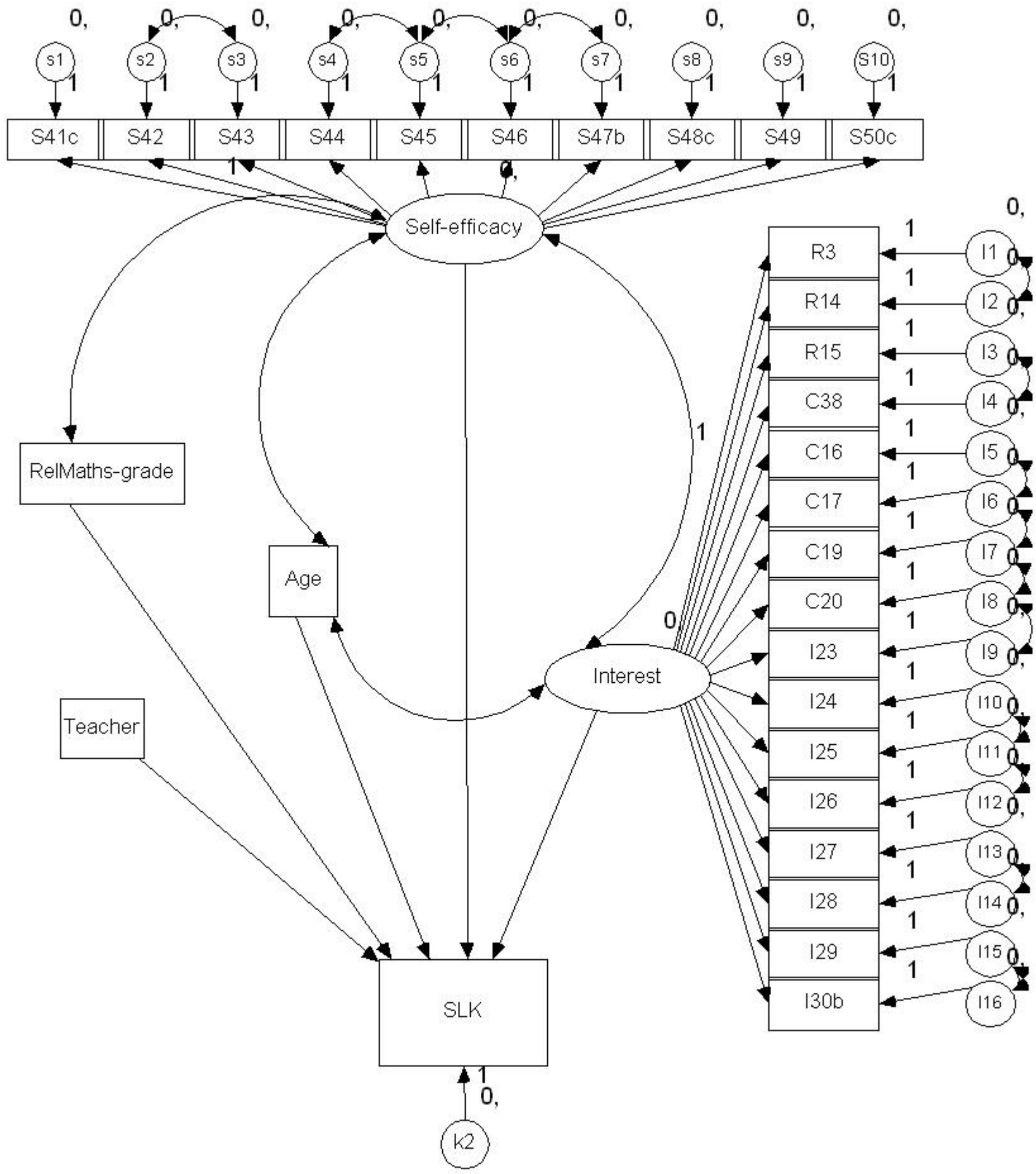


Figure 6.13. Measurement and structural components of achievement model

6.6 Chapter summary

In this chapter the results of the study were presented. In the first section of the chapter, evidence was provided to support the validity of interpretations made from the two developed instruments, SLIM and SESL. The subsequent sections of the chapter then addressed each of the study's research questions. In particular the instruments were used to explore factors that influenced middle school students' interest and achievement in statistical literacy. Students' self-competency beliefs, as measured by their self-efficacy, were the strongest predictor of interest from the variables available. This finding supports the external validity of SLIM in that the literature consistently demonstrates the close link between such beliefs and interest (Marsh et al., 2005; Trautwein et al., 2006). In addition to this, the results suggest that the influence of interest on achievement, at least for this sample of students, is quite minor, with prior achievement, self-efficacy, and teacher related factors, playing a much more substantial role.

The implications of the results presented in this chapter are relevant to researchers and educators in statistical literacy. These, together with future directions for research are discussed in the next chapter.

Chapter 7

Study summary and discussion

The discussion in this chapter commences with a review of the study's results as they relate to each of the research questions. At a more general level, it then proposes an interest hierarchy associated with the development of statistical literacy and addresses reported gender differences in the responses to some of the SLIM items. Following this, the discussion examines the implications of the study's results for both teachers and researchers. It then addresses the limitations of the study and concludes with suggestions for further research.

7.1 Discussion of results

Detailed results of the study are provided in Chapter 6, the following discussion, therefore, provides a summary of these results as they relate to each of the research questions.

The development of a valid measure of interest

The review of the literature, reported in Chapter 3, noted that there was very little research available that specifically addressed the development of interest or indeed any affect in a middle school statistics context. This may be an outcome of the relatively minor emphasis that, until recently, has been placed on statistics education in many mathematics syllabi (Holmes, 2003; Watson, 2006) and it has resulted in a lack of appropriate instrumentation in this context. The literature review also found, however, that there were many studies that addressed affect in an undergraduate statistics context (e.g. Budé et al., 2007; Estrada et al., 2008; Tempelaar, 2006), primarily through attitudinal studies, reflecting the relative emphasis placed on undergraduate statistics. Several of these studies made use of the previously validated "Survey of Attitudes Towards Statistics (SATS)." The instrument, however, was developed specifically for undergraduate and graduate students and as a result was not appropriate for middle school students, especially in light of the finding that younger students tend to be more emotionally unstable than adults (Larson et al., 2002). This study has sought to overcome the lack of appropriate instrumentation through the development of SLIM, a measure of middle school students' interest in statistical literacy. The following discussion addresses the extent to which students' responses to the Likert-type self-descriptions comprising SLIM, reflect a valid measure of their interest in statistical literacy. It commences with a summary of validity issues related to SLIM and then explores specific issues, such as the dimensionality of SLIM, the influence of context on students' responses, and the nomenclature associated with SLIM.

Overall summary. Based on Messick's (1995) six forms of validity evidence, the results presented in the previous chapter suggest that interpretations based on SLIM should be valid. The 16 items comprising SLIM conformed to the requirements of the Rasch measurement model, in that item fit statistics were within acceptable limits. In addition to this, the items sampled the identified topics of statistical literacy (Watson, 2006) and reflected each of the identified interest elements, as shown in Figure 5.1. The measure explained 67% of the variance in students' responses, which is a larger proportion than the 60% regarded by Linacre (2006a) as "good" and its reported person separation reliability was .88.

Issues related to multidimensionality. There was some evidence of multidimensionality in the instrument, with a comparison of deviance test suggesting the existence of dimensions aligning with each of the three interest elements. The evidence of multidimensionality may have been the result of the structure of the questionnaire or instead reflected Hattie's (2009) "rope" model analogy, used to delineate the operational model. Linacre (1998, p. 1) cautioned that "empirical data are always manifestations of more than one latent dimension." The problem, therefore, is not so much whether multidimensionality exists, but whether its existence has an adverse influence on the efficacy of the instrument. Given that the three identified dimensions were highly correlated, the apparent multidimensionality may not have adverse consequences on the use of SLIM. In any case, further investigation of the dimensionality of the instrument is required.

Issues related to the context assessed in self-descriptions. As reported in Chapters 5 and 6, many of the reflective interest items did not conform to the requirements of the Rasch measurement model and were removed from SLIM. It was expected that such items would assess higher levels of interest in statistical literacy than either the curiosity interest or importance interest items. At these high levels of interest, however, students' identities have an important influence on their interest assessments (Renninger, 2009). In particular, the contexts associated with interest self-descriptions and the extent to which students identify with these contexts, appear to play a prominent role in their interest responses. This result has occurred in other interest-based studies, with Häussler (1987) reporting that contexts in science education can explain up to 60% of the variation in students' interest responses. Yet context is important in statistical literacy (Watson, 2006). Consequently a tension exists between the inclusion of contexts and with them the maintenance of content validity, and the need to conform to the requirements of the measurement model. In order to resolve the tension, SLIM has included reflective interest items that are quite general, yet still represent the levels of valuing associated with highly interested students. An interest in "learning more about statistics" (item R14), for example, assessed a desire to re-engage with statistics but in a general context. In addition, SLIM contains specific context laden items that assess lower levels of interest. The importance of understanding "graphs that appear on the internet or in newspapers" (item I28), for example, has a very specific media context but assesses interest at those low levels that are associated with importance (Boekarts & Boscolo, 2002; Ryan & Deci, 2000a).

Issues related to nomenclature. Given that statistical literacy is defined by Gal (2003) as an ability to interpret and critically evaluate messages containing statistical elements, it is acknowledged that SLIM does not assess a student's interest in acquiring such an ability, but rather his or her level of interest in the underlying concepts and learning activities associated with the acquisition of this literacy. Although interest has an emotional component, the interest assessed by SLIM reflects the valuing that is associated with individual interest.

The antecedents of middle school students' interest

The literature review, reported in Chapter 3, suggested that students' interest in statistical literacy should be influenced by a number of factors broadly grouped into those related to the individual (Krapp, 2007) and those related to the situation (Mitchell, 1993). This study has focussed on individual measures and the following discussion examines these. It commences with a brief review of the instrument SESL that was specifically developed during the study to assess students' self-competency beliefs. The discussion then addresses the influence of students' self-competency beliefs, frames of reference, and other individual factors on their interest. It also examines some results related to the influence of the school and teacher. The influence of achievement on interest is addressed in the following section.

A brief review of SESL. The results presented in the previous chapter suggest that the SESL scale provides a valid measure of middle school students' self-efficacy in statistical literacy. The ten items comprising SESL explained 71% of the variance in student responses and reported a person separation reliability of $R_p = .84$. As reported, the items in SESL conformed to the requirements of the measurement model, in that all fit statistics were within acceptable limits.

Self-competency beliefs. In mathematics there is a known association between students' self-competency beliefs, in the form of their mathematics self-concept, and their interest (Marsh et al., 2005; Trautwein et al., 2006). The positive association between students' interest in statistical literacy and their self-efficacy in statistical literacy, reported in Section 6.4, was therefore expected. In addition to this, the reported association between students' interest and the square of their self-efficacy, supports Silvia's (2003) contention that self-efficacy should be related to interest quadratically.

Frames of reference. The frames of reference (FoR) that students used to arrive at their self-competency beliefs were also considered. As proposed by Marsh (1986), two frames of reference were examined in the study: external – comparison of competency with peers – and internal – comparison of competency in the subject with competency in other subjects. The analysis reported in Section 6.4, suggested that apart from students who considered that they are worse at mathematics than their peers, students' interest assessments were relatively independent of their responses to either of the two questions that assessed FoR. The average value of Interest for students with a negative assessment on the external FoR, of whom 67% were female, was significantly lower than the average value of Interest for students with either neutral or positive assessments. Within the first group, positive changes on the internal FoR also had an influence on Interest. The average level of Interest for students with a negative assessment on the external FoR but a positive assessment on the internal FoR, was significantly greater than the average level for students with negative assessments on both FoRs. These results suggest that self-competency perceptions have their greatest influence on interest or rather lack of interest, for those students with relatively negative self-competency perceptions.

A difficulty with the analysis was that students' FoR assessments were with respect to their mathematics performance, while Interest was with respect to statistical literacy. Students with relatively positive mathematics self-competency beliefs may have been able to disentangle their interest in mathematics from their interest in statistical literacy, which in fact may span a number of subject domains. Consequently their mathematics self-competency beliefs had a minimal influence on their interest in statistical literacy. Those students with relatively negative mathematics self-competency beliefs, however, may not have been able to distinguish between the two domains. It is possible that this group of students, dominated by girls in the study, provided low interest assessments for statistical literacy because they did not feel competent in the mathematics classroom. Given the finding by Smith et al. (2007) that women who are anxious about their performance are susceptible to stereotype threats, it is also possible that these students were adversely influenced by stereotypes suggesting mathematics is a male domain.

Other individual factors. It was expected that other individual factors would also contribute to students' interest in statistical literacy. In regard to age, older students tend to report lower levels of interest in learning than younger students (Dotterer et al., 2009). Such a result occurred in this study, where in Section 6.1 a slight negative association was reported between students' ages and Interest ($r = -.10$, $p = .01$). In addition to this, it was expected that gender would influence interest, although it was unclear how this might occur in a middle school statistics context. In the regression models, reported in Section 6.4, gender was not a significant predictor of the variable Interest, although the results of DIF analyses indicated that gender did appear to influence students' responses to some items in SLIM. Given that Frenzel, Goetz, Pekrun, and Watt (2010) recently reported higher levels of boys' interest in the mathematics domain, this lack of influence of gender on interest in statistical literacy, may point to distinct differences between the statistical literacy and mathematics domains. The relationship between gender and interest in statistical literacy is addressed further in the general discussion.

Situational factors. It was expected that situational factors, primarily those related to the teacher, would influence students' interest in statistical literacy. In the study, however, no specific teacher or school factors were measured. Through the use of hierarchical linear models, however, the grouping due to schools was found to influence the nature of the relationship between interest and self-efficacy. Random school effects in both the intercept term and the coefficient of Self-efficacy were found to significantly improve model fit. The lack of school-specific variables in the study, however, prevented further exploration of this finding. In addition, the variable Teacher represented a broad measure of teacher and/or school influence, in that students attending StatSmart schools, who did the post-test, had been taught for 6 months or more by a teacher undertaking professional development in statistics. The variable Teacher, although significantly influencing students' achievement, did not predict their interest. This suggests that any influence the teacher has on students' interest is either relatively minor or more long-term, in that students' interest is relatively stable and possibly influenced by a composite of all their previous teachers' efforts. In regard to the relative strength of the teacher's influence on their students' interest, Frenzel et al. (2010) reported that although students' perceptions of their teachers' enthusiasm for teaching influenced students' interest in mathematics ($b = 0.06$), students' perceptions of their peers' valuing of mathematics had a much greater influence ($b = 0.33$).

The influence of the mathematics classroom. It is not at all clear whether middle school students themselves can disassociate their interest in statistical literacy from their beliefs and attitudes towards mathematics in general. A large number of students in the study expressed the opinion that the statistics encountered in mathematics classes was no more interesting than the other work encountered in these classes. Similarly a large number of students found that the statistics encountered in mathematics classes was no more interesting than those encountered in other classes. The results related to students' frames of reference indicated that students with negative self-competency beliefs in mathematics were more likely to report lower interest in statistical literacy than students with positive self-competency beliefs. As discussed, such negative beliefs may arouse negative emotions that fail to distinguish between the mathematics and statistical literacy domains.

The relationship between interest and achievement

The Model of Domain Learning (Alexander, 2003) predicts that interest will grow as knowledge in a domain increases, suggesting a positive association between achievement and interest. Indeed a meta-analysis of 31 studies reported an average correlation of .32 between measures of interest and achievement in mathematics (Schiefele, 1992). More recently, however, Marsh et al. (2005) reported weak and non-significant associations between interest and achievement in mathematics. Further, Trautwein et al. (2006, p. 803) reported that in mathematics, self-competency beliefs were a "potent predictor of interest and almost completely mediated the effects of achievement." In as much as the variable shown in Table 4.5 as RelMaths-grade was a measure of students' prior achievement in statistical literacy, the results reported in Section 6.5 confirm these more recent findings of Marsh et al. (2005) and Trautwein et al. (2006). Students' prior achievement did influence their interest in statistical literacy but only through their self-efficacy beliefs. In addition to this there was a weak non-significant association between students' interest and their SLK score ($r = .11$, $p = .06$). Given the domain similarities between the variables Interest and SLK, it was surprising that this association was not stronger. It is possible that unlike their mathematics achievement, reflected in the variable RelMaths-grade, students did not view the StatSmart tests as of importance. Trautwein et al. (2006) reported higher associations between students' interest and mathematics grades, which they argued represented high-stakes assessment, than between their interest and performance in a standardised mathematics test used in their study, which they argued represented low-stakes assessment.

The results of the study suggest that gender might have some influence on the relationship between interest and achievement, in that there was a weak but significant association between girls interest in statistical literacy and their statistical literacy knowledge, yet no such association for boys. The finding is surprising, given that Schiefele et al. (1992) reported the association between interest and achievement was stronger for males than for females. In the regression models, however, there was no evidence in this study to suggest that gender influenced the relationship between the variables Interest and SLK. It appears that greater statistical power is required to explore the influence of gender on this relationship.

7.2 General discussion

Although a statistical literacy hierarchy has been identified (Watson & Callingham, 2003), it was noted in Chapter 2 that this hierarchy does not include the affective development of students. The first part of this general discussion, therefore, proposes an associated interest hierarchy that could be used to map students' affective development in statistical literacy. The second part of the discussion then addresses issues related to the reported gender differences in responses to some SLIM items.

The statistical literacy interest hierarchy

The Rasch analysis of students' responses to SLIM, has allowed for the placement of interest self-descriptions and person interest measures on the one hierarchical scale, thus enabling their meaningful comparison. As shown in Figure 6.1 of Chapter 6, the clustering of item thresholds suggests that the statistical literacy interest hierarchy can be divided into five broad bands. These bands are the result of one large break in the hierarchical order associated with the item difficulties and the five-point category structure used in the instrument. The following discussion examines smaller breaks in the hierarchy of item difficulties and proposes that it can be logically partitioned into four divisions. In light of these proposed divisions, the discussion then revisits the five-band hierarchy shown in Figure 6.1.

The four-division hierarchy of items. The identification of these divisions and the items within them, commenced with a scan of the hierarchy of item difficulties, as shown in Table 6.2, for clusters of items and obvious discontinuities, such as separations exceeding two standard errors. The final divisions were then based on clusters of items grouped logically according to substantive theory. This process resulted in the four divisions shown in Table 7.1, which also reports the context and content emphasized in each division.

Table 7.1

The four-division hierarchy of interest items

Division	Description	Context and content	Items
4	Interest in statistical activities and a desire to re-engage	No contexts	R15, C38, R14, R3
3	The importance of and a desire to find out about statistical literacy	Wider contexts. Inference and interpretation of data.	*C19*, C17, C16, C20, I23, I25
2	Importance of statistical literacy	Self-related contexts. Data interpretation and chance.	I24, *I26*, I30b
1.	Importance of mastering simple tasks related to statistical literacy.	Self-related contexts. Data presentation.	I28, I27, I29

The three items in Division 1 had difficulties of a similar magnitude and all assessed aspects related to task mastery. As shown in Table 6.2, their difficulties ranged from -0.54 to -0.46 logits, a relatively short interval given the reported standard errors in item difficulties were 0.04 logits. The items in the division assessed the importance of mastering simple tasks related to statistical literacy. As an example, “arranging data into tables” (item I29) is a routine statistical task. Its endorsement by many students suggests it was viewed as very relevant. Yet such endorsement is likely to reflect an extrinsic motivation – getting good marks in school – and consequently low levels of interest in statistical literacy (Boekarts & Boscolo, 2002; Ryan & Deci, 2000a).

The three items in Division 2 also had difficulties of a similar magnitude. Their difficulties ranged from -0.35 to -0.25 logits, a relatively short interval in terms of the standard error. The item difficulty gap separating Divisions 1 and 2 was 0.11 logits. Although quite small, this gap does exceed two standard errors in item difficulty. The items in Division 2 assessed the importance of statistical literacy but primarily in self-related contexts. Knowing “how to calculate the chance of being injured from risky behavior” (item I24) assessed statistical literacy in a context very much associated with the self. Endorsement of such an item is likely to reflect immediate goals related to the self – being safe – but goals that are more distant for students than those associated with getting good marks.

With the exception of item C19, the difficulties of the six items in Division 3 were clustered about zero, ranging from -0.07 to 0.09 logits. The item difficulty gap separating Divisions 2 and 3 was 0.18 logits, a large interval in terms of the standard error. The items in Division 3 appeared to assess both curiosity and importance interest elements. They also tended to assess statistical literacy in wider contexts than those associated with the self. Such contexts are arguably less personally relevant to students and given the reported association between personal relevance and interest (Hulleman & Harackiewicz, 2009), their endorsement reflects higher levels of interest. Understanding “news reports that use averages” (item I23) assessed statistical literacy in a media context. Fewer students could endorse this item and thus see it as personally relevant. Endorsement of such an item is likely to reflect distant goals – such as being an effective citizen – and demonstrates increasing levels of interest in statistical literacy.

The difficulties of the four items in Division 4 ranged from 0.39 logits through to 0.76 logits. The item difficulty gap separating Divisions 3 and 4 was 0.30 logits, a very large gap in terms of the standard error. All of these items assessed reflective interest, with the more difficult items assessing a desire to re-engage with statistical literacy. Items in this latter group, such as a wanting to find out “all there is to know about statistics” (item C38) and “getting a job that involves statistics” (item R15) represent high levels of interest in statistical literacy (Hidi & Renninger, 2006).

Not all items conformed to the logical grouping described above and these are italicized in the table. Item C19, for example, assessed statistical literacy in a political context. The reported difficulty of the item was 0.43 which placed it in Division 4. The context associated with the item, however, did not appear to be personally relevant to these students. Similarly, being able to “believe scientific claims that are based on data” (item I26) was placed relatively low in the hierarchy, yet appeared to assess a context wider than one associated with the self, perhaps reflecting the goal to be an effective citizen.

Overview of the statistical literacy interest hierarchy. The five stages marked on Figure 6.1 broadly align with the acclimation and competence stages of the Model of Domain Learning (Alexander, 2003), providing a five-stage hierarchy. The exact alignment, however, is a suggestion for further research and the terms used below to describe each stage are tentative . As is seen from the figure, which is shown again as Figure 7.1, the placement of item thresholds reflects both the four division hierarchy of items and the five category structure used with the instrument.

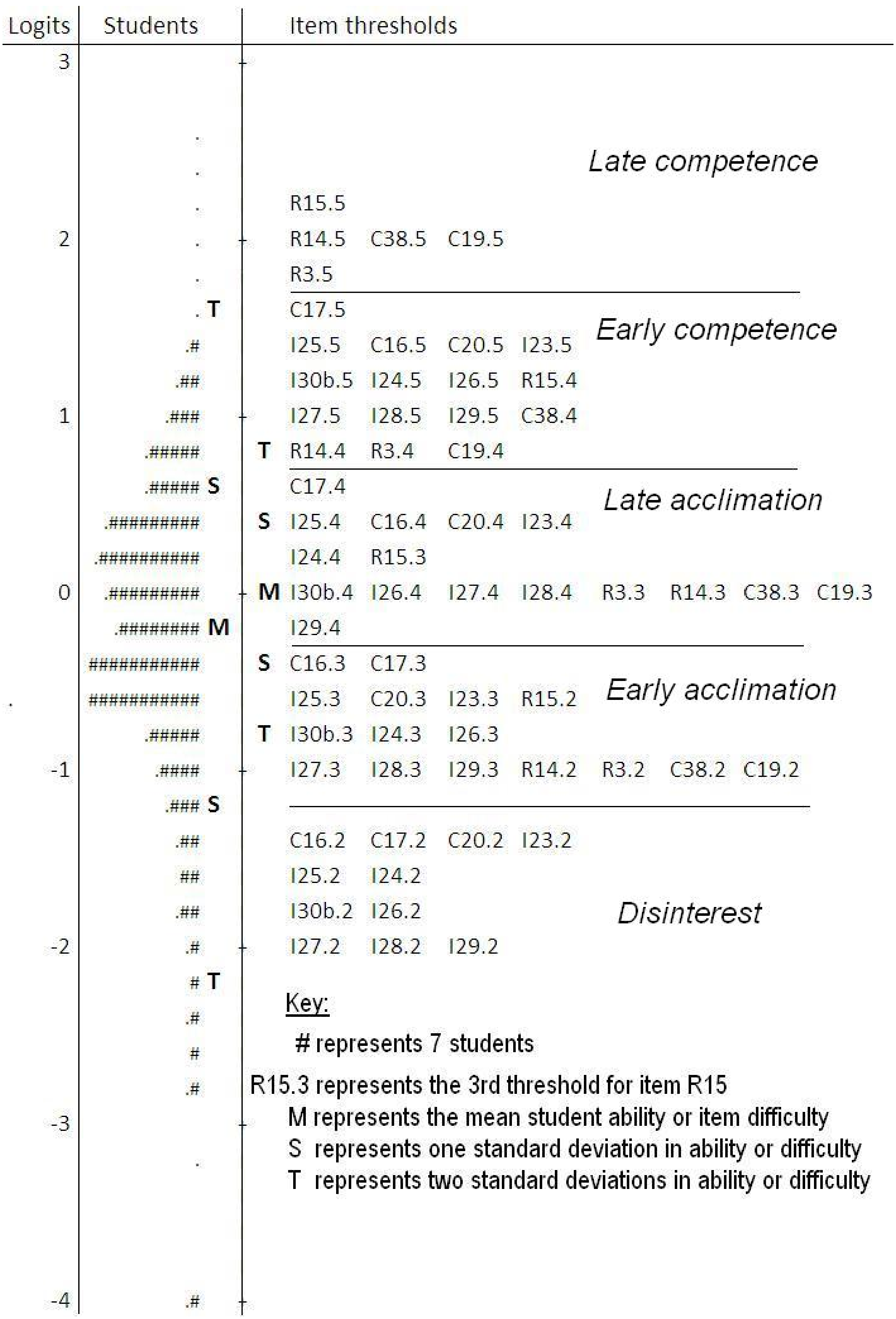


Figure 7.1. Proposed statistical literacy interest hierarchy

The lowest stage on the hierarchy, termed “Disinterest”, represents very low, if any, interest in statistical literacy. As is seen from the figure, many students in this stage of interest development were likely to respond with a 1, the lowest category, to all SLIM self-descriptions. Even near the upper reaches of this stage, at levels of interest above -2 logits, students barely acknowledged the importance of task-mastery, providing a response of 2 to Division 1 items. Near the boundary of this stage, at an interest level of approximately -1.2 logits, students were likely to respond with a 2 to Division 2 and 3 items.

The second lowest stage, termed “Early acclimation”, represents low interest for statistical literacy. Students in this stage of interest development were likely to acknowledge some importance in mastering tasks associated with statistical literacy, responding with a 3 to Division 1 items. Students near the top of this stage were also likely to see some importance in gaining statistical literacy, responding with a 3 to Division 2 and and some Division 3 items.

The third lowest stage, termed “Late acclimation”, represents moderate interest for statistical literacy. Students in this stage of interest development were likely to positively endorse the importance of mastering tasks associated with statistical literacy, responding with a 4 to Division 1 items. Students near the top of this stage were also likely to endorse the importance of statistical literacy in wider contexts, responding with a 3 or 4 to Division 2 and 3 items.

The second highest stage, termed “Early competence”, represents high interest for statistical literacy. Students in this stage of interest development were likely to completely endorse the importance of mastering tasks associated with statistical literacy, responding with a 5 to Division 1 items. Students near the top of this stage were also likely to positively endorse the importance of statistical literacy in wider contexts, responding with a 4 or 5 to Division 2 and 3 items. Students in this stage also started to show appreciable levels of interest in re-engaging with statistical literacy, responding with a 3 or 4 to most Division 1 items.

The highest stage on the hierarchy, termed ”Late competence”, represents very high levels of interest. Students in this stage completely endorsed the importance of statistical literacy and had a desire to re-engage in the domain. They were likely to respond with a 5 to all Division 1, 2, and 3 items and most Division 4 items.

Gender considerations

This study found that although overall levels of interest were similar for boys and girls, there were items that attracted more interest from boys and others more interest from girls. Boys were more likely than girls to find an interest in “working on problems involving data and statistics” (item R3). Such an interest might reflect findings in the sciences that boys show a general interest in technical objects (Jenkins & Pell, 2006) or it could reflect gender stereotypes associated with mathematics. Girls, on the other hand, were more likely to want to find out “how a survey will be used to predict the next election” (item C17) and “whether a survey reported on the radio or TV about students was correct” (item C20). These items were the only two that specifically used the term “survey” and it is possible that these results reflect the reported general interest that girls have for social applications (Häussler, Hoffman, Langeheine, Rost, & Sievers, 1998) and their predicted need to find a sense of self through a connection with others (Powell, 2004).

These reported gender differences might also reflect known gender stereotypes for mathematics and language, in that mathematics is a stereotypically male domain and language a female domain (Smith et al., 2007). Apart from knowledge of statistical concepts, statistical literacy also requires language and mathematical skills (Gal, 2002), which may invoke different levels of interest according to whether gender stereotypes are operating.

Students’ achievement-related goals could explain the way that gender stereotypes influence their interest. Hyde and Durik (2005) reported that students are more likely to show higher levels of achievement-related goals in domains where their gender is stereotypically favoured. In particular, they reported that boys show higher levels of both performance and mastery achievement goals in mathematics and girls show higher levels of performance and mastery goals in reading and language. In addition to the reported influence of gender stereotypes, there is a reported positive association between students' mastery goal orientation and their interest (Harackiewicz et al., 2008; Pekrun et al., 2009). As as a result, the reported tendency in this study for boys to find more interest in doing problems might be influenced by gender stereotypes and reflect their perception that such tasks are inherently mathematical. Similarly, the reported tendency for girls in this study to find more interest in surveys might reflect gender stereotypes and their perception that tasks associated with surveys require inherent language and reading skills.

7.3 Study implications

The development of SLIM and the proposed interest hierarchy have implications for researchers, in that the evaluation of teaching interventions in statistical literacy could include affective data. The discussion in this section, however, focusses on the study's implications for teachers and curriculum designers. It examines the importance of self-efficacy on the positive development of interest, the need for students to see the personal relevance of contexts in their learning of statistical literacy, and the unique role that statistical literacy may have in minimising the harmful affects of gender stereotypes.

Addressing students self-competency beliefs

The close association between students' self-efficacy and their interest, reported in this study, is a reminder of the inter-relatedness of cognitive and affective development. Teachers who wish to raise the level of interest that their students have for statistical literacy can do so through addressing their students' self-competency beliefs. Although these beliefs are based on students' task-mastery, they are also influenced by the support and encouragement of significant others (Bandura, 1997). Support of students in their concept development, both through encouragement and the impartation of skills and strategies, is therefore likely to impact positively upon their interest.

The personal relevance of contexts

Contexts play an important role in the acquisition of statistical literacy. Students with high levels of statistical literacy are able to interact critically with a range of contexts (Watson & Callingham, 2003). Contexts also appear to play a key role in students' interest in statistical literacy. The interest hierarchy, presented earlier, predicts that students who can see the personal relevance of statistical literacy in wider contexts are likely to have higher levels of interest. In a science education context, Hulleman and Harackiewicz (2009) found that asking students to write about the personal relevance of what they were learning helped to increase their levels of interest, especially for those students with low self-competency beliefs. In teaching statistical literacy, teachers need a wide range of contexts at their disposal. Some of these may appeal to their students and some may not. Nevertheless encouraging students to reflect on the personal relevance of the context may go some way towards increasing their level of interest in statistical literacy.

Addressing gender stereotypes

Hyde (2005) argued that gender stereotypes are harmful, in that girls might be less inclined to pursue mathematical careers and boys less inclined to pursue careers involving language and reading. Statistical literacy is unique in that it is based, in part, on a number of mathematical and language skills (Gal, 2002; Watson, 2006). The placement of this literacy in the curriculum and its subsequent teaching have the potential to address the more harmful aspects of gender stereotyping. The draft Australian Curriculum (National Curriculum Board, 2009) situates statistics firmly within the mathematics syllabus. "Statistics and Probability" is now one of only three proposed content strands in the mathematics syllabus and there is a strong emphasis on the cross-curricular nature of numeracy, and thus statistical literacy. Given the earlier suggestion that gender stereotypes may influence students' interest, the teaching of statistical literacy in the mathematics classroom should not emphasise mathematical skills at the expense of literacy skills. As an example, learning activities associated with the acquisition of statistical literacy can be embedded in media contexts (Watson, 2006). In this way boys should appreciate the relevance of language skills to statistical literacy. Similarly, the teaching of statistical literacy in non-mathematical domains such as the social sciences should not emphasise language skills at the expense of mathematical skills. The social science teacher must be able to integrate associated mathematical concepts so that girls can appreciate their relevance in non-mathematical domains.

7.4 Limitations of the study

The generalisability of interpretations based on the study partly rest upon the use of randomness in sample selection. Random selection of students did not occur and indeed ethical considerations make it very difficult to achieve randomness in studies involving children. The use of a large representative sample, as was the case in the study, addressed this limitation. Due to the lack of randomness, however, all cited p-values in the study are notional, as are claims of statistical significance.

During the modelling process, it was necessary to use the responses of students for whom all variables were known. The models of achievement, for example, were based on a sub-sample of 295 students. The use of sub-samples in such cases was unavoidable, but it did reduce the available statistical power and compromise the representativeness of the particular sub-sample.

In applying statistical models to the data it is acknowledged that "all models are wrong, but some are useful" (Box & Luceno, 1999). All of the models used in the study are based on assumptions, some of which were not fully met. Where-ever possible, however, violations in model assumptions were addressed, in many cases through the use of more complex models.

In order to avoid respondent burden, the number of items in the questionnaire was deliberately minimized. This resulted in a number of single-item measures that were used to answer research question 2. Whereas such measures lack the substance of the multi-item measures developed in the study, they have been used as means of exploring relationships between interest and other key adolescent developmental factors. Further work in this area using multiple-item measures instead, is required.

7.5 Recommendations and future research

The evidence presented in the study suggests that interpretations made from the Statistical Literacy Interest Measure are valid. The results also suggest the need for further item development in the instrument, particularly those assessing lower levels of interest. The statistical literacy interest hierarchy currently commences with importance interest items at its lowest levels. Arguably these reflect an integrated-extrinsic motivation, which resides near the top of Ryan and Deci's (2000a, p. 61) "Taxonomy of human motivation." Further item development for SLIM could consider self-descriptions that assess less integrated forms of extrinsic motivation, such as introjected regulation, where students perform tasks to avoid guilt or satisfy parental expectations. These self-descriptions could possibly use the common stems “I need to know” or “I should know how to”.

The issue of context plays a key role in differentiating the levels associated with the statistical literacy hierarchy, as described by Callingham and Watson (2005). Students in the lower levels of the statistical literacy hierarchy typically have an informal engagement with context. As they progress through the hierarchy their engagement becomes more formal, consistent and finally critical. The interest hierarchy presented earlier in the chapter also points to a key role for context, in that an ability to see the personal relevance of wider contexts appears to be associated with higher levels of interest. In addition, the topics associated with statistical literacy also appear to influence students’ interest, in that items assessing data presentation were in general much lower on the hierarchy than those assessing beginning inference. Consequently the interaction of content and context on students’ interest needs to be explored further. Given that the statistical literacy hierarchy reported in Watson and Callingham (2003) was identified on the basis of 3852 student responses and 80 test items, such an exploration may require a larger study, with more self-descriptions, than the study reported here. A larger study would allow a more accurate specification of the relationship between interest and achievement in statistical literacy.

Students’ goal orientations appear to play a key role in the development of their interest, in that students with high levels of mastery goals are more likely to report higher levels of interest than students with low levels of such goals (Harackiewicz, et al. 2008). In addition, there appear to be gender differences in the way achievement goals influence motivation, with Hyde and Durik (2005) reporting that the motivational benefits of adopting performance goals were stronger for boys than for girls. The finding in the study of no association between interest and achievement for boys, yet a weak positive association for girls is surprising. It could be a feature of the particular sample, or it could suggest the possibility that gender might also influence the motivational benefits of interest in statistical literacy, possibly through the goal-orientation of students. Although recent research, described in Wigfield and Cambria (2010), has explored the relationship between achievement goal-orientation and interest, it has not explicitly explored gender differences. In light of the results of the study, such research would be beneficial to educators.

The Model of Domain Learning (MDL) predicts that students' acquisition of statistical literacy will depend on their interest in statistical literacy and their ability to acquire and use appropriate strategic skills. As noted, though, empirical studies involving the MDL have mostly been based on adult learners in a tertiary context. There is a need to establish the viability of this model in a middle school context, where interest typically shows a declining trend (Dotterer, et al., 2009). Given that Watson and Callingham (2003) have mapped out the statistical literacy hierarchy, and that this study has laid the foundations for a valid measure of interest in statistical literacy and with it a proposed interest hierarchy, further research is required to investigate the third component of the MDL, namely the strategic skills employed by middle school students as they progress through the statistical literacy hierarchy. The development of an instrument to assess strategic skill usage could allow the MDL to be tested in the middle school.

7.6 Concluding comments

This study has explored interest as a source of motivation for children. As noted, several studies have documented the decline in levels of students' interest across the entire school period and this trend was also observed in the study, where a negative association between interest and age was reported. Yet interest is an important predictor of re-engagement and as a result, teachers need to ensure that their students' interest develops positively. Given the importance of statistical literacy as a key life-skill, the study has developed an instrument that could enable teachers and others to monitor their students' interest in statistical literacy. Based on the results of the study, the Statistical Literacy Interest Measure (SLIM) should provide teachers with valid information about their students' interest in statistical literacy.

At a more general level, it was noted that very little research has explored the influence of affect in a secondary school statistics context. Given the increased emphasis that statistics education now appears to have in the proposed Australian curriculum, both SLIM and SESL are timely additions to the repertoires of researchers seeking to explore further the development of middle school students' statistical literacy.

References

Adams, R.J., & Wu, M.J. (1997). Multi-level item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, *22*(1), 47–76.

Aiken, K.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interaction*. Newbury Park: SAGE Publications.

Alexander, P.A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher*, *32*(8), 10–14.

Alexander, P.A., Jetton, T.L., & Kulikowich, J.M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology*, *87*(4), 559–575.

Alexander, P.A., Sperl, C.T., Buehl, M.M., Fives, H., & Chiu, S. (2004). Modeling domain learning profiles from the field of special education. *Journal of Educational Psychology*, *96*(3), 545–557.

Allredge, J.R., Johnson, H.D., & Sanchez, J.J. (2006). Does viewing videos of statistics in action affect student attitudes? In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Available from: www.stat.auckland.ac.nz/~iase.

Anderson, E.B. (1997). The Rating Scale Model. In W.J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 67-84). New York: Springer.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573.

Arbuckle, J. (2008). Amos (17.0.0). [Computer Software]. Crawfordville FL: Amos Development Corporation.

Australian Bureau of Statistics. (1997). *Australian social trends no. 4102.0*. Available from: http://www.abs.gov.au/.

Australian Bureau of Statistics. (2008). *Schools Australia no. 4221.0*. Available from: http://www.abs.gov.au/.

Bakker, A., Derry, J., & Konold, C. (2006). Using technology to support diagrammatic reasoning about center and variation. In A. Rossman & B. Chance (Eds.), *Procedings of the Seventh International Conference on Teaching Statistics.* Available from: www.stat.auckland.ac.nz/~iase.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Ben-Zvi, D., & Garfield, J. (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. 3-15). Dordrecht: Kluwer Academic Publishers.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Berlyne, D. (1960). *Conflict, arousal, and curiosity.* New York: McGraw-Hill.

Bidgood, P. (2006). Creating statistical resources from real datasets: The STARS project. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Available from: www.stat.auckland.ac.nz/~iase.

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

Bikner-Ahsbahs, A. (2004). Interest-dense situations and their mathematical valences. In *Proceedings of the Tenth International Congress on Mathematical Education*. Retrieved from: http://www.icme-organisers.dk/tsg24/Documents/BiknerAshbahs.doc.

Boekarts, M., & Boscolo, P. (2002). Interest in learning, learning to be interested. *Learning and Instruction*, *12*(4), 375–382.

Bollen, K., & Barb, K. (1981). Pearson's r and coarsely categorized measures. *American Sociological Review*, *46*(2), 232–239.

Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bong, M., & Skaalvik, E.M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40.

Box, G., & Luceno, A. (1999). *Quality quandaries: Six sigma, process drift, capability indices, and feedback adjustment*. Retrieved from: http://cqpi.engr.wisc.edu/system/files/r176.pdf.

Budé, L., Van De Wiel, M.W.J., Imbos, T., Candel, M.J.J.M., Broers, N.J., & Berger, M.P.F. (2007). Students' achievements in a statistics course in relation to motivational aspects and study behaviour. *Statistics Education Research Journal*, *6*(1), 5–21.

Byrne, B. (2001). *Structural equation modeling with AMOS*. Mahwah, NJ: Lawrence Erlbaum Associates.

Callingham, R., & Watson, J.M. (2005). Measuring statistical literacy. *Journal of Applied Measurement*, *6*(1), 1–29.

Callingham, R., & Watson, J. (2007). Overcoming research design issues using Rasch measurement: The StatSmart project. In P. Jeffery (Ed.), *Proceedings* *of the Australian Association for Research in Education Annual Conference.* Available from: http://www.aare.edu.au/07pap/cal07042.pdf.

Carmichael, C.S., & St. Hill, R. (2006). Towards quality in multiple-choice assessment. *International Journal of Business and Management Education*, *13*(8), 33–47.

Carmona, J. (2004). Mathematical background and attitudes towards statistics in a sample of undergraduate students. In *Proceedings of the Tenth International Conference on Mathematics Education.* Available from: www.stat.auckland.ac.nz/~iase.

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3). Available from: http://www.amstat.org/publications/jse/.

Cobb, G.W., & Moore, D.S. (1997). Mathematics, statistics and teaching. *American Mathematical Monthly*, *104*(9), 804–823.

Cobb, P., & Hodge, L. (2002). Learning, identity, and statistical data analysis. In B. Phillip (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Available from: www.stat.auckland.ac.nz/~iase.

Cook, R. (1977). Detection of influential observations in linear regression. *Technometrics*, *19*(1), 15–18.

Coulson, J. (Ed.) (1969). *The little Oxford dictionary.* Oxford: Clarendon Press.

Csikszentmihalyi, M. (2002). *Flow: The classic work on how to achieve happiness.* London: Rider.

Curtis, D., & Boman, P. (2007). X-ray your data with Rasch. *International Education Journal*, *8*(2), 249–259.

D'Andrea, L., & Waters, C. (2002). Teaching statistics using short stories: Reducing anxiety and changing attitudes. In B. Phillip (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Available from: www.stat.auckland.ac.nz/~iase.

Deci, E.L. (1992). The relation of interest to the motivation of behaviour: A self-determination theory perspective. In K.A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 43-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination.* New York: Plenum Press.

Delbridge, A.J., Bernard, J.R., Blair, D., & Ramson, W.S. (Eds.) (1987). *The Macquarie dictionary.* Sydney: The Macquarie Library.

Department of Education, Science and Technology (2005). *Schools assistance regulations.* Retrieved from Department of Education, Science and Training website: http://www.dest.gov.au.

Dewey, J. (1910). *Educational essays.* Bath: Cedric Chivers Ltd.

Dotterer, A., McHale, S., & Crouter, A. (2009). The development and correlates of academic interests from childhood through adolescence. *Journal of Educational Psychology*, *101*(2), 509–519.

Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.

Egan, Q., & Archer, P. (1985). The accuracy of teachers' ratings of ability: A regression model. *American Educational Research Journal*, *22*(1), 25–34.

Estrada, A., Batanero, C., Fortuny, J.M., & Diaz, C. (2008). A structural study of future teachers' attitudes towards statistics. In *Proceedings of joint ICME*

Study and IASE Round Table Conference. Retrieved from: http://www.ugr.es/~icmi/iase_study/Sample%20Paper2.pdf.

Faraway, J. (2006). *Extending the linear model with R: Generalized linear, mixed effects and non-parametric models*. New York: Chapman and Hall.

Finney, S.J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*(2), 161–186.

Finzer, W. (2006). What does dragging this do? The role of dynamically changing data and parameters in building a foundation for statistical understanding. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Available from: www.stat.auckland.ac.nz/~iase.

Fox, L.H. (1982). *The study of social processes that inhibit or enhance the development of competence and interest in mathematics among highly able young women*. Retrieved from ERIC database. (ED222037).

Fredricks, J.A., & Eccles, J.S. (2002). Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, *38*(4), 516–533.

Frenzel, A.C., Goetz, T., Pekrun, R., & Watt, H.M.G. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507–537.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–51.

Gal, I. (2003). Teaching for statistical literacy and services of statistical agencies. *The American Statistician*, *57*(2), 80–84.

Garfield, J.B. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22–38.

Gehlbach, H., Brown, S., Ioannou, A., Boyer, M., Hudson, N., & Niv-Solomon, A. (2008). Increasing interest in social studies: Social perspective taking and self-efficacy in stimulating simulations. *Contemporary Educational Psychology*, *33*(4), 894–914.

Goetz, T., Frenzel, A., Hall, N., & Pekrun, R. (2008). Antecedents of academic emotions: Testing the internal/external frame of reference model for academic enjoyment. *Contemporary Educational Psychology*, *33*(1), 9–33.

Goldin, G. (2002). Affect, meta-affect, and mathematical belief structures. In G.C. Leder, E. Pehkonen, & G. Torner (Eds.), *Beliefs: A hidden variable in mathematics education* (pp. 59-72). Dordrecht: Kluwer Academic Publishers.

Haertel, E.H. (1997). Reliability. In R. Brennan (Ed.), *Educational measurement* (pp. 65-110). Westport, CT: Praeger Publishers.

Harackiewicz, J., Durik, A., Barron, K., Linnenbrink-Garcia, L., & Tauer, J. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest, and performance. *Journal of Educational Psychology*, *100*(1), 105–122.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Abingdon, UK: Routledge.

Häussler, P. (1987). Measuring students' interest in physics-design and results of a cross-sectional study in the Federal Republic of Germany. *International Journal of Science Education*, *9*(1), 79–92.

Häussler, P., Hoffman, L., Langeheine, R., Rost, J., & Sievers, K. (1998). A typology of students' interest in physics and the distribution of gender and age within each type. *International Journal of Science Education*, *20*(2), 223–238.

Hay, I., & Ashman, A. (2003). The development of adolescents' emotional stability and general self-concept: The interplay of parents, peers, and gender. *International Journal of Disability, Development and Education*, *50*(1), 77–91.

Hidi, S., & Harackiewicz, J.M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, *70*(2), 151–179.

Hidi, S., & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127.

Hoffman, L. (2002). Promoting girls' interest and achievement in physics classes for beginners. *Learning and Instruction*, *12*(4), 447–465.

Hoge, R., & Coladarci, T. (1989). Teacher-based judgements of academic ability: A review of the literature. *Review of Educational Research*, *59*(3), 297–313.

Holland, J. (1985). *Making vocational choices: A theory of vocational personalities and work environments*. Englewood Cliffs, NJ: Prentice-Hall.

Holmes, P. (1986). A statistics course for all students aged 11-16. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics*, (pp. 194–196). Available from: www.stat.auckland.ac.nz/~iase.

Holmes, P. (2003). 50 years of statistics teaching in English schools: Some milestones. *The Statistician*, *52*(4), 439–474.

Hu, L., & Bentler, P. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Hulleman, C., Durik, A., Schweigert, S., & Harackiewicz, J. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Pscyhology*, *100*(2), 398–416.

Hulleman, C., & Harackiewicz, J. (2009). Promoting interest and performance in high school science classes. *Science*, *236*(5958), 1410–1412.

Hunter, J.P., & Csikszentmihalyi, M. (2003). The positive psychology of interested adolescents. *Journal of Youth and Adolescence*, *32*(1), 27–35.

Hutchison, D. (2009). Designing your sample efficiently: Clustering effects in education surveys. *Educational Research*, *51*(1), 109–126.

Hyde, J.S., & Durik, A.M. (2005). Gender, competence, and motivation. In A.J. Elliot & C.S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 375-391). New York: Gilford Press.

Hyde, J. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592.

Isaac, J.D., Sansone, C., & Smith, J.L. (1999). Other people as a source of interest in an activity. *Journal of Experimental Social Psychology*, *35*(3), 239–265.

Izard, C.E. (1977). *Human emotions*. New York: Plenum Press.

Jenkins, E., & Pell, R. (2006). *The Relevance of Science Education (ROSE) project in England: A summary of findings*. Retrieved from http://www.ils.uio.no/english/rose/.

Kane, M.T. (2006). Validation. In R.L. Brennan (Ed.), *Educational measurement* (pp. 17-64). Westport, CT: American Council on Education and Praeger Press.

Kashdan, T., & Silvia, P. (2009). Curiosity and interest: The benefits of thriving on novelty and challenge. In C.R. Snyder & S. Lopez (Eds.), *Oxford handbook of positive psychology* (pp. 367-374). New York: Oxford University Press.

Keeves, J., & Alagumalai, S. (1997). New approaches to measurement. In G.N. Masters & J.P. Keeves (Eds.), *Advances in measurement in educational research and assessment* (pp. 23-42). Oxford: Pergamon.

Kiefer, A., & Sekaquaptewa, D. (2007). Implicit stereotypes, gender identification, and math-related outcomes. *Psychological Science*, *18*(1), 13–18.

Kish, L. (1965). *Survey sampling.* New York: Wiley.

Klem, A.M., & Connell, J.P. (2004). Relationships matter: Linking teacher support to students engagement and achievement. *The Journal of School Health*, *74*(7), 262–273.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*(5), 448–470.

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction*, *12*(4), 383–409.

Krapp, A. (2007). An educational-psychological conceptualisation of interest. *International Journal for Educational and Vocational Guidance*, *7*(1), 5–21.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology.* Beverly Hills: SAGE Publications.

Kunter, M., Baumert, J., & Köller, O. (2007). Effective classroom management and the development of subject-related interest. *Learning and Instruction*, *17*(5), 494–509.

Lane, D.M. (2006). Interative simulations in the teaching of statistics: Promise and pitfalls. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Available from: www.stat.auckland.ac.nz/~iase.

Larson, R., Moneta, G., Richards, M., & Wilson, S. (2002). Continuity, stability, and change in daily emotional experience across adolescence. *Child Development*, *73*(4), 1151–1165.

Lawless, K.A., & Kulikowich, J.M. (2006). Domain knowledge and individual interest: The effects of academic level and specialization in statistics and psychology. *Contemporary Educational Psychology*, *31*(1), 30–43.

Lee, C., & Famoye, F. (2006). Teaching statistics using a real time online database created by students. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Available from: www.stat.auckland.ac.nz/~iase.

Leong, J. (2006). *High school students' attitudes and beliefs regarding statistics in a service-learning-based statistics course.* Unpublished PhD thesis, Georgia State University, Atlanta, GA, USA.

Linacre, J. (1994). Sampe size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

Linacre, J. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*(3), 266–283.

Linacre, J. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, *3*(2), 103–122.

Linacre, J. (2006a). *A user's guide to Winsteps: Program manual.* Chicago: Winsteps.com.

Linacre, J. (2006b). Winsteps (3.61.2): Rasch measurement computer program. [Computer Software]. Chicago: Winsteps.com.

Litman, J. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, *44*(7), 1585–1595.

Lo Bianco, J. (2000). Multiliteracies and multilingualism. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 91-105). New York: Routledge.

Lock, R.H. (2006). Teaching an introductory statistics class based on sports examples. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Available from: www.stat.auckland.ac.nz/~iase.

Lopez, F.G., Brown, S.D., Lent, R.W., & Gore, P.A. (1997). Role of social-cognitive expectations in high school students' mathematics-related interest and performance. *Journal of Counseling Psychology*, *44*(1), 44–52.

Low, K., & Rounds, J. (2007). Interest change and continuity from early adolescence to middle adulthood. *International Journal for Educational and Vocational Guidance*, *7*(1), 23–36.

Lubinski, D., Benbow, C., & Morelock, M. (2000). Gender differences in engineering and the physcial sciences among the gifted: An inorganic-organic distinction. In K.A. Heller, F.J. Monks, R.J. Sternberg, & R.F. Subotnik (Eds.), *International handbook of giftedness and talent* (pp. 633-648). Oxford: Elsevier.

Ma, X. (2006). Cognitive and affective changes as determinants of taking advanced mathematics courses in high school. *American Journal of Education*, *113*(1), 123–149.

Ma, X., & Kishor, N. (1997). Assessing the relationship between the attitude towards mathematics and achievement in mathematics: A meta-analysis. *Journal for Research in Mathematics Education*, *28*(1), 26–47.

Ma, X., & Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *American Journal of Education*, *110*(3), 256–280.

Magee, B. (1973). *Popper*. London: Fontana/Collins.

Manor, O., Matthews, S., & Power, C. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology*, *19*(1), 149–157.

Marcoulides, G., Gottfried, A., Gottfried, A., & Oliver, P. (2008). A latent transition analysis of academic intrinsic motivation from childhood through adolescence. *Educational Research and Evaluation*, *14*(5), 411–427.

Marsh, H.W., Trautwein, U., Ludtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416.

Marsh, H. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*(1), 129–149.

McLeod, D.B. (1992). Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575-596). New York: MacMillan Publishing Co.

McNeely, C., & Falci, C. (2004). School connectedness and the transition into and out of health-risk behaviour among adolescents. *Journal of School Health*, *74*(7), 284–292.

McPhan, G., Morony, W., Pegg, J., Cooksey, R., & Lynch, T. (2008). *Maths? Why Not?* Retrieved from Australian Department of Education, Employment and Workplace Relations website: http://www.dest.gov.au/.

Meletiou-Mavrotheris, M., Lee, C., & Fouladi, R.T. (2007). Introductory statistics, college student attitudes and knowledge – a qualitative analysis of the impact of technology-based instruction. *International Journal of Mathematical Education in Science and Technology*, *38*(1), 65–83.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*(9), 741–749.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*(3), 424–436.

Mitchell, M. (1997). Situational interest in the statistics classroom. In *Proceedings of the Annual meeting of the American Educational Research Association.* Retrieved from ERIC database. (ED413345).

Mitchell, M., & Gilson, J. (1997). Interest and anxiety in mathematics. In *Proceedings of the Annual meeting of the American Educational Research Association.* Retrieved from ERIC database. (ED412116).

Moore, D.S. (1988). Should mathematicians teach statistics? *The College Mathematics Journal, 19*(1), 3–7.

Moore, D.S., & Cobb, G.W. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly, 107*(7), 615–630.

Murdock, D. (2009, July 13). Global cooling chills summer 2009. *National Review Online.* Available from: http://article.nationalreview.com.

Murphey, P.K., & Alexander, P.A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *The Journal of Experimental Education, 70*(3), 197–214.

Mvududu, N. (2003). A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course. *Journal of Statistics Education, 11*(3). Available from: http://www.amstat.org/publications/jse/.

National Curriculum Board (2009). *Shape of the Australian curriculum: Mathematics*. Retrieved from Australian Curriculum Assessment and Reporting Authority website: http://www.acara.edu.au.

Netemeyer, R.G., Bearden, W.O., & Sharma, S. (2003). *Scaling procedures: Issues and applications.* Thousand Oaks, CA: SAGE Publishing.

Oyserman, D. (2004). Self-concept and identity. In M.B. Brewer & M. Hewstone (Eds.), *Self and social identity* (pp. 5-24). Malden, MA: Blackwell Publishing.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Pscyhology Review*, *18*(4), 315–341.

Pekrun, R., Elliot, A., & Maier, M. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Pscyhology*, *101*(1), 115–135.

Perline, R., Wright, B., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement*, *3*(2), 237–255.

Perney, J., & Ravid, R. (1990). The relationship between attitudes towards statistics, math self-concept, test anxiety and graduate students' achievement in an introductory statistics course. In *Proceedings of the Annual Meeting of the American Educational Research Association.* Retrieved from ERIC database. (ED318607).

Pickard, A. (2007). *Research methods in information.* London: Facet Publishing.

Powell, K. (2004). Developmental psychology of adolescent girls: Conflicts and identity issues. *Education*, *125*(1), 77–87.

Printrich, P.R., & De Groot, E.V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40.

R Development Core Team (2009). R (2.10.1). [Computer Software]. Vienna, Austria: R Foundation for Statistical Computing.

Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Measurement Transactions*, *19*(1), 1012.

Ramdas, L. (1990). Women and literacy: A quest of justice. *Convergence*, *23*(1), 27–40.

Rao, C. (1975). Teaching of statistics at the secondary level: An interdisciplinary approach. *International Journal of Mathematical Education in Science and Technology*, *6*(2), 151–162.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: SAGE Publications.

Reeve, J. (1989). The interest-enjoyment distinction in intrinsic motivation. *Motivation and Emotion*, *13*(2), 83–103.

Renninger, K.A. (2000). Individual interest and its implications for understanding intrinsic motivation. In C. Sansone & J.M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 373-404). San Diego: Academic Press.

Renninger, K. (2009). Interest and identity in instruction: An inductive model. *Educational Psychologist*, *44*(2), 105–118.

Robbins, S.B., Le, H., Davis, D., Lauver, K., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288.

Rosenberg, E.L. (1998). Levels of affect and the organization of affect. *Review of General Psychology*, *2*(3), 247–270.

Rottinghaus, P., Larson, L., & Borgen, F. (2003). The relation of self-efficacy and interests: A meta-analysis of 60 samples. *Journal of Vocational Behavior*, *62*(4), 221–236.

Rumsey, D.J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10*(3). Available from: http://www.amstat.org/publications/jse/.

Ryan, K., & Ryan, A. (2005). Psychological processes underlying stereotype threat and standardised math test performance. *Educational Psychologist*, *40*(1), 53–63.

Ryan, R.M., & Deci, E.L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

Ryan, R.M., & Deci, E.L. (2000b). When rewards compete with nature: The undermining of intrinsic motivation and self-regulation. In C. Sansone & J.M. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 13-54). San Diego: Academic Press.

Schau, C., Stevens, J., Dauphinee, T.L., & Del Vecchio, A. (1995). The development and validation of the survey of attitudes towards statistics. *Educational and Psychological Measurement*, *55*(5), 868–875.

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Pscyhologist*, *26*(3), 299–323.

Schiefele, U. (1992). Topic interest and levels of text comprehension. In K.A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 151-182). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schiefele, U., & Csikszentmihalyi, M. (1995). Motivation and ability as factors in mathematics experience and achievement. *Journal for Research in Mathematics Education*, *26*(2), 163–181.

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K.A. Renninger, S. Hidi, & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 183-212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review 6*(4). Retrieved from: http://web.augsburg.edu/~schield/.

Schou, S.B. (2007). A study of student attitudes and performance in an online introductory business statistics class. *Electronic Journal for the Integration of Technology in Education 6*. Available from: http://ejite.isu.edu/.

Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review*, *13*(3), 211–224.

Schunk, D. (1996). *Learning theories: An educational perspective*. Englewood Cliffs, NJ: Prentice-Hall.

Sciutto, M. (1999). Student-centered methods for decreasing anxiety and increasing interest level in undergraduate statistics courses. *Journal of Instructional Psychology, 22*(3). Available from Academic Search Premier.

Silvia, P.J. (2001). Interest and interests: The psychology of constructive capriciousness. *Review of General Psychology*, *5*(3), 270–290.

Silvia, P.J. (2003). Self-efficacy and interest: Experimental studies of optimal incompetence. *Journal of Vocational Behavior*, *62*(4), 237–249.

Smith, E. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement*, *2*(3), 281–311.

Smith, J., Sansone, C., & White, P. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Pscyhology*, *99*(1), 99–114.

Smith, R. (1991). The distributional properties of Rasch item fit statistics. *Educational and Psychological Measurement*, *51*(3), 541–565.

Smith, R., & Miao, C. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement theory into practice*. Greenwich: Ablex.

Sorge, C., & Schau, C. (2002). Impact of engineering students' attitudes on achievement in statistics: A structural model. In *Proceedings of the Annual Meeting of the American Educational Research Association*. Retrieved from: http://evaluationandstatistics.com/references.html.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stevens, T., & Olivaréz, A. (2005). Development and evaluation of the Mathematics Interest Inventory. *Measurement and Evaluation in Counseling and Development*, *38*(3), 141–152.

Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Educational Technology and Society*, *7*(3), 9–20.

Tempelaar, D. (2006). A structural equation model analyzing the relationship of students' statistical reasoning abilities, their attitudes toward statistics, and learning approaches. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Available from: www.stat.auckland.ac.nz/~iase.

Thomas, R. (2005). *High stakes testing: Coping with collateral damage.* Mahwah, NJ: Lawrence Erlbaum Associates.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis.* Washington, DC: American Psychological Association.

Townsend, K.C., & McWhirter, B.T. (2005). Connectedness: A review of the literature with implications for counseling, assessment, and research. *Journal of Counseling and Development*, *83*(2), 191–201.

Trautwein, U., Ludtke, O., Köller, O., Marsh, H.W., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806.

Trewin, D. (2005). Improving statistical literacy: The respective roles of schools and the national statistical offices. In M. Coupland, J. Anderson, & T. Spencer (Eds.), *Proceedings of the 20th Biennial Conference of the Australian Association of Mathematics Teachers.* Adelaide: AAMT.

Tsai, Y., Kunter, M., Lüdtke, O., Trautwein, U., & Ryan, R. (2008). What makes lessons interesting? The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, *100*(2), 460–472.

Turner, N. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, *58*(4), 541–568.

Usher, E., & Pajeres, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology*, *31*(2), 125–141.

Wallman, K.K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1–8.

Watson, J.M. (1997). Assessing statistical literacy using the media. In I. Gal & J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 177-121). Amsterdam: IOS Press.

Watson, J.M. (2006). *Statistical literacy at school: Growth and goals.* New Jersey: Lawrence Erlbbaum Associates.

Watson, J.M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.

Watt, H.M.G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th through 11th grade Australian students. *Child Development*, *75*(5), 1556–1574.

Watt, H. (2005). Explaining gendered math enrolments for NSW Australian secondary school students. *New Directions for Child and Adolescent Development*, *110*(1), 15–29.

Watt, H. (2008). A latent growth curve modeling approach using an accelerated longitudinal design: The ontogeny of boys' and girls' talent perceptions and intrinsic values through adolescence. *Educational Research and Evaluation*, *14*(4), 287–304.

Whitlock, J.L. (2006). Youth perceptions of life at school: Contextual correlates of school connectedness in adolescence. *Applied Development Science*, *10*(1), 13–29.

Wigfield, A., Byrnes, J.P., & Eccles, J.S. (2006). Development during early and middle adolescence. In P.A. Alexander & P.H. Winne (Eds.), *Handbook of educational psychology* (pp. 87-113). Mahwah, NJ: Lawrence Erlbaum Associates.

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review*, *30*(1), 1–35.

Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.

Wigfield, A., Tonks, S., & Eccles, J.S. (2004). Expectancy value theory in cross-cultural perspective. In D.M. McInerney & S. Van Etten (Eds.), *Big theories revisited: Volume 4 of sociocultural influences on motivation and learning*. Greenwich: Information Age Publishing Inc.

Wild, C. (1994). Embracing the "wider view" of statistics. *The American Statistician*, *48*(2), 163–171.

Wolfe, E., & Smith, E. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part 1 – instrument development tools. *Journal of Applied Measurement*, *8*(1), 97–123.

Wolfe, E., & Smith, E. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part 2 – validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

Wright, B., & Stone, M. (1999). *Measurement Essentials*. Wilmington, DE: Wide Range Inc.

Wright, B. (2000). How to set standards. *Rasch Measurement Transactions*, *14*(1), 740.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (1998). Conquest (2.0): Generalised item response modeling software. [Computer Software]. Melbourne: ACER.

Wu, M., & Adams, R. (2006). Modelling mathematics problem solving item responses using a multidimensional irt model. *Mathematics Education Research Journal*, *18*(2), 93–113.

Appendix A

Study questionnaire

Student details

Please complete the following details. (Print your answers neatly.)

1. Family name:

2. Given name:

3. Age (in years and months):

4. Are you are boy or girl?

5. Year level (or grade) at school:

6. School name:

Statistical Literacy Interest Inventory

The following survey seeks to find out how you feel about using statistics, which includes activities such as doing surveys, making graphs and tables, working out averages, calculating chance. These are not just done in your maths class! Each question is written in the form of a description and you need to indicate how similar you are to each description. Please answer using a number from 1 to 5, when 1 stands for "doesn't describe me at all" and 5 stands for "describes me well". Use a number between 1 and 5 if the description is similar to you some of the time.

How similar are you to the descriptions below? (Circle the number of your choice)

	I'm interested in:	
R1	Doing magazine or online surveys.	1 2 3 4 5
R2	Surveys that find out how people feel about things.	1 2 3 4 5
R3	Working on problems involving data and statistics.	1 2 3 4 5
R4	Looking up unusual statistics.	1 2 3 4 5
R6b	Using averages to compare sports teams or players.	1 2 3 4 5
R7	The average rainfall for my home area.	1 2 3 4 5
R9	Reading graphs in newspaper and magazine reports.	1 2 3 4 5
R10	Conducting surveys of other students at my school.	1 2 3 4 5
R11	Working out the probabilities (or chances) for dice, coins and spinners.	1 2 3 4 5
R12b	Using computer programs to help me investigate problems involving data.	1 2 3 4 5
R13	Using statistics to prove a point or win an argument.	1 2 3 4 5
R14	Learning more about statistics.	1 2 3 4 5
R15	Getting a job that involves statistics.	1 2 3 4 5
	I would like to know:	
C16	How scientists calculate the chance of rain.	1 2 3 4 5
C17	How a survey can be used to predict who will win the next election.	1 2 3 4 5
C19	How politicians make decisions that are based on data.	1 2 3 4 5
C20	Whether a survey reported on the radio or TV about students was correct.	1 2 3 4 5
C21	Whether a game I was playing that used dice or spinners was fair.	1 2 3 4 5
C22	How a graph could be used to compare my sports team with other teams.	1 2 3 4 5
C38	All there is to know about statistics.	1 2 3 4 5

	It's important to me personally that I:	
I23	Can understand news reports that use averages.	1 2 3 4 5
I24	Know how to calculate the chance of being injured from risky behavior.	1 2 3 4 5
I25	Understand the words used in statistics.	1 2 3 4 5
I26	Can believe scientific claims that are based on data.	1 2 3 4 5
I27	Use the correct graph when displaying my data.	1 2 3 4 5
I28	Can understand graphs that appear on the internet or in newspapers.	1 2 3 4 5
I29	Can arrange data into tables.	1 2 3 4 5
I30b	Can use data to investigate questions that I might have.	1 2 3 4 5

	Other descriptions:	
R31	I get so involved when I work with data that sometimes I lose all sense of time.	1 2 3 4 5
R36	I like to work on statistics problems in my spare time.	1 2 3 4 5

	Additional items:	
IE42	Compared to others in my class I am good at maths.	1 2 3 4 5
IE43	Out of all my subjects I usually get my best marks in maths.	1 2 3 4 5
IE44	I find statistics more interesting that other work we do in maths.	1 2 3 4 5
IE45	The statistics that I do in maths classes is more interesting than the statistics that I do in other subjects.	1 2 3 4 5

Self-efficacy items

	I am confident that I am able to:	
S41b	Solve problems that use averages.	1 2 3 4 5
S42	Find when a newspaper article has used the wrong type of average.	1 2 3 4 5
S43	Explain to a friend how probability (or chance) is calculated.	1 2 3 4 5
S44	Show data correctly on a bar chart.	1 2 3 4 5
S45	Explain the meaning of a graph in a newspaper or on the internet.	1 2 3 4 5
S46	Find a mistake in someone else's graph.	1 2 3 4 5
S47b	Explain when conclusions that are based on surveys might be wrong.	1 2 3 4 5
S48c	Look up the correct number from a table of numbers.	1 2 3 4 5
S49	Explain how to select a fair sample of students for a school survey.	1 2 3 4 5
S50c	Work out the most likely outcome from a game involving chance.	1 2 3 4 5

Maths survey

M1	I like to answer questions in maths classes.	1	2	3	4	5
M2	I like maths.	1	2	3	4	5
M3	I am interested in maths.	1	2	3	4	5
M4	I find that knowing a lot about maths is helpful.	1	2	3	4	5
M5	I feel good when it comes to working on maths.	1	2	3	4	5
M6	I want to know all about how to do maths problems.	1	2	3	4	5
M7	I feel excited when a new maths topic is announced.	1	2	3	4	5
M8	I want to learn more about maths.	1	2	3	4	5
M9	I choose to work on maths.	1	2	3	4	5
M10	I want to know all about maths.	1	2	3	4	5
M11	Compared with other students in my maths class I expect to do well.	1	2	3	4	5
M12	I'm certain I can understand the ideas taught in my maths class.	1	2	3	4	5
M13	I expect to do very well in my maths class.	1	2	3	4	5
M14	Compared with others in my class, I think I'm a good maths student.	1	2	3	4	5
M15	I am sure I can do an excellent job on the problems and tasks assigned for my maths class.	1	2	3	4	5
M16	I think I will receive a good grade for maths.	1	2	3	4	5
M17	My study skills are excellent compared with others in my maths class.	1	2	3	4	5
M18	Compared with other students in my class I think I know a great deal about maths.	1	2	3	4	5
M19	I know that I will be able to learn the material for my maths class.	1	2	3	4	5

Appendix B

Main study results

The following is a list of tables that appear in this appendix. This list also provides a brief description of each table.

- Table B.1 – SLIM item statistics based on pilot study.
- Table B.2 – SESL item statistics based on pilot study.
- Table B.3 – SLIM item statistics based on pooled sample.
- Table B.4 – SLIM category statistics based on pooled sample.
- Table B.5 – SLIM threshold estimates.
- Table B.6 – results of exploratory factor analysis of SLIM items.
- Table B.7 – gender differences for SLIM items.
- Table B.8 – year level differences for SLIM items.
- Table B.9 – StatSmart attendance differences for SLIM items.
- Table B.10 – SESL item statistics.
- Table B.11 – SESL category statistics.
- Table B.12 – gender differences for SESL items.
- Table B.13 – year level differences for SESL items.
- Table B.14 – StatSmart attendance differences for SESL items.

Tables B.1 and B.2 report the items and fit statistics for SLIM and SESL respectively that are based on the pilot study. In particular the tables report the item code, the number of valid student responses to the items (N), the estimated difficulty of each item (δ_i), the standard error of this difficulty estimate (SE[δ_i]), the infit statistic (u_i), the standardised version of the infit statistic (Z_u), the outfit statistic (v_i), and its standardised version (Z_v).

Table B.1

SLIM selected statistics based on pilot study

Item ID	N	δ_i	SE(δ_i)	u_i	Z_u	v_i	Z_v
R31	78	0.94	0.14	0.95	-0.23	0.89	-0.44
C38	81	0.71	0.13	0.79	-1.32	0.79	-1.13
R15	220	0.61	0.08	0.97	-0.24	0.95	-0.40
C19	220	0.49	0.07	0.94	-0.62	0.84	-1.52
R2	220	0.32	0.07	1.12	1.32	1.25	2.21
R14	221	0.30	0.07	0.84	-1.86	0.95	-0.49
R9	220	0.24	0.07	1.22	2.32	1.23	2.16
R11	221	0.24	0.07	0.96	-0.42	1.09	0.93
R3	220	0.16	0.07	0.98	-0.15	1.22	2.09
C17	221	0.13	0.07	0.95	-0.52	0.93	-0.66
R12b	81	0.07	0.11	1.29	1.92	1.37	2.16
I23	221	-0.01	0.07	0.87	-1.53	0.93	-0.76
I25	221	-0.07	0.07	0.82	-2.20	0.86	-1.56
C20	221	-0.18	0.07	1.00	0.02	0.96	-0.40
C16	221	-0.19	0.07	1.19	2.10	1.47	4.40
C21	221	-0.19	0.07	1.17	1.92	1.15	1.54
I28	221	-0.50	0.07	0.90	-1.18	0.89	-1.22
I30b	167	-0.51	0.08	1.10	1.04	1.08	0.73
I27	221	-0.63	0.07	1.06	0.74	1.12	1.25
I24	221	-0.64	0.07	1.00	0.06	0.96	-0.36
I29	221	-0.69	0.07	1.03	0.34	1.12	1.28
I26	221	-0.81	0.07	1.18	1.97	1.16	1.63

Table B.2

SESL selected statistics based on pilot study

Item ID	N	δ_i	SE(δ_i)	u_i	Z_u	v_i	Z_v
S42	221	0.80	0.07	0.96	-0.45	0.98	-0.15
S47b	81	0.64	0.12	0.77	-1.62	0.85	-0.95
S43	221	0.30	0.07	0.99	-0.09	0.98	-0.15
S45	220	0.21	0.07	0.93	-0.72	0.93	-0.75
S46	220	0.03	0.07	1.04	0.47	1.01	0.15
S49	221	0.03	0.07	0.95	-0.49	0.92	-0.82
S50b	167	-0.27	0.08	1.25	2.28	1.28	2.35
S41b	80	-0.44	0.12	1.09	0.61	1.18	1.10
S44	221	-0.51	0.08	1.18	1.87	1.13	1.34
S48b	81	-0.79	0.13	0.63	-2.66	0.63	-2.50

Table B.3 reports the items and fit statistics for SLIM that are based on the pooled sample. As noted, responses from 17 students were removed from the analysis because they had answered fewer than eight of the 16 items. As a result the total number of student responses available for analysis was 774. The table reports the number of valid student responses to the items (N), the estimated difficulty of each item (δ_i), the standard error of this difficulty estimate (SE[δ_i]), the infit statistic (u_i), the standardised version of the infit statistic (Z_u), the outfit statistic (v_i), and its standardised version (Z_v).

Table B.4, reports the category statistics for SLIM. In particular it reports the number of responses in each category (N), the percentage response for each category, the estimated value of the thresholds (τ_k), and the standard error of this estimate [SE(τ_k)]. These statistics are based on student responses from the pooled sample.

Table B.3

SLIM selected statistics

Item ID	N	δ_i	SE(δ_i)	u_i	Z_u	v_i	Z_v
R15	766	0.76	0.04	1.14	2.61	1.20	3.12
C38	633	0.53	0.04	1.05	0.92	1.02	0.34
C19	771	0.43	0.04	0.99	-0.21	0.92	-1.37
R14	772	0.42	0.04	0.87	-2.76	0.97	-0.53
R3	772	0.39	0.04	1.03	0.71	1.17	3.01
C17	770	0.09	0.04	1.13	2.65	1.09	1.62
C16	772	0.00	0.04	1.19	3.85	1.25	4.55
C20	774	-0.05	0.04	1.11	2.33	1.09	1.69
I23	773	-0.05	0.04	0.89	-2.40	0.92	-1.54
I25	765	-0.07	0.04	0.79	-4.80	0.81	-4.00
I24	773	-0.25	0.04	1.15	3.13	1.16	3.07
I26	769	-0.33	0.04	1.05	1.14	1.06	1.26
I30	714	-0.35	0.04	0.92	-1.61	0.91	-1.82
I28	772	-0.46	0.04	0.86	-2.98	0.87	-2.66
I27	767	-0.51	0.04	0.94	-1.33	0.99	-0.26
I29	771	-0.54	0.04	0.97	-0.62	1.00	0.09

Table B.4

Category statistics for SLIM

	Responses per category		Thresholds	
Category	N	(%)	τ_k	SE(τ_k)
1	2451	20	None	
2	2711	22	-1.44	0.03
3	3080	25	-0.46	0.02
4	2369	19	0.41	0.02
5	1369	11	1.48	0.03
(no response)	244	2		

Table B.5 shows the threshold estimates for SLIM, where τ_i are defined in Section 4.4 of Chapter 4.

Table B.5

Threshold estimates for SLIM

Item ID	τ_2	τ_3	τ_4	τ_5
R15	-0.68	0.30	1.17	2.24
C38	-0.91	0.07	0.94	2.01
C19	-1.01	-0.03	0.84	1.91
R14	-1.02	-0.04	0.83	1.90
R3	-1.05	-0.07	0.80	1.87
C17	-1.35	-0.37	0.50	1.57
C16	-1.44	-0.46	0.41	1.48
C20	-1.49	-0.51	0.36	1.43
I23	-1.49	-0.51	0.36	1.43
I25	-1.51	-0.53	0.34	1.41
I24	-1.69	-0.71	0.16	1.23
I26	-1.77	-0.79	0.08	1.15
I30	-1.79	-0.81	0.06	1.13
I28	-1.90	-0.92	-0.05	1.02
I27	-1.95	-0.97	-0.10	0.97
I29	-1.98	-1.00	-0.13	0.94

Exploratory factor analysis

The number of factors extracted was determined using parallel analysis, a method which reportedly indicates the correct number of factors more frequently than either Kaiser's rule or the Scree test (Thompson, 2004; Turner, 1998). In this instance the analysis suggested three factors could be extracted. The solution was then rotated using the varimax solution. Loadings that were smaller than 0.3 were ignored, as for a sample of this size smaller loadings are not significantly different from zero (Stevens, 2002). The three factor solution, shown in Table B.6, explained 61% of the variance.

Table B.6

Results of exploratory factor analysis

Item	Component 1	Component 2	Component 3
R3			0.76
R14			0.73
R15			0.72
C38		0.48	0.57
C16		0.70	
C17		0.80	
C19		0.81	
C20		0.74	
I23	0.63		
I24	0.54	0.39	
I25	0.66		0.37
I26	0.68		
I27	0.74		
I28	0.78		
I29	0.77		
I30b	0.74		

Table B.7, reports the estimated item difficulties for both males and females based on responses from the pooled sample. Given that there were 16 pairwise comparisons, the Bonferroni adjustment reduced the critical value to $0.05/16 = 0.003$ at the 5% level. In particular, the table reports the estimated item difficulties based on male responses (δ_m), the standard error of these estimates (SE[δ_m]), the estimated item difficulties based on female responses (δ_f), the standard error of these estimates (SE[δ_f]), the difference in item estimates ($\delta_m - \delta_f$), the standard error of this difference (SE[$\delta_m - \delta_f$]), the t-statistic for this difference (t) and its estimated p-value (p). Statistically significant differences are emboldened.

Table B.7

SLIM item difficulties by gender

	Male		Female		Difference			
Item	δ_m	SE(δ_m)	δ_f	SE(δ_f)	$\delta_m - \delta_f$	SE($\delta_m - \delta_f$)	t	p
R3	0.17	0.06	0.58	0.06	-0.41	0.08	-5.14	**0.00**
R14	0.37	0.06	0.47	0.05	-0.10	0.08	-1.26	0.21
R15	0.67	0.06	0.83	0.06	-0.17	0.08	-2.00	0.05
C16	0.06	0.06	-0.05	0.05	0.11	0.08	1.46	0.14
C17	0.23	0.06	-0.03	0.05	0.26	0.08	3.37	**0.00**
C19	0.50	0.06	0.37	0.05	0.13	0.08	1.70	0.09
C20	0.11	0.06	-0.19	0.05	0.30	0.08	3.97	**0.00**
I23	-0.07	0.06	-0.04	0.05	-0.04	0.08	-0.49	0.62
I24	-0.14	0.06	-0.34	0.05	0.20	0.08	2.61	0.01
I25	-0.09	0.06	-0.04	0.05	-0.05	0.08	-0.66	0.51
I26	-0.38	0.06	-0.29	0.05	-0.09	0.08	-1.13	0.26
I27	-0.49	0.06	-0.52	0.05	0.03	0.08	0.44	0.66
I28	-0.53	0.06	-0.40	0.05	-0.13	0.08	-1.68	0.09
I29	-0.51	0.06	-0.56	0.05	0.05	0.08	0.70	0.48
I30b	-0.40	0.06	-0.31	0.05	-0.09	0.08	-1.14	0.25
C38	0.48	0.06	0.58	0.06	-0.10	0.09	-1.15	0.25

Table B.8 reports statistics for items showing significant DIF by year level. Given that there were 160 pairwise comparisons only the ten most extreme differences are reported. In this instance, the Bonferroni adjustment reduced the critical value to $0.05/160 = 0.0003$ at the 5% level. The table reports the item code, the year level groups being compared, the difference in the item difficulties ($\delta_1 - \delta_2$), the standard error of this difference [SE($\delta_1 - \delta_2$)], the t-statistic for this difference (t), and its associated p-value. Positive differences indicate that the first group found the item more difficult than the second group. Statistically significant differences are emboldened.

Table B.8

SLIM item difficulties by year level

Item	Year levels	$\delta_1 - \delta_2$	SE($\delta_1 - \delta_2$)	t	p
R3	7, 9	-0.29	0.10	-2.85	0.0045
R15	7, 8	-0.37	0.10	-3.57	0.0004
R15	7, 9	-0.56	0.11	-5.14	**0.0000**
R15	7, 10	-0.64	0.20	-3.24	0.0014
I24	7, 10	-0.51	0.18	-2.91	0.0040
I26	7, 8	0.44	0.10	4.50	**0.0002**
I26	7, 9	0.60	0.10	5.89	**0.0000**
I26	7, 10	0.46	0.18	2.62	0.0092
I28	7, 9	0.28	0.10	2.74	0.0063
C38	7, 9	-0.38	0.12	-3.27	0.0012

Table B.9, reports the estimated item difficulties for students attending StatSmart schools and those not attending these schools. Given that there were 16 pairwise comparisons, the Bonferroni adjustment reduced the critical value to $0.05/16 = 0.003$ at the 5% level. In particular the table reports the estimated item difficulties based on StatSmart responses (δ_s), the standard error of these estimates (SE[δ_s]), the estimated item difficulties based on Non-StatSmart responses (δ_n), the standard error of these estimates (SE[δ_n]), the difference in item estimates ($\delta_s - \delta_n$), the standard error of this difference (SE[$\delta_s - \delta_n$]), the t-statistic for this difference (t) and its estimated p-value (p). Statistically significant differences are emboldened.

Table B.9

SLIM item difficulties by attendance or otherwise at StatSmart school.

	StatSmart		Non-StatSmart		Difference			
Item	δ_s	SE(δ_s)	δ_n	SE(δ_n)	$\delta_s - \delta_n$	SE($\delta_s - \delta_n$)	t	p
R3	0.39	0.05	0.37	0.06	0.02	0.08	0.21	0.83
R14	0.38	0.05	0.48	0.06	-0.10	0.08	-1.21	0.23
R15	0.75	0.05	0.76	0.07	-0.01	0.09	-0.16	0.87
C16	0.08	0.05	-0.12	0.06	0.20	0.08	2.56	0.01
C17	0.03	0.05	0.17	0.06	-0.14	0.08	-1.75	0.08
C19	0.35	0.05	0.55	0.06	-0.20	0.08	-2.44	0.02
C20	-0.02	0.05	-0.10	0.06	0.09	0.08	1.10	0.27
I23	-0.19	0.05	0.16	0.06	-0.35	0.08	-4.46	**0.00**
I24	-0.14	0.05	-0.42	0.06	0.28	0.08	3.53	**0.00**
I25	-0.12	0.05	0.01	0.06	-0.12	0.08	-1.57	0.12
I26	-0.18	0.05	-0.55	0.06	0.38	0.08	4.78	**0.00**
I27	-0.49	0.05	-0.54	0.06	0.05	0.08	0.67	0.50
I28	-0.48	0.05	-0.43	0.06	-0.06	0.08	-0.74	0.46
I29	-0.53	0.05	-0.55	0.06	0.02	0.08	0.30	0.77
I30b	-0.35	0.05	-0.36	0.07	0.01	0.08	0.18	0.86
C38	0.50	0.05	0.63	0.09	-0.14	0.10	-1.36	0.18

Table B.10 shows the items and fit statistics for SESL that are based on student responses from the pooled sample. As noted, responses from four students were removed from the analysis because they had answered fewer than five of the nine items. As a result, the total number of student responses available was 787. The table reports the number of valid student responses to the items (N), the estimated difficulty of each item (δ_i), the standard error of this difficulty estimate (SE[δ_i]), the infit statistic (u_i), the standardised version of the infit statistic (Z_u), the outfit statistic (v_i), and its standardised version (Z_v).

Table B.10

SESL selected statistics

Item ID	N	δ_i	SE(δ_i)	u_i	Z_u	v_i	Z_v
S42	783	0.75	0.04	0.93	-1.50	1.00	0.00
S47b	645	0.50	0.05	0.80	-3.90	0.83	-3.00
S43	785	0.14	0.04	1.01	0.20	0.96	-0.70
S45	781	0.07	0.04	0.91	-1.80	0.90	-1.90
S46	783	0.06	0.04	0.98	-0.40	0.96	-0.90
S48c	419	0.05	0.06	0.96	-0.60	0.93	-1.00
S49	783	-0.07	0.04	1.03	0.70	1.03	0.60
S50c	423	-0.41	0.06	1.27	3.70	1.24	3.20
S41c	646	-0.48	0.05	1.11	1.90	1.10	1.60
S44	785	-0.61	0.04	1.10	2.00	1.02	0.50

Table B.11 shows the category statistics for SESL. In particular it reports the number of responses in each category (N), the percentage response for each category, the estimated value of the thresholds (τ_k), and the standard error of this estimate (SE[τ_k]).

Table B.12 reports the estimated item difficulties by gender for students from the pooled sample. Given that there were 10 pairwise comparisons, the Bonferroni adjustment reduced the critical value to $0.05/10 = 0.005$ at the 5% level. In particular the table reports the estimated item difficulties based on female

Table B.11

Category statistics for SESL

Category	Responses per category N	(%)	Thresholds τ_k	SE(τ_k)
1	1007	13	None	
2	1323	17	-1.71	0.04
3	1790	23	-0.60	0.03
4	1614	21	0.47	0.03
5	913	12	1.83	0.04
(no response)	993	13		

responses (δ_f), the standard error of these estimates (SE[δ_f]), the estimated item difficulties based on male responses (δ_m), the standard error of these estimates (SE[δ_m]), the difference in item estimates ($\delta_f - \delta_m$), the standard error of this difference (SE[$\delta_f - \delta_m$]), the t-statistic for this difference (t) and its estimated p-value (p).

Table B.12

SESL item difficulties by gender

	Female		Male		Difference			
Item	δ_f	SE(δ_f)	δ_m	SE(δ_m)	$\delta_f - \delta_n$	SE($\delta_f - \delta_m$)	t	p
S41b	-0.40	0.06	-0.57	0.07	0.18	0.10	1.86	0.06
S42	0.78	0.06	0.72	0.06	0.06	0.09	0.69	0.49
S43	0.09	0.06	0.19	0.06	-0.10	0.08	-1.17	0.24
S44	-0.64	0.06	-0.57	0.07	-0.08	0.09	-0.87	0.38
S45	0.06	0.06	0.07	0.06	-0.01	0.08	-0.12	0.91
S46	0.12	0.06	-0.01	0.06	0.13	0.08	1.52	0.13
S47b	0.48	0.06	0.52	0.07	-0.04	0.09	-0.48	0.63
S48c	0.02	0.08	0.08	0.08	-0.06	0.11	-0.50	0.62
S49	-0.17	0.06	0.06	0.06	-0.23	0.08	-2.77	0.01
S50c	-0.27	0.08	-0.57	0.09	0.30	0.12	2.54	0.01

Table B.13 reports the six most extreme differences in item difficulty estimates based on year level at school. Given that there were 100 pairwise comparisons, the Bonferroni adjustment reduced the critical value to 0.0005 and consequently no comparisons were statistically significant. Three of those reported concern students from the only Year 6 class, which may be atypical.

Table B.13

SESL item difficulties by year level

Item	Year levels	$\delta_1 - \delta_2$	SE($\delta_1 - \delta_2$)	t	p
S50c	6, 10	-0.98	0.31	-3.18	0.002
S50c	6, 9	-0.74	0.28	-2.68	0.008
S50c	6, 8	-0.73	0.28	-2.61	0.010
S50c	7, 10	-0.50	0.20	-2.50	0.013
S46	7, 9	-0.29	0.11	-2.67	0.008
S41c	7, 9	0.30	0.12	2.43	0.012

Table B.14, reports the estimated item difficulties for students attending StatSmart schools and those not attending these schools. Given that there were 10 pairwise comparisons, the Bonferroni adjustment reduced the critical value to $0.05/10 = 0.005$ at the 5% level. In particular the table reports the estimated item difficulties based on StatSmart responses (δ_s), the standard error of these estimates (SE[δ_s]), the estimated item difficulties based on Non-StatSmart responses (δ_n), the standard error of these estimates (SE[δ_n]), the difference in item estimates ($\delta_s - \delta_n$), the standard error of this difference (SE[$\delta_s - \delta_n$]), the t-statistic for this difference (t) and its estimated p-value (p).

Table B.14

SESL item difficulties by attendance or otherwise at StatSmart school.

	StatSmart		Non-StatSmart		Difference			
Item	δ_s	SE(δ_s)	δ_n	SE(δ_n)	$\delta_s - \delta_n$	SE($\delta_s - \delta_n$)	t	p
S41c	-0.48	0.06	-0.47	0.09	-0.01	0.11	-0.11	0.91
S42	0.71	0.05	0.82	0.07	-0.12	0.09	-1.32	0.19
S43	0.12	0.05	0.15	0.07	-0.03	0.09	-0.33	0.74
S44	-0.56	0.06	-0.69	0.07	0.13	0.09	1.48	0.14
S45	0.01	0.05	0.15	0.07	-0.13	0.09	-1.55	0.12
S46	0.12	0.05	-0.02	0.07	0.14	0.09	1.64	0.10
S47b	0.49	0.05	0.52	0.09	-0.02	0.11	-0.20	0.84
S48c	0.07	0.06	-0.03	0.12	0.10	0.14	0.69	0.49
S49	-0.09	0.05	-0.03	0.07	-0.06	0.09	-0.66	0.51
S50c	-0.38	0.07	-0.51	0.13	0.13	0.14	0.94	0.35

Appendix C

Results of linear models

Diagnostic plots

Figures C.1, C.2 and C.3 show the diagnostic plots for the models reported as Equations 6.1, 6.4 and 6.5 respectively. The top plot in each figure shows the sample quantiles against the quantiles estimated from a theoretical normal distribution. These plots assess the normality of the residuals and should ideally be linear in each case. The second plot in each figure shows the residuals against the predicted values. These plots assess the homogeneity of the variance in the residuals and should ideally show uniform scatter across the range of predicted values.

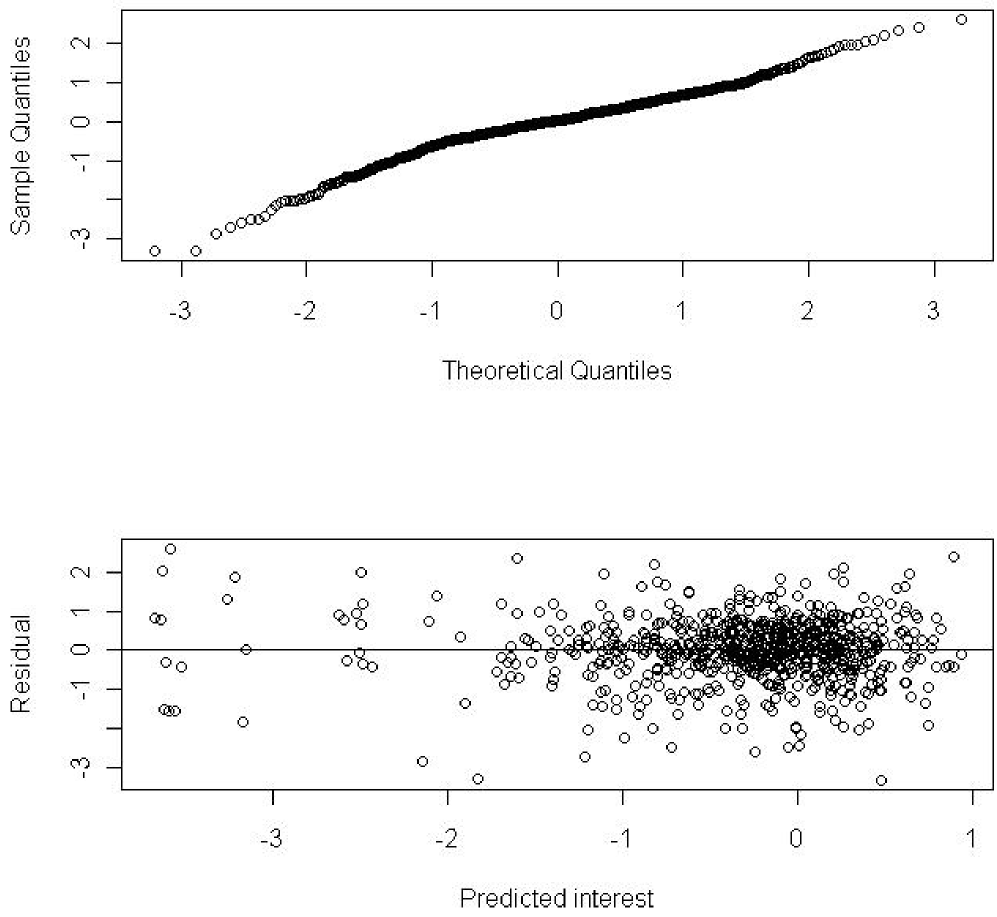
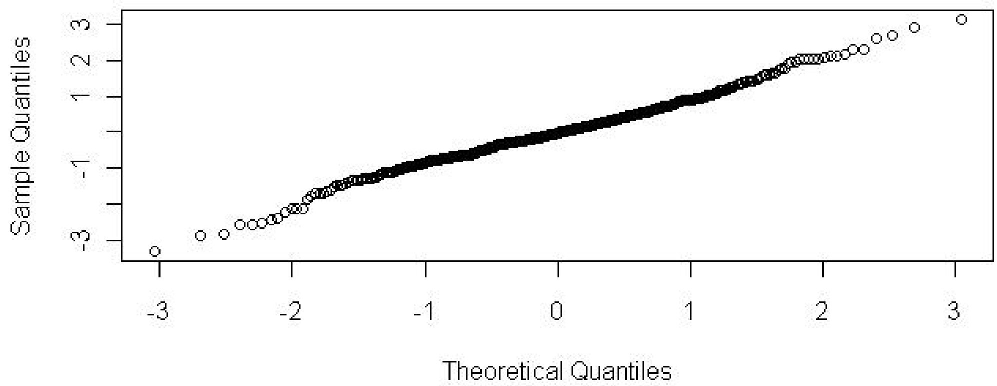


Figure C.1. Diagnostic plots for interest model shown in Equation 6.1



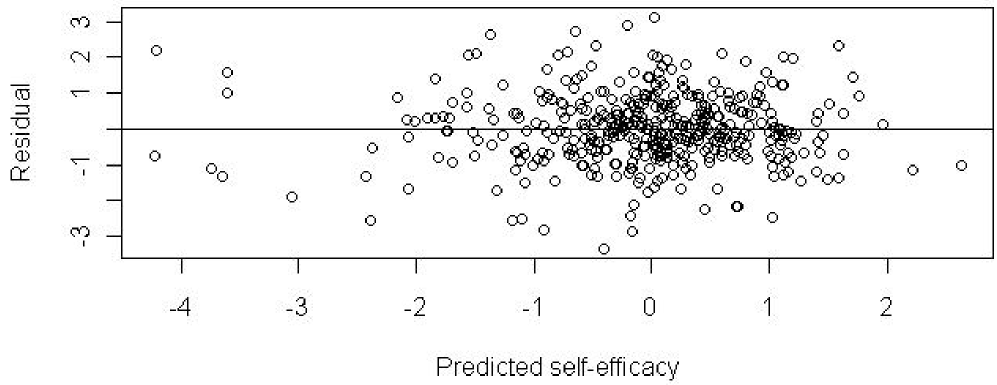


Figure C.2. Diagnostic plots for self-efficacy model shown in Equation 6.4

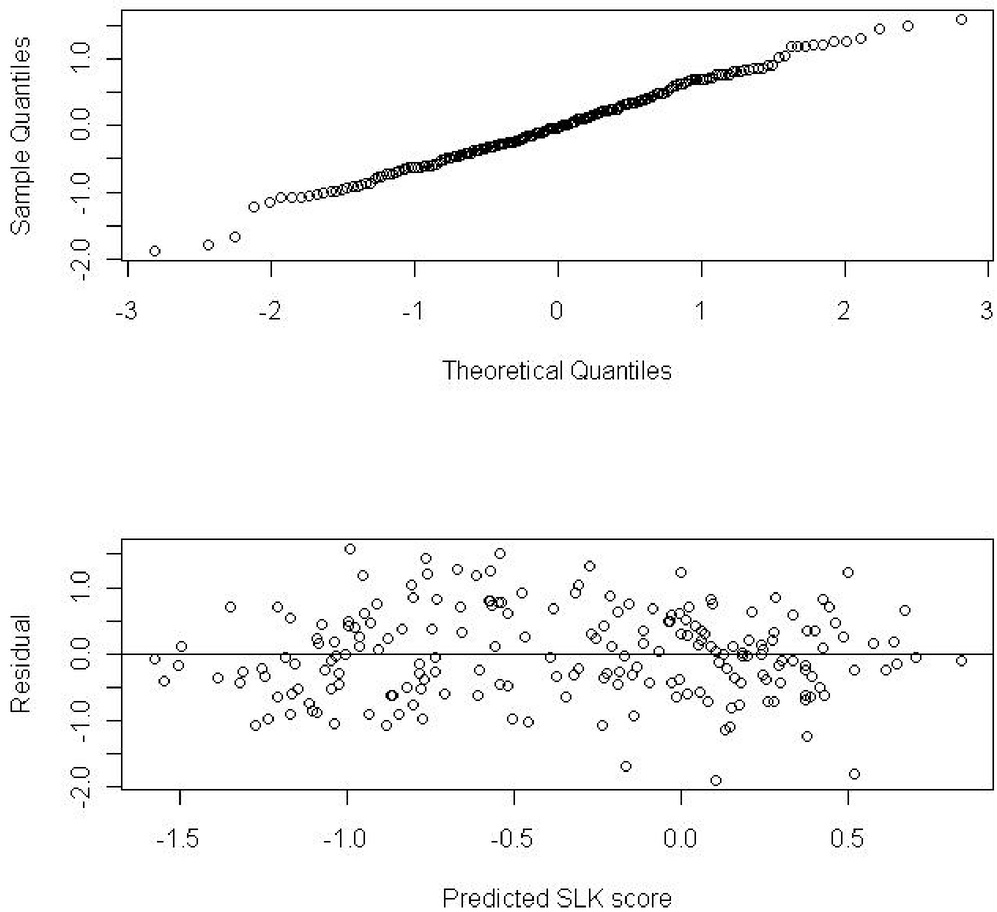


Figure C.3. Diagnostic plots for SLK-score model shown in Equation 6.5

Appendix D

Ethical clearance documents

Letter to principals

[Insert date]

Dear [Insert name],

The development of middle school children's' interest in statistical literacy

The purpose of this letter is to invite your school's participation in a research project that seeks to identify factors that contribute to students' interest in statistical literacy. This project forms the basis of PhD study conducted by Colin Carmichael, who is a registered teacher. It will involve a sample of your students in years 7 to 9 completing a short attitudinal questionnaire, of no more than 20 minutes, during their mathematics class.

Background: Interest is an emotion that is often present in self-motivated learning. It is the doing of something for its inherent value. Research suggests that students' interest in learning declines after they commence school reaching a minimum during adolescence. This study seeks to address the issue of interest in learning through the development of an instrument that can reliably assess the level of students' interest in a specific area of learning, viz. statistical literacy. Statistical literacy is the ability of a citizen to interpret messages that contain statistical elements, for example media claims that are based on survey data. Statistical literacy is acquired by students in many subjects, however most concepts are covered in the mathematic curriculum.

Benefit for the school: The researcher will be happy to demonstrate to interested staff resources that can be used to develop students' statistical literacy. The school will also be able to access their students' aggregate data as compared to other students who are involved in the project.

Confidentiality: All schools and students involved in the project are guaranteed confidentiality. Only the researchers will have access to the information collected. All information will be coded and no individual students or their schools will be named during the project or in any forthcoming reports. No identifying conversations or photographs will be used in any reports. The data will be secured and stored by the principal researcher for a period of 5 years. After this time they will be destroyed as confidential waste.

Freedom to refuse or withdraw: Participation of schools and students in all aspects of this project is entirely voluntary and evidenced by signing a consent form. A school or student can refuse to participate without any effects. Where a participant (school or student) elects to withdraw from the study, the data supplied to date will also be withdrawn. Parent/guardians and students are also free to withdraw their data at anytime. Parents of students in the classes of the school's participating teachers will receive information letters about the project, along with a "Consent for Participation" form, which they will be encouraged to complete and return. Only students who agree to participate and from whom parental consent forms have been obtained, will participate in the study. The letters will be provided by the researchers and distributed through the school, with researchers having no knowledge of parents' identities.

Concerns or complaints: This project has been approved by the Human Research Ethics Committee of the University of New England (Approval No. HE08/037, Valid to 31/03/2010). Should you have any complaints concerning the manner in which this research is conducted, please contact the Research Ethics Officer at the following address:

Research Services

University of New England

Armidale, NSW 2351.

Telephone: (02) 6773 3449

Fax: (02) 6773 3543

Email: Ethics@une.edu.au

Results of investigation: Students will not be given individual results during the project; however teachers and schools can be given feedback if requested. If you have any other questions about this research please don't hesitate to contact me on:

Colin Carmichael

2273 Gore Highway

Southbrook

Q.4363

Ph: (07) 4691 0558

Email: ccarmich@une.edu.au

Yours sincerely,

Colin Carmichael

STATEMENT OF INFORMED CONSENT FOR SCHOOLS

This form requests your permission for your school to take part in the research into the development of middle school students' interest in statistical literacy. The study is explained in the accompanying information letter.

Do you understand the nature of the research sufficiently well to make a free informed decision on behalf of your school? Yes or No

Are you satisfied that the circumstances in which the research is being conducted provide for the physical, emotional and psychological safety of your school, staff and students? Yes or No

I, ...(Print name), agree that:

1. I have read and understood the enclosed information sheet explaining the project and its purpose.
2. I understand that all identifiable information obtained will be treated as strictly confidential and that all research data will be securely stored on the University of New England premises for a period of 5 years, and will then be destroyed as confidential waste.
3. I agree that information collected during the study may be used in publications provided that involvement of the school, its teachers and students cannot be identified.
4. Any questions that I have asked have been answered to my satisfaction.
5. I agree to allow my school to participate in this study and understand that I may withdraw my school at any time without any consequences.

I give my permission for

....(Print school's name)

to take part in the research project.

Signed ....Date ....

Letter to parents

[Insert date]

Dear Parent/Guardian,

The development of middle school children's' interest in statistical literacy

I am writing to draw your attention to a research project with which your child's school is involved. ¡Name of School¿ has agreed to participate in the project that aims to measure students' interest in statistics. Students participating in the project will be required to undertake a short questionnaire (no more than 20 minutes). This survey will be conducted during a normal mathematics class and should not cause any distress or upset to your child. The study forms a part of a PhD research project that is conducted by Colin Carmichael, who is a registered teacher.

Background: It is often easier to learn something when you are interested. Unfortunately, many students report low levels of interest in learning. This study seeks to address this issue through the development of a test that can accurately measure students' interest. This test will then be used to evaluate the interestingness of learning materials. The study looks at students' interest in statistical literacy, which involves reading and understanding messages that contain statistical elements, such as graphs. In the current information age, it is essential that our students are able to understand such information and understanding is easier if they are interested.

Confidentiality: All schools and students involved in the project are guaranteed confidentiality and anonymity. Only the investigators will have access to the information collected. All information will be coded and no individual students or their schools will be named during the project or in any of the forthcoming reports. The data will be secured and stored by the researcher for a period of 5 years after which time it will be destroyed as confidential waste.

Freedom to refuse or withdraw: Participation of schools in all aspects of this project is entirely voluntary. Parent/guardians and students are also free to withdraw their data at any time during the study. This information letter includes a consent form. If you agree to your child participating in this study please complete and return the consent form to your child's school. If you do not agree to your child participating in this study, he/she will complete other activities, as directed by his/her teacher, during the survey period.

Concerns or complaints: This project has been approved by the Human Research Ethics Committee of the University of New England (Approval No. HE08/037, Valid to 31/03/2010). Should you have any complaints concerning the manner in which this research is conducted, please contact the Research Ethics Officer at the following address:

Research Services
University of New England
Armidale, NSW 2351.
Telephone: (02) 6773 3449
Fax: (02) 6773 3543
Email: `Ethics@une.edu.au`

Results of investigation: Students will not be given individual results during the project; however teachers and schools can be given feedback if requested. If you have any other questions about this research please don't hesitate to contact me on:

Colin Carmichael

2273 Gore Highway

Southbrook, Q.4363

Ph: (07) 4691 0558.

Email: ccarmich@une.edu.au

Yours sincerely,

Colin Carmichael

STATEMENT OF INFORMED CONSENT

This form requests your permission for your child to complete a short survey related to their interest in statistical literacy. If you agree to your child's participation, please complete and sign the form below.

Do you understand the nature of the research sufficiently well to make a free informed decision on behalf of your child? Yes or No

Are you satisfied that the circumstances in which the research is being conducted provide for the physical, emotional and psychological safety of your child? Yes or No

I, ...(Print name), agree that:

1. I have read and understood enclosed information sheet explaining the project and its purpose.
2. I understand that all identifiable information obtained will be treated as strictly confidential and that all research data will be securely stored on the University of New England premises for a period of 5 years, and will then be destroyed.
3. I agree that information collected during the study may be used in publications provided that involvement of the school and my child cannot be identified.
4. Any questions that I have asked have been answered to my satisfaction.
5. I agree to allow my child to participate in this study and understand that I may withdraw my child at any time without any consequences.

I give my permission for......................................(Print child's name) to participate in this research project.

Signed..Date......................

Statement by child:

I have discussed participation with my parent / guardian, understand what participation involves, and agree to participate in this study:

Name of child:........................

Signature of child...........................

Letter to parents of interviewees

Dear Parent/Guardian,

The development of middle school children's interest in statistical literacy

I would like your permission to conduct an interview with [Students' name].

This interview is an extension of the survey completed by students at [School name] for the Interest in Statistical Literacy project. This study is aligned with the StatSmart project in which your child's school is currently participating. Students will be asked to explain some of their responses and comment on their responses to the interest survey that they completed earlier this year.

This interview would involve your child, along with two or three others, withdrawing from their mathematics class to be interviewed for up to 40 minutes. The session will be audio taped for later transcription and analysis by the researcher. No identifying conversations of students will be used in reports.

In general students enjoy taking part in this type of interview, partly because they get individualised attention from someone interested in what they think. The interview protocol explores factors that influence students' interest in statistical literacy. Students are told at the beginning that they are free to withdraw at any time they wish, and that the results are not used for school assessment but are held confidential. The audiotapes and transcripts will be stored under secure conditions at the University of Tasmania.

This study has been approved by the Social Sciences Human Research Ethics Committee. If you have concerns or complaints about the conduct of this study you should contact the Executive Officer of the HREC (Tasmania) Network on (03) 6226 7479 or email:human.ethics@utas.edu.au. The Executive Officer is the person nominated to receive complaints from research participants. You will need to quote *HREC project number: H9151.*

Your child's participation in the interview is entirely voluntary, and refusing to participate will not have any adverse effect on your child's schooling. Please discuss participation with your child, and if you are happy for your child to participate, and your child agrees to do so, please each give your consent by signing and returning the form below. I would appreciate this very much.

`Colin Carmichael`

`Ph: (07) 4691 0558`

`Email: colin.carmichael@utas.edu.au`

Yours sincerely,
Colin Carmichael

STATEMENT OF INFORMED CONSENT

This form requests your permission for your child to complete a short survey related to their interest in statistical literacy. If you agree to your child's participation, please complete and sign the form below.

Do you understand the nature of the research sufficiently well to make a free informed decision on behalf of your child? Yes or No

Are you satisfied that the circumstances in which the research is being conducted provide for the physical, emotional and psychological safety of your child? Yes or No

I, ...(Print name), agree that:

1. I have read and understood enclosed information sheet explaining the project

and its purpose and that allowing my child to participate in an audio-taped 40-minute interview about his/her interest in statistical literacy and classroom experiences.

2. I understand that all identifiable information obtained will be treated as strictly confidential and that all research data will be securely stored on the University of Tasmania premises for a period of 3 years, and will then be destroyed.

3. I agree that information collected during the study may be used in publications provided that involvement of the school and my child cannot be identified.

4. Any questions that I have asked have been answered to my satisfaction.

5. I agree to allow my child to participate in this interview and understand that I may withdraw my child at any time without any consequences.

I give my permission for.......................................(Print child's name)
to participate in this research project.

Signed..Date......................

Statement by child: I have discussed participation with my parent / guardian, understand what participation involves, and agree to participate in this study:

Name of child:........................

Signature of child...........................