*Mathematics Education Research Journal* 2010, Vol. 22, No. 3, 9-39

# Measuring Middle School Students' Interest in Statistical Literacy

Colin Carmichael, Rosemary Callingham, Ian Hay, and Jane Watson
*University of Tasmania*

The following paper describes the development of an instrument designed to assess middle school students' interest in statistical literacy. The paper commences with a review of the literature as it relates to interest in this context and then proposes a theoretical model upon which the proposed instrument is based. The Rasch Rating Scale model is then applied to student responses to items in the instrument and fit statistics are analysed in order to assess the extent to which these responses conform to the requirements of the measurement model. The paper then presents evidence, including interview data, to support the validity of interpretations that can be made from the proposed instrument. The findings suggest that the proposed instrument provides a theoretically sound measure of middle school students' interest for statistical literacy that will be useful for the evaluation of interventions aimed at developing these students' statistical literacy.

Student engagement with learning tasks is considered higher when the learner is interested in that task. Indeed, Schiefele and Csikszentmihalyi (1995) reported that high levels of student interest were positively associated with academic achievement, deeper levels of cognitive processing, the use of self-regulatory learning strategies and students' ratings on the quality of their learning experiences. Interested students are also more likely to seek out opportunities that allow their engagement with the object of interest. As a result of this, student interest is known to be a predictor of subject choice, with McPhan, Morony, Pegg, Cooksey, and Lynch (2008) reporting that interest was one of five factors that explained students' choice of senior mathematics course in an Australian context, the others being their mathematics self-concept, their previous achievement in mathematics, and their perceptions regarding the usefulness and difficulty of mathematics. Such interest, however, is formed through the years preceding senior school, where interest towards mathematics appears to reach a minimum (Dotterer, McHale, & Crouter, 2009; Watt, 2004). Such low levels of interest for mathematics during adolescence are arguably contributing to a decline in the number of students studying higher levels of mathematics during their senior school years (Forgasz, 2006), which in turn is contributing to the reported shortage of skilled mathematicians (Australian Academy of Science, 2006). Addressing this declining participation rate involves strategies for sustaining and increasing students' interest for mathematics.

Extending the above argument, the recognised shortage of statisticians in Australia (Trewin, 2005) is assumed to be influenced by low levels of students' interest for statistics, particularly during early adolescence or their middle school years. In the Australian school context, statistical concepts have been introduced in the chance and data strand of the mathematics curriculum and presumably reinforced in other discipline areas such as science. For this reason students' interest for statistics is assumed to be influenced by their interest for mathematics. The assumed links between interest and achievement in statistics is, however, under-researched, in part because of a lack of appropriate instrumentation to measure statistics in a middle school context. Consequently, there is scope for the development of new instruments that can assess students' interest specifically for statistics (Carmichael, Callingham, Watson, & Hay, 2009). Previous attempts to use general mathematical interest scales to assess students' specific mathematical interest domains have been criticised with Ma and Kishor (1997) arguing that the general attitudinal instruments in mathematics provided at best only a crude approximation to the students' "true" attitudes to mathematics because of the content diversity associated with mathematics and the multidimensionality of students' mathematics development.

Thus, this paper reports the development of an instrument to provide a valid measure of middle school students' interest for the concepts underlying statistical literacy, where this literacy is defined as an ability to interpret and critically evaluate messages containing statistical elements (Gal, 2003). In an Australian context, most children should have encountered the requisite concepts for statistical literacy by the time they have completed their 11th year of school education at an age of approximately 16 years. The specific aims for this paper are:

1. To describe a theoretical model that can be used as the basis for better understanding middle school students' interest development;
2. To describe the development of the Statistical Literacy Interest Measure (SLIM), an instrument that can measure this interest; and,
3. To provide evidence to support the validity of interpretations based on SLIM.

# Theoretical Background

## *Interest*

The Macquarie Dictionary (Delbridge, Bernard, Blair, & Ramson, 1987, p. 910) defines interest as "the feeling of one whose attention or curiosity is particularly engaged by something." Interest is a positive affect that is directed towards some object, termed the interest object. Collections of such interest objects are often referred to as an individual's interests. Interest is regarded as having both trait and state characteristics (Schiefele, 1991). At the trait level individual interest is described as a "person's relatively enduring predisposition to reengage particular content over time" (Hidi & Renninger, 2006, p. 113). It is a close personal attachment to, or a valuing of, an interest object. Interest at the state level is more transitory but is associated with higher levels of emotion. This state can be induced by aspects of the environment and in such instances is termed situational interest. Such interest is very similar to the concept of flow (Csikszentmihalyi, 2002), a state where learners become so absorbed in the learning task that they lose all sense of time.

## *Operational Model of Interest in Statistical Literacy*

Based on motivation theory (Schunk, 1991), for many students in a middle school context, their interests in and their knowledge of statistical literacy are dynamic and interactive, that is, their content knowledge is influencing their interest, and their interest is influencing their content knowledge. Thus, because of this assumed interaction, the following discussion seeks to clarify and define a construct called statistical literacy interest. In regard to this, it is suggested in the following section that there are three main elements associated with students' interest and these are: importance interest, reflective interest, and curiosity interest. Along with the different types of interest 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 for a specific topic (Schiefele, Krapp, & Winteler, 1992). As such, students' responses to interest surveys typically reflect their valuing of the context or activity described in the survey items. This valuing typically relates to personal valuing that is influenced by the individual's past experience, current interests, knowledge, and goals, as well as 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.” It targets both the specific situations that students might encounter, such as “working out the probabilities for dice,” and also a student's desire to reengage in statistics, such as “getting a job involving statistics.” It is assumed that students who endorse the latter have those predispositions to reengage 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 reengagement.

It is also possible for students to anticipate and to reflect upon their interest towards or valuing of content knowledge that 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. It can be regarded as a form of epistemic curiosity (Litman, 2008). It is argued that students who would like to find out about the underlying concepts for 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, Boekaerts 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 (2000) 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 reengage.

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 about students' self-concept, where it is seen as 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 an 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, Baumert, & Schnabel, 2001; Tsai, Kunter, Ludtke, 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.

The two proposed content components relate to the actual subject matter and the contexts and activities associated with the learning of this subject matter. The topics associated with statistical literacy are identified by Watson (2006) as: sampling or data collection, graphs, averages, chance, beginning inference, and variation. The latter, however, is difficult to assess for, as Watson (2006, p. 219) acknowledges, "many curriculum documents do not even mention the word." Accordingly the topics used in this model are restricted to the first five of Watson's topics, reflecting an earlier classification by Holmes (1986). A review of the statistics education literature (Carmichael et al., 2009) suggests that the activities and data-contexts that students encounter will contribute to their interest. These include, for example, the use of technology (Mitchell, 1993), and sports-related data (Finzer, 2006).

In regard to this taxonomy grid model, shown in Table 1, the three elements of interest are: importance, curiosity, and reflective interest, and these can be constructed along the horizontal axis. On the vertical axis the two content components are the specific topics investigated within statistical literacy, and the activities and contexts associated with the learning of statistical literacy. This taxonomy became the theoretical starting point to generate a bank of items to populate the grid. These items became the initial item bank for the survey instrument.

Table 1
*Taxonomy Grid Model for Item Generation of Student Interest in Statistical Literacy Survey*

| Horizontal axis – interest elements: Reflective | Importance | Curiosity |
|---|---|---|
| Vertical axis – content components: | | |
| 1. Topics in statistical literacy: | | |
| Data collection. | | |
| Graphs. | | |
| Averages. | | |
| Chance. | | |
| Inference. | | |
| 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. | | |

## Methodology

The methodology used in this study involved a number of stages. Initially a bank of items was written to reflect the theoretical model outlined. These items were then reviewed by an expert panel and tested on a sample of middle school students. After some early changes to the items, subsequent quantitative testing of the items occurred. Follow-up interviews were also conducted with a number of participating students.

### *Creation of Item Bank*

A set of 34 self-description survey items was written in order to reflect the proposed theoretical model as outlined in Table 1. The different elements of interest were assessed through the use of different common stems. Reflective interest items used the common stem "I'm interested in," curiosity interest items used the common stem "I would like to know," and the importance interest items used the common stem "It's important to me personally that I can." A sample of these is shown in Table 2, which also details the element and context/activity that each item is thought to assess.

In addition to these self-descriptions a further set of six general interest items was developed. These items avoided specific contexts, topics, and activities, instead concentrating on students' interest in statistics in general. Examples included: "I am interested in learning more about statistics" and "It's important to me personally that I can used data to investigate questions that I might have."

Table 2
*Sample of Items reflecting the Operational Model*

| Interest | Topic | Context/activity | Item |
|---|---|---|---|
| Reflective | Data collection | Classroom | I'm interested in conducting surveys of other students at my school. |
| Importance | Averages | Media | It's important to me personally that I can understand news reports that use averages. |
| Curiosity | Graphs | Sports | I would like to know how a graph could be used to compare my sports team with other teams. |

## *Expert Review of Items*

The items were initially reviewed by a panel of academics with extensive experience in scale development and/or statistical literacy. The panel was asked to provide feedback regarding the appropriateness of items and also the layout and readability of the survey. Following this initial review, the revised items were then reviewed by a group of 45 practising teachers of middle school students. 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 rewritten to include the term *chance*, as it was felt that students were more familiar with the latter.

Based on the results of these reviews 30 of the original 40 items were deemed suitable for initial testing. Consequently these self-descriptions were compiled into a questionnaire that used a 5-point Likert scale, ranging from 1 (*Not me at all*) to 5 (*Describes me well*). All statements were expressed in a positive way as the practice of mixing negatively and positively worded statements is thought to reduce validity (Netemeyer, Bearden, & Sharma, 2003).

## *The Student Sample*

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 this study deemed to be middle school students were those who were in years 7, 8, or 9 of school, although some older and younger students were also enlisted. Year 7 students in the sample were either enrolled in a secondary school ($N$ = 153) and presumably taught by a mathematics specialist, or in dedicated middle school ($N$ = 97) and presumably taught by a generalist middle school teacher. Schools were invited to participate and targeted in order that as closely as possible the resulting sample would reflect the major demographic features of the population of Australian middle school students. This population, in turn, can be assumed to consist of approximately equal proportions of each gender, equal proportions of students in each of years 7, 8, and 9, and a range of school types including both Government and Independent schools. A total of 1384 students from 16 consenting schools across four Australian states were then invited to participate in the study. The results reported here are based on a total of 791 complete responses, a response rate of 57%. Of these students, 39% attended Government Schools, somewhat less than the population proportion estimated to be 61% (Australian Bureau of Statistics, 2008).

Data collection occurred in three stages over 12 months. The initial stage was undertaken using a sample of Queensland middle school students. The second stage occurred 6 months later and involved a sample of middle school students from schools participating in the "StatSmart Project" (Callingham & Watson, 2007). This project, in turn, is based in Tasmania, Victoria and South Australia, and seeks to examine the influence of teacher professional development on student outcomes in a statistical literacy context. The final stage involved students from both StatSmart and Non-StatSmart schools. A breakdown of schools and students in each stage is shown in Table 3.

Table 3
*Students' and Schools' Details for Each Stage of the Study*

| Stage | Students | | | Schools | |
|---|---|---|---|---|---|
| | Number | Mean age (yrs) | Males (%) | Government | Independent |
| Initial | 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 |

## *Quantitative Analysis of Items*

Quantitative analysis of the items employed the Rasch Rating Scale model (Andrich, 1978). Apart from assessing the unidimensional nature of the construct, this model allows an exploration of the hierarchical structure: Provided its assumptions are met, it creates an interval scale upon which both the interest level of students and the interestingness of items can be placed. The analysis in this study relates primarily to the extent to which students' responses to items conform to the requirements of the Rasch model. This, in turn, is assessed through the analysis of item fit-statistics.

Students' responses for all items in this study were analysed using the Rasch modelling program *Winsteps* (Linacre, 2006). Although this program produces a number of model fit statistics this study reports only the *infit* statistic ($u_i$), because it is less susceptible to the influence of outliers than other statistics (Bond & Fox, 2007). In line with a recommendation by Keeves and Alagumalai (1999), items with reported infit statistics between 0.77 and 1.3 are regarded as displaying satisfactory fit.

## *Initial Testing of Items*

During the initial stage, items were further reviewed on the basis of teacher feedback and/or quantitative analysis. As an example, 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. Since its reported infit statistic ($u_i = 1.52$) exceeded the upper limit, the item was written in a more general form as: "I'm interested in using averages to compare sports teams or players." As another example, feedback from participating teachers revealed that younger students 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 and replaced by an item that asked students for their level of interest in using data to investigate questions.

Initial testing of items also revealed a lack of spread in their relative difficulty, where the difficulty in this case is a scaled measure of the interestingness of the item based on students' responses to the Likert scale. There was an absence of items that could be endorsed by most students and also a lack of items that reflected apparent upper levels of interest. To rectify this situation, additional items with a general context were included. For example, student endorsement of the item "I'd like to know all about statistics" was thought to be indicative of higher levels of valuing towards statistics. As a result of this review, three items were modified and a further three were replaced; however 24 remained unchanged, which is ample for the linking of student responses across all stages of the study (Wright & Stone, 1999). The 30 items used as the basis for the quantitative analysis are shown in Appendix 1, which also shows the identifier code used for each item.

In order to obtain a measure of external validity, students in the initial stage also completed items from the Mathematics Interest Inventory (Stevens & Olivarez, 2005). Of the 27 items in the inventory, students in this study completed 10, which reportedly load onto one factor that assesses a positive attachment to mathematics. In addition to this measure of mathematics interest, all students in the study completed nine items comprising the Students' Self-Efficacy for Statistical Literacy Scale (Carmichael & Hay, 2009).

## *Final Analysis of Items*

Given that most items remained the same for the entire study student responses from all three stages were pooled for the quantitative analysis. This pooled sample was then randomly split into two similar samples, with details shown in Table 4.

Table 4
*Students' and Schools' Details for Random Samples*

| | Students | | | Schools | |
|---|---|---|---|---|---|
| | Number | Mean age (yrs) | Males (%) | StatSmart | Non-StatSmart |
| Sample 1 | 410 | 13.6 | 44.6 | 7 | 9 |
| Sample 2 | 381 | 13.5 | 47.2 | 7 | 9 |

The subsequent analysis, based on the responses of the 410 students in Sample 1, explores the degree of fit between the data and the model. An iterative approach is used, whereby items that display significant misfit are removed from the analysis and the model is reapplied to a smaller subset of items. Responses from the 381 students in the second sample are used to confirm the results of this analysis.

## *Follow-up Interviews*

A sample of 17 students was selected from those who had completed the interest survey. The students, from two participating schools, were selected in order to represent a range of levels of interest for statistical literacy. Interviews were semi-structured and were conducted in groups of between 2 and 4 students; details of the groups are provided in Table 5.

Students were asked a number of questions, but of interest to this study are the following:

1. What are some of the things you do in maths when you learn about statistics?
2. Which of these are of interest to you?
3. What are some of the things you do in other classes when you learn about statistics?
4. Which of these are of interest to you?

Students were also asked to explain why they responded as they did to specific items in the interest inventory, the purpose of these data being to provide further evidence for the validity of the proposed instrument.

Table 5
*Details of Students and Schools used in Qualitative Study*

| Group | Students | School |
|---|---|---|
| 1 | 3 boys and 1 girl from a mixed ability Year 7 class | Independent, co-educational from Qld. |
| 2 | 2 girls from a mixed ability Year 8 class | |
| 3 | 2 girls and 1 boy from a high ability Year 9 class | |
| 4 | 2 boys and 2 girls from a low ability Year 8 class | Government, co-educational from Tasmania |
| 5 | 2 boys and 2 girls from a high ability Year 10 class | |

Interviews took between 30 and 40 minutes. They were recorded and subsequently transcribed. A content analysis of the data (Miles & Huberman, 1984) was then performed and some of the results are reported in this paper.

## Results

### *Item Fit*

Using the iterative approach described above, 16 of the 30 items were found to display satisfactory fit. Collectively these 16 items, which are shown emboldened in Appendix 1, produce a measure of interest that explains 67% of the variance in student responses and reports an internal reliability of $\alpha = 0.91$. These items are ordered by difficulty in Table 6, which reports their estimated difficulties ($\delta_1$), the interest element they are thought to assess, their standard errors ($SE[\delta_1]$), and their infit statistics ($u_i$). The table also displays the same statistics, but obtained from the second sample of students. In addition to this the table reports the difference in difficulty estimates between the two samples ($\delta_1 - \delta_2$), the standard error of this difference ($SE[\Delta\delta]$), and the corresponding *t*-statistic for this difference.

Table 6 *Item Statistics for SLIM Obtained from both Samples of Students*

| Item | Interest element | Sample 1 | | | Sample 2 | | | Differences | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\delta_1$ | $SE(\delta_1)$ | $u_i$ | $\delta_2$ | $SE(\delta_2)$ | $u_i$ | $\delta_1 - \delta_2$ | $SE(\Delta\delta)$ | $t$ |
| 15 | Reflective | 0.80 | 0.06 | 1.16 | 0.70 | 0.06 | 1.12 | 0.10 | 0.08 | 1.18 |
| 38 | Curiosity | 0.55 | 0.06 | 1.06 | 0.52 | 0.06 | 1.04 | 0.03 | 0.08 | 0.35 |
| 14 | Reflective | 0.42 | 0.05 | 0.85 | 0.43 | 0.06 | 0.90 | -0.01 | 0.08 | -0.13 |
| 19 | Curiosity | 0.42 | 0.05 | 0.97 | 0.45 | 0.06 | 1.00 | -0.03 | 0.08 | -0.38 |
| 3 | Reflective | 0.36 | 0.05 | 1.06 | 0.42 | 0.06 | 1.01 | -0.06 | 0.08 | -0.77 |
| 17 | Curiosity | 0.14 | 0.05 | 1.14 | 0.04 | 0.05 | 1.10 | 0.10 | 0.07 | 1.41 |
| 20 | Curiosity | -0.01 | 0.05 | 1.07 | -0.10 | 0.05 | 1.17 | 0.09 | 0.07 | 1.27 |
| 16 | Curiosity | -0.04 | 0.05 | 1.18 | 0.04 | 0.05 | 1.21 | -0.08 | 0.07 | -1.13 |
| 25 | Importance | -0.06 | 0.05 | 0.77 | -0.07 | 0.05 | 0.80 | 0.01 | 0.07 | 0.14 |
| 23 | Importance | -0.08 | 0.05 | 0.88 | -0.03 | 0.05 | 0.90 | -0.05 | 0.07 | -0.71 |
| 24 | Importance | -0.21 | 0.05 | 1.17 | -0.30 | 0.05 | 1.14 | 0.09 | 0.07 | 1.27 |
| 26 | Importance | -0.30 | 0.05 | 1.01 | -0.36 | 0.05 | 1.11 | 0.06 | 0.07 | 0.85 |
| 30 | Importance | -0.38 | 0.05 | 0.89 | -0.33 | 0.06 | 0.96 | -0.05 | 0.08 | -0.64 |
| 28 | Importance | -0.50 | 0.05 | 0.94 | -0.42 | 0.05 | 0.78 | -0.08 | 0.07 | -1.13 |
| 29 | Importance | -0.54 | 0.05 | 1.00 | -0.55 | 0.06 | 0.94 | 0.01 | 0.08 | 0.13 |
| 27 | Importance | -0.56 | 0.05 | 0.97 | -0.46 | 0.06 | 0.89 | -0.10 | 0.08 | -1.28 |

An inspection of the item difficulty hierarchy shown in Table 6 indicates that students find it most difficult to endorse items that assess a desire to reengage with statistics, such as getting a job involving statistics (item 15) and learning more about statistics (item 14). Wanting to know all about statistics (item 38), although classified as a curiosity item, can also be regarded as a desire to reengage. Students find it easier to endorse an interest in specific situations, such as working on problems involving data and statistics (item 3). They find it even easier to endorse a desire to find out about statistical literacy in specific situations, with items assessing curiosity interest (16, 17, 19, and 20) being less difficult than items assessing reengagement, but more difficult than the importance-interest items. At the bottom of the hierarchy, students find it easiest to endorse the importance of mastering simple tasks such as using the correct graph (item 27), arranging data into tables (item 29), and understanding graphs in the media (item 28). These all assess, to an extent, the importance of task mastery. Students find it less easy to endorse the importance of statistical literacy in wider contexts; such as believing scientific claims based on data (item 26), knowing how to calculate the risk of injury (item 24), and understanding news reports that use averages (item 23). With the exception of items 19 and 25, the statistical literacy interest hierarchy commences at its lowest level with the importance of task-mastery, then, the importance of statistical literacy in a wider context, the desire to find out about statistical literacy, an interest in the situation, and, at its highest level, a general desire to reengage in statistical literacy.

The results obtained from students in Sample 2 indicate that the measure appears to be invariant across samples. Within the bounds of measurement error, the hierarchy obtained from Sample 1 is equivalent to that obtained from Sample 2. Reported fit statistics are also satisfactory across both samples.

## *Dimensionality*

A basic assumption of the Rasch model is that the underlying trait is unidimensional. In order to assess this, it is recommended that a principal component analysis (PCA) of the residuals be examined for apparent structure (Linacre, 1998). This was done for responses from Sample 1 and the results are displayed in Figure 1, which positions each item on the plot by its difficulty and the magnitude of its loading on the principal component of its residuals. This plot indicates some structure in the residuals with most of the importance items (23 to 30) grouped together in the one quadrant.

Given this apparent structure in the residuals, it was decided to test the data for evidence of multidimensionality. An exploratory factor analysis suggested the presence of three factors aligning with the three elements of interest. Significant factor loadings for this analysis are reported in

Appendix 2. As is discussed, however, previous research suggests that these elements should contribute to the one dimension. In order to test this assumption, a multidimensional Rasch model (Adams, Wilson, & Wu, 1997) was applied to the 16 items using the software package *Conquest* (Wu, Adams, Wilson, & Haldane, 1998). More specifically, items 3, 4, 15, and 38 were assigned to a reflective interest dimension; items 16, 17, 19, and 20 were assigned to a curiosity interest dimension; and the remaining items were assigned to an importance interest dimension. In comparison to a single dimensional 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 at the 1% level. The same procedure was repeated on Sample 2 responses and similar results were obtained. Thus the evidence suggests the presence of three dimensions, although these are highly correlated with all correlations exceeding .75. This apparent multidimensionality, however, 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.

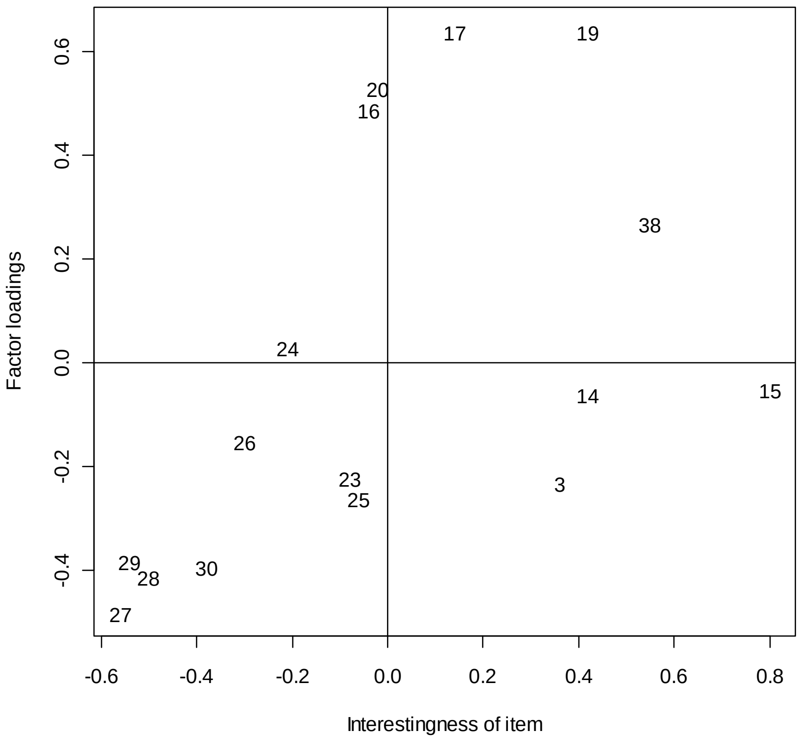


*Figure 1.* Plot of PCA of residuals against difficulty for each item in SLIM.

## *Relationship with Other Measures*

As reported, students in the initial stage of the study completed a mathematics interest inventory. These results were analysed using the Rasch Rating Scale model and were found to create a measure of mathematics interest that explains 83% of the variance and reports an internal reliability of $\alpha = .94$. Based on their responses to this inventory and SLIM, students were assigned mathematics and statistical literacy interest scores respectively. On the basis of these scores, there is evidence that students' interest for mathematics and their interest for statistical literacy are moderately and positively associated $(r = .571, p = .000)$.

Students' responses to the self-efficacy scale were also analysed using the Rasch Rating Scale model and were found to create a measure that explains 70% of the variance in student responses and reports an internal reliability of $\alpha = .93$. All students were thus assigned a self-efficacy score and a statistical literacy interest score. On the basis of these scores, there is evidence that students' self-efficacy for statistical literacy and their interest for statistical literacy are moderately and positively associated $(r = .625, p = .000)$.

## *Differential Item Functioning (DIF)*

It is important that items do not function differentially for different subgroups undertaking the instrument. In this instance items were tested for evidence of broad differences in responses by gender. This test is routinely done by Winsteps, and involves the estimation of item difficulties separately for each gender and then the comparison of these estimates. In this instance the test was applied to student responses from Sample 1 and after applying a Bonferroni adjustment, one item displayed evidence of DIF at the 5% level of significance. There was evidence that boys tend to find more interest in working on problems (item 3) than girls. The same item was also identified when the process was applied to student responses from Sample 2.

## *Follow-up Interviews*

The data presented in this section achieve two aims. In the first instance they are used to "paint" pictures of students who are typically above average, average, and below average in relation to their interest for statistical literacy. Secondly, the data are used to highlight different motivations for students' responses to the items in the interest inventory.

Three students in the sample, with levels of interest greater than 1 logit, were considered to have above average interest. These students tended to acknowledge a positive emotional aspect to their learning of statistics. For example, a Year 10 male student with an interest of 2.4 logits acknowledged that *CensusAtSchool* – an opportunity for random sampling of other Australian students provided by the Australian Bureau of Statistics – was more "fun … because it's different people and so much information." Similarly, a Year 7 student with an interest of 1.39 logits argued that doing online surveys "… was just a fun thing for me to do." Eleven students, with interest levels between -1 and 1 logits, were considered to have average interest. These students were more likely to provide ambivalent responses during the interview. For example, a Year 8 student with an interest of 0.18 logits felt that "…sometimes its interesting finding out the facts." Similarly, another student, with an interest of -0.32 logits, regarded a statistical software package he had used as "…the least boring one." Those students with below average interest, in this case levels of interest less than -1 logits, tended to be influenced more by their self-competency beliefs about mathematics in general. When asked what experiences she had found most interesting when learning statistics, one girl with an interest of -1.8 logits answered "… probably when we had enough free time to play Red-river," a game that was unrelated to the learning of statistics. Another Year 8 girl with an interest of -2.1 logits had negative experiences with mathematics in general, stating "I used to be good at everything. Now I'm not, except for sport." These students also preferred the statistics that they encountered in other subjects. This appeared to be more due to a preference for the other subjects than the statistics that they encountered, with one girl claiming in other subjects "… you get up and do stuff."

The reasons students provided for their different responses were varied. In general, their responses to the importance interest items were governed by seemingly extrinsic motivations. For example, in relation to the importance of using the correct graph (item 29), one Year 10 girl remarked "I didn't want to, like, stuff something up." Similarly, a Year 8 girl on why she had answered the importance items as she had, said "… probably because it would help me in the future." Students' responses to the reflective interest items, however, were more influenced by their perceived identity and goal aspirations. A Year 9 student found doing online surveys (item 1) to be of interest because "… it helps you figure out different things about yourself." A Year 10 student was interested in using statistics to prove a point (item 11) because "… it's a sort of, ahm, sort of powerful skill."

Students' responses appeared to be governed by their ages and therefore their exposure to statistical concepts. Older students tended to give responses that reflect the true nature of statistics. For example one Year 10 boy did not like the inherent uncertainty associated with statistical investigations claiming that in such cases "…you're not really solving a problem." The younger students tended to be more influenced by their mathematics experiences. When asked why he had responded positively to an interest in working on problems involving data and statistics (item 3), one Year 7 boy replied "…I just like solving stuff like that" and another added "I just like maths."

## *Middle School Students' Interest in Statistical Literacy*

One of the benefits of using the Rasch measurement model is that it is possible to locate both the interest level of students and the interestingness of items on the one scale. Figure 2 shows this information for students in Sample 1. The first column of the figure shows the logit scale, while the second column shows the interest level of students which, as can be seen, ranges from approximately -4 logits up to 2.6 logits. The third column of the figure shows the item thresholds, where there are four thresholds for each item, one less than the number of Likert categories. The threshold denoted 25.4, for example, is the point on the scale where a student with that level of interest is equally likely to respond with a 3 or 4 to item 25. This figure also shows in the first column the position on the logit scale of the mean interest level and then separations based on the sample standard deviation.

An inspection of Figure 2 indicates that several students have interest levels that are more than 2 standard deviations below the mean: They do not have measurable levels of valuing for statistical literacy. On the 5-point Likert scale, such students typically provide responses of 1 (*not me at all*) to all interest self-descriptions. At an interest level of approximately -2 logits, students are able to see some importance in understanding and completing basic tasks associated with statistical literacy. On the 5-point scale, such students are likely to respond with a 2 to items 27 and 29. At this level there is some measurable, albeit low, level of valuing for statistical literacy. The interview data suggest the valuing of such students is strongly influenced by their experiences of mathematics, which may lead them to disengage with the learning of statistics.

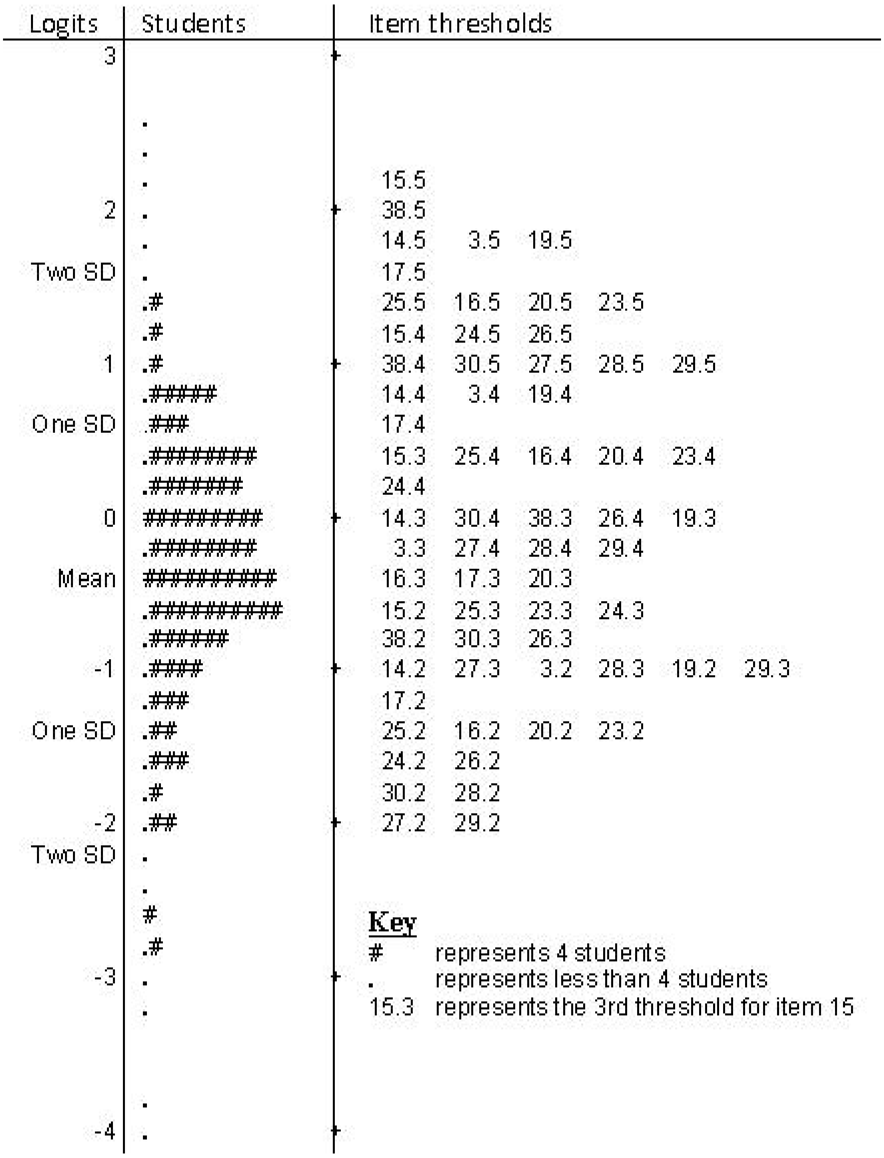


*Figure 2*. Item person map for SLIM.

Students with interest levels near the mean can readily see the importance of understanding the basic tasks associated with statistical literacy. For example, students with an interest level greater than -0.2 logits are likely to respond with a 4 on the 5-point scale, to items 27, 28, and 29.

These students can also see some importance in using statistical literacy in wider contexts and have some desire to find out about statistical literacy. For example, they are likely to respond with a 3 to items 16, 17, and 20. They do not, however, want to reengage with statistical literacy and are likely to respond with a 2 to item 15. The interview data suggest that such students tend to be ambivalent towards the learning of statistics, finding it of interest sometimes.

At the top end of the scale, students with interest levels that are greater than one standard deviation above the mean can see the importance of statistical literacy and have a desire to find out about statistical literacy in a variety of contexts. On the 5-point scale, they typically respond with a 4 to items such as 16, 17, and 20. These students also express some desire to reengage with statistical literacy and are likely to respond with a 3 to item 15. A small number of students have interest levels that exceed two standard deviations above the mean. Such students strongly endorse the importance of statistical literacy and have a desire to reengage with the domain. On the 5-point scale, these students are likely to respond with a 5 (*describes me well*) to most self-descriptions, although perhaps a 4 on items that assess reengagement. The interview data confirm that highly interested students tend to acknowledge a positive emotional attachment to statistics and experience "fun" in some of their learning (Schiefele, 1991).

# Discussion

## *Evidence for the Validity of Interpretations based on SLIM*

There are six aspects, or forms of evidence, to a validity argument: content, substantive, structural, generalisability, external, and consequential (Messick, 1995). The following discussion addresses each of these aspects in relation to SLIM, the measure reported in this paper.

Arguments that relate to the relevance, representativeness, and technical quality of the items all contribute to content evidence. The expert review of items, and their subsequent refinement, contributed to their relevance. The self-descriptions, shown in Appendix 1, represent the five identified topics underlying statistical literacy, for example: data collection (item 20), graphs (item 27), averages (item 23), chance (item 24), and inference (item 17). They also include self-descriptions that assess interest in the interpretation of media reports (items 20, 23, and 28), fundamental to statistical literacy, as well as interest in the learning of statistics (items 3, 25, 27, 29, and 30). Of concern is the lack of specific items near the top of the hierarchy, where only general self-descriptions were able to fit the requirements of the measurement model. As is discussed later in this paper, this may be more due to the nature of the construct, rather than deficiencies in the instrument itself. There is also a cluster of items near the bottom of the hierarchy that assess the importance of mastering simple tasks (items 27, 28, 29, and 30). This suggests some redundancy and the possibility that some of these items could be removed without adversely affecting the quality of the measure. In a Rasch measurement paradigm, evidence to support the technical quality of items is provided in the reported fit statistics (Smith, 2001). The fit statistics reported in Table 6 are all within an acceptable range, thus demonstrating that most students responded to the items in a similar way.

Substantive evidence refers to the extent to which underlying theories predict the observed outcomes. The focus, in this instance, is the agreement between the observed and expected hierarchy of item difficulties. It was expected that students would find it easier to endorse items assessing importance than those assessing reflective interest. With the exception of item 25, items assessing the importance of understanding basic concepts, such as graphs (item 28), are the easiest. Such items are likely to assess students' valuing of task-mastery, and accordingly represent low levels of the valuing associated with interest. Items that assess the importance of using statistical literacy in wider contexts are more difficult. As is expected, items that are the most difficult to endorse are those assessing a desire to reengage in statistics, such as getting a job that involves statistics (item 15). As expected, items assessing an interest in the situation, such as working on problems (item 3), represent lower levels of individual interest than those assessing reengagement. With the exception of item 19, which has a political context, the items assessing epistemic curiosity in specific situations lie between those assessing importance and those assessing reflective interest, on the interest hierarchy.

In regards to structural evidence, it appears that the underlying interest construct may consist of three highly correlated dimensions that align with each of the three elements of interest used in this study. Further testing needs to establish whether this is a statistical artefact that has occurred through the use of different common item stems. 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.

The evidence presented in Table 6 supports the generalisability of SLIM. The items in SLIM appear to be invariant across two randomly selected samples, suggesting the measure can be validly used to assess interest in other samples of middle school students.

The reported positive associations between SLIM and other measures provide external evidence for its validity. Given that Australian middle school students learn their statistical literacy primarily in their mathematics classes, it is expected that their interest for statistical literacy should be positively associated with their interest for mathematics in general. As is reported in Carmichael et al. (2009), students' competency beliefs are known to be positively associated with their interest. It is therefore expected that students' self-efficacy for statistical literacy will be positively associated with their interest.

Consequential evidence concerns the future impact on students who may use the instrument. Given that SLIM will likely be used to evaluate interventions, it is important that items do not differentiate between subgroups of students (Smith, 2001). As is reported, one item displays evidence of differential item functioning by gender. Boys tend to find more interest in doing problems involving statistics (item 3) than girls. Given that only one of the 16 items was problematic, it is arguably of minor consequence.

## *The Nature of the Construct*

Many of the items assessing reflective interest elicited student responses that were quite erratic and these items were discarded if they did not meet the requirements of the measurement model. Although the interview data suggest that students' responses to these items were closely associated with the self, these results may also reflect the influence of context. For example, many students who were otherwise disinterested in statistical literacy provided positive responses to items with a sports context, suggesting context had a major influence on their response. Such findings are supported in other domains, with Haussler (1987) reporting that students' interest in the context associated with physics self-descriptions can explain up to 60% of the variation in their overall responses. Yet context is essential to statistical literacy. Arguably general self-descriptions that lack context assess more of an interest in doing general mathematical computations and problems, than an interest in the learning of statistics. This is an inherent problem with measuring interest in the current context. This problem can be overcome through extending the measure to include curiosity and importance, as was done in this study. Although most of the reflective items in the interest inventory do not meet the requirements of the measurement model, most of the importance items do, as do several of the curiosity items. Students' responses to these items, although representing lower levels of the valuing that is associated with interest, were more consistent.

The findings of this study support the research of Ma and Kishor (1997) who argued that the general attitudinal instruments in mathematics provide at best only a crude approximation to students' "true" attitudes to mathematics because of the content diversity associated with mathematics and the multidimensionality of students' mathematics development.

Furthermore the research supports the notion that students' interest in a content area can be seen as both multidimensional and having an interlinking hierarchy, in that the different strands come together to form a general or overall construct. While Hattie's (2009) "rope" model analogy has it limitations, the researchers in this study have investigated students' interest in statistical literacy at both the individual strand item level and the interlinked factor level of the construct.

## Limitations of the Study and Recommendations for Future Research

A limitation of this study is the lack of randomness in sample selection and the relatively small sample size. This could influence the generalisability of results and also produce a certain bias, in that students with an above average interest for statistics may be more willing to undertake such a statistics survey than those with a below average interest. Future research, using a larger stratified sample, is required to explore the presence or otherwise of such bias.

The interview data suggest the possibility that the responses of younger students, with arguably fewer experiences in statistics, and students with low competency beliefs in mathematics, may be more guided by their general mathematics interest than their interest for statistical literacy. While based on only a small sample of students, these validity issues do suggest the need for further investigation of the instrument.

In regards to the apparent multidimensionality, Linacre (1998, p.1) cautioned that "empirical data are always manifestations of more than one latent dimension." It is recommended that the instrument is further tested in order to ascertain whether the apparent multidimensionality has an adverse consequence on its use. In addition to this it is recommended that further item development occur, in particular the inclusion of more items that try to tap an individual's level of personal valuing that is associated with reflective interest. How SLIM relates to other psychosocial variables, such as goal orientation, is also relevant to research investigating the development of statistical literacy. Interest in learning is known to be positively associated with the adoption of mastery learning goals (Wigfield & Cambria, 2010). In addition to this, Hyde and Durik (2005) have suggested that there are interaction effects between gender and students' style of goal orientation in that the motivational benefits of adopting performance goals appear to be stronger for boys than for girls.

## Conclusion

Recent developments in Australia suggest that statistical literacy will play a more prominent role in the school curriculum than it has in the past. In the proposed Australian National Curriculum, probability and statistics is one of only three content strands within the mathematics syllabus. Yet statistical literacy must be attained in a wide variety of contexts, and like numeracy, it too requires "… an across curriculum commitment" (Council of Australian Governments, 2008, p. xii). Given this climate of change in Australia, the research presented in this paper is timely. A growing emphasis on statistical literacy implies a growing need for research on how students come to learn and value such literacy.

This paper has reported the development of SLIM, a measure of middle school students' interest in statistical literacy. More specifically the paper has proposed a taxonomy grid model to explain the nature of students' interest, described the development of SLIM, and then presented evidence that interpretations made from the use of the instrument will be valid. The Rasch analysis of student responses has confirmed the hierarchical nature of the statistical literacy interest construct in that most middle school students appear to value statistical literacy as important, but fewer value it to the extent that they wish to reengage. The research presented in this paper suggests that SLIM has a sound theoretical basis and displays satisfactory psychometric properties. The instrument should be useful for researchers seeking to explore the influence of students' interest in the emerging yet distinct domain of statistical literacy.

## Acknowledgements

The research was funded by Australian Research Council grant number LP0669106, with support from the Australian Bureau of Statistics.

## References

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

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

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

Australian Academy of Science. (2006). Mathematics and statistics: Critical skills for Australia's future. *The National Strategic Review of Mathematical Sciences Research in Australia.* Retrieved April 22, 2008 from: http://www.review.ms.unimelb.edu.au/Report.html.

Australian Bureau of Statistics. (2008). *4221.0 Schools Australia*. Retrieved April 22, 2008 from: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4221.02008?OpenDocument

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

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Callingham, R., & Watson, J. M. (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*. Fremantle: ACER. Available at http://www.aare.edu.au/07pap/cal07042.pdf

Carmichael, C.S., Callingham, R., Watson, J.M., & Hay, I. (2009). Factors influencing the development of middle school students' interest in statistical literacy. *Statistics Education Research Journal, 8*(1), 62-81.

Carmichael, C.S., & Hay, I. (2009). The development and validation of the Students' Self-Efficacy for Statistical Literacy Scale. In R. Hunter, B. Bicknell & T. Burgess (Eds.), *Proceedings of the 32nd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 97-104). Wellington: MERGA Inc.

Council of Australian Governments. (2008). *National numeracy review report.* Department of Education, Employment and Workplace Relations. Retrieved September 1, 2008 from: http://www.coag.gov.au/reports/docs/national_review.pdf

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

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

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

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

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*: IASE. Available at www.stat.auckland.ac.nz/~iase

Forgasz, H. (2006). Australian year 12 "Intermediate" level mathematics enrolments 2000-2004: Trends and patterns. In P. Grootenboer, R. Zevenbergen & M. Chinnappan (Eds.), *Proceedings of the 29th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 211-220). Canberra: MERGA Inc.

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

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

Haussler, 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.

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

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). Victoria (BC): IASE.

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.

Kashdan, T. B., & Silvia, P. J. (2009). Curiosity and interest: The benefits of thriving on novelty and challenge. In C. R. Snyder & S. J. Lopez (Eds.), *Oxford handbook of positive psychology* (2nd ed., pp. 367-374). New York: Oxford University Press.

Keeves, J. P., & Alagumalai, S. (1999). 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.

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.

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

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

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

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

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

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

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Newbury Park: SAGE Publications.

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

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

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

Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*(3 & 4), 299-323.

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). New Jersey: Lawrence Erlbaum Associates

Schunk, D.H. (1991). Self-efficacy and achievement motivation. *Educational Psychologist, 26,* 207-231.

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

Stevens, T., & Olivarez, A. (2005). Development and evaluation of the mathematics interest inventory. *Measurement and Evaluation in Counseling and Development, 38*(3), 141-152.

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

Trewin, D. (2005). Improving statistical literacy: The respective roles of schools and the National Statistical Offices. In M. Coupland, J. Anderson & T. Spencer (Eds.), *Twentieth biennial conference of the Australian Association of Mathematics Teachers* (pp. 11-19). Adelaide: AAMT.

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

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

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.

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.

Wright, B., & Stone, M. (1999). *Measurement essentials* (2nd ed.). Wilmington, Delaware: Wide Range Inc.

Wu, M. J., & Adams, R. J. (2006). Modelling mathematics problem solving item responses using a multidimensional IRT model. *Mathematics Education Research Journal, 18*(2), 93-13.

Wu, M. J., Adams, R. J., Wilson, M. R., & Haldane, S. (1998). Conquest [Computer Software]. Melbourne: ACER.

---

## *Authors*

Colin Carmichael, Faculty of Education, University of Southern Queensland, West St. Toowoomba, QLD, 4350. <Colin.Carmichael@usq.edu.au>. The research reported in this paper was carried out while the author was at the University of Tasmania.

Rosemary Callingham, School of Education, University of Tasmania, Locked Bag 1307, Launceston, TAS, 7250. <rosemary.callingham@utas.edu.au>

Ian Hay, School of Education, University of Tasmania, Locked Bag 1308, Launceston, TAS, 7250. <ian.hay@utas.edu.au>

Jane Watson, School of Education, University of Tasmania, Private Bag 66, Hobart, TAS, 7001. <jane.watson@utas.edu.au>

# Appendix 1: Items in the interest inventory[1]

I'm interested in:

1. Doing magazine or online surveys.
2. Surveys that find out how people feel about things.
3. **Working on problems involving data & statistics.**
4. Looking up unusual statistics.
6. Using averages to compare sports teams or players.
7. The average rainfall for my home area.
9. Reading graphs in newspaper and magazine reports.
10. Conducting surveys of other students at my school.
11. Working out the probabilities (or chances) for dice, coins and spinners.
12. Using computer programs to help me investigate problems involving data.
13. Using statistics to prove a point or win an argument.
14. **Learning more about statistics.**
15. **Getting a job that involves statistics.**

I'd like to know:

16. **How scientists calculate the chance of rain.**
17. **How a survey can be used to predict who will win the next election.**
19. **How politicians make decisions that are based on data.**
20. **Whether a survey reported on the radio or TV about students was correct.**
21. Whether a game I was playing that used dice or spinners was fair.
22. How a graph could be used to compare my sports team with other teams.
38. **All there is to know about statistics.**

It's important to me personally that I:

23. **Can understand news reports that use averages.**
24. **Know how to calculate the chance of being injured from risky behavior.**

---

[1] Bold items are those used for the interest measure SLIM.

**25. Understand the words used in statistics.**

**26. Can believe scientific claims that are based on data.**

**27. Use the correct graph when displaying my data.**

**28. Can understand graphs that appear on the internet or in newspapers.**

**29. Can arrange data into tables.**

**30. Can use data to investigate questions that I might have.**

Other descriptions:

31. I get so involved when I work with data that I sometimes lose all sense of time.

36. I like to work on statistics problems in my spare time

## Appendix 2: SLIM item loadings on three components after rotation

| Item | Component 1 | Component 2 | Component 3 |
|---|---|---|---|
| 3 | | | 0.74 |
| 14 | | | 0.75 |
| 15 | | | 0.71 |
| 38 | | 0.49 | 0.57 |
| 16 | | 0.70 | |
| 17 | | 0.78 | |
| 19 | | 0.81 | |
| 20 | | 0.73 | |
| 23 | 0.63 | | |
| 24 | 0.53 | | |
| 25 | 0.61 | | 0.35 |
| 26 | 0.60 | | |
| 27 | 0.73 | | |
| 28 | 0.74 | | |
| 29 | 0.74 | | |
| 30 | 0.77 | | |

Notes: Only factor loadings exceeding 0.3 are reported.
Solution follows a varimax rotation.