The development and validation of the Statistical Literacy Interest Measure (SLIM)

Colin Carmichael (School of Education)

Abstract

This paper describes the development of an instrument designed to assess the interest that middle school students (years 7 to 9) have towards statistical literacy. In particular the paper presents a theoretical framework for interest in statistical literacy which is subsequently used as the basis for item development. Initial pilot results that are analysed using a Rasch measurement perspective are then presented.

A statistically literate adult should be able to interpret and critically evaluate messages that contain statistical elements (Gal 2003:80). For example, they should be able to recognize bias as a possible source of error in media reports of survey data. Watson (2006) proposed a model to describe the development of statistical literacy in primary and middle school students. In this model she identified task motivation as one component that contributes to statistical literacy. Within this category, Watson argued that statistically literate students should acquire certain dispositions that include: skepticism, imagination, curiosity and awareness. It is argued that positive affect, and in particular interest, is a major source of this task motivation and that the necessary dispositions identified by Watson will be acquired by students as they experience positive affect in their interactions with data.

This paper reports the development of SLIM, an instrument designed to assess levels of student interest in statistical literacy. The paper commences with a review of the theoretical background relating to interest and its influence on learning. It then reports on a model of interest that is used as the basis for item construction. Finally this paper reports some preliminary findings on the reliability and validity of SLIM.

Theoretical background

The Macquarie Dictionary defines interest as ‘the feeling of one whose attention or curiosity is particularly engaged by something’. Interest is a positive affect that is specifically directed towards some object, termed the *object of interest*. It is regarded as having both trait and state characteristics (Schiefele 1991:302). At the trait level *individual interest* is described as a ‘person’s relatively enduring predisposition to reengage particular content over time’ (Hidi & Renninger 2006:113). Interest at the state level is more transitory and is typified by a positive emotion akin to excitement. 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 and in such instances is termed *actualized interest*. It is believed that individual interest can emerge from situational interest (Hidi & Renninger 2006; Mitchell 1997). Thus the requisite dispositions for statistical literacy may emerge from students experiencing the state of interest during their learning.

In a learning context, a student’s interest can explain some of their motivation to engage in learning activities (Deci 1992). In fact students, who are motivated out of an interest in the subject, are known to produce qualitatively superior learning outcomes to their uninterested peers. For example, Schiefele and Csikszentmihalyi (1995) reported that levels of student interest were positively associated with deeper levels of cognitive processing, the use of self-regulatory learning strategies and students’ ratings on the quality of their learning experience. Further, there is a significant correlation between levels of student interest and their academic achievement (Schiefele, Krapp & Winteler 1992).

A model of interest

The interest assessed in a school context using self-report questionnaires will be a mix of individual and situational interest. Consequently at an operational level interest is thought to have both a value and emotion dimension (Schiefele 1991:302). The value dimension, which is assessed through items that ask students to reflect on the importance to them personally of the interest object, reflects a measure of enduring or individual interest. The emotion dimension, which is assessed through items that ask students to reflect on the interest that the object evokes, reflects a measure of the more transient situational interest. A third dimension is included in the current model, which assesses students' interest in finding out about information that for many may be unknown. This dimension is appropriate in a middle school context where students' academic interests are emerging and like all interest, is associated with specific interest objects. It is a very specific form of epistemic curiosity (Litman & Spielberger 2003) that should reflect students' current levels of interest.

It is also believed that in a school situation, students' self-reported interest will have two components (Hoffman 2002). The first is their interest in the specific subject matter and the second their interest in the learning of this subject matter. Thus it is proposed that a model of interest in statistical literacy, shown in Table 1, will have three dimensions and two components.

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. This subject matter may for convenience be presented in topics which are identified by Watson (2006) as: sampling or data collection; graphs; average; chance; beginning inference; and, variation. A student's interest in the learning of statistical literacy, however, will also be influenced by the contexts in which the material is presented and the activities that they encounter.

Context plays a particularly key role in the development of statistical literacy. Watson and Callingham (2003) argued that as students progress through the statistical literacy hierarchy they become more adept at applying their statistical knowledge to wider contexts until finally they are able to critically interact with these contexts. It is argued that it is in these wider and often more realistic contexts that students will experience more interest. Such reasoning is supported by similar studies in the physics education context, with Haussler et al. (1998) reporting that contexts were able to explain up to 60% of the variation in student interest scores. 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.

The activities that students encounter in the learning of statistical concepts will also influence self-reports of their interest. Deci (1992:49) argued that a person will experience interest when they encounter novel activities in a context that allows for the satisfaction of their basic psychological needs; competence, autonomy and social-relatedness. Novelty can often be created through the use of technology and in fact several authors have argued that technology can enhance students' interest in both a secondary mathematics (Mitchell 1993) and tertiary statistics (Bakker, Derry & Konold 2006) context. A student's need for social relatedness and autonomy is dependent on the classroom environment and may be met if the student is permitted to work in a group and with some degree of autonomy. In a secondary mathematics context, for example, Mitchell (1993) reported that group work could trigger students' situational interest. He also reported that if students encountered activities that they considered were meaningful, then their interest would develop into a more enduring individual interest. It is argued that in a statistics context, data exploratory activities that enable students to answer meaningful questions will enhance their interest.

Table 1: Model of interest in statistical literacy

Value: importance to them personally.	Emotion: reflective interest in doing statistics.	Epistemic curiosity: desire to know about statistics.
Interest in statistical literacy subject matter: i. Sampling ii. Graphs iii. Averages iv. Chance v. Inference vi. Variation		
Interest in the learning of statistical literacy. Evoked from: i. Classroom and school contexts that may involve the use of data relating to the students themselves. ii. Wider contexts, including sports and social issues that may or may not be presented in the media. And/or evoked from learning activities that include: i. The use of technology. ii. Group as opposed to individual learning activities iii. The use of data to answer meaningful questions.		

Methodology

The methodology associated with the creation of an instrument to measure interest in statistical literacy is principally one of establishing the validity of such an instrument. As Kane (2006:22) cautions, however, the establishment of validity is an argument that requires a research program rather than single empirical study. Consequently this paper reports the first tentative steps in a research program aimed to establish the validity of SLIM.

Creation of item bank

The model of interest, as discussed above, delineates the *universe of generalization* (Kane 2006:33) which is the set of all possible self-descriptions that could be used to assess interest in statistical literacy. From this universe of generalization a bank of 40 such descriptions was written in order to reflect the proposed components and dimensions of the theoretical model. A sample of these self-descriptions is shown in Table 2 which also details the dimension and context that each item is thought to assess.

Table 2: sample of items from the universe of generalization

Dimension	Topic	Context/activity	Item
Emotion	Data collection	Social issues	I'm interested in surveys that find out how people feel about things, e.g. school.
Value	Inference	Sports	It's important to me personally that I can use data to investigate questions (e.g. which athlete has performed best).
Epistemic curiosity	Chance	Social issues	I would like to know how scientists calculate the chance of rain.

All of the self-descriptions were then reviewed by a panel including academics and practising teachers. This resulted in 23 suitable self-descriptions that were subsequently presented with a five point Likert scale ranging from 1 (statement doesn't describe me at all) to 5 (statement describes me well).

Piloting SLIM

The second step in the development of SLIM was to administer the instrument to a suitable sample of students. Before doing this, however, additional items from the previously validated *Mathematics Interest Inventory* (Stevens & Olivarez 2005) were also included in the questionnaire in order to provide a measure of concurrent validity. The Mathematics Interest Inventory (MII) contains 10 self-descriptions, for example 'I am interested in maths' and 'I want to know all about maths'.

Analysis of pilot data

Kane (2006:23) recommended that a key step in the validation process is to establish the link between an individual's observed performance and the scoring method used. In many studies, the ordinal data produced from Likert scales is erroneously treated as interval data. Michell (1997:361), however, has argued strenuously against this practice, claiming that it has resulted in 'a systemically sustained blind spot'. Fundamental measurement, as is conducted in the sciences, requires both order and additivity (Michell 1990:chap.3). It is difficult to meet such requirements with psychological variables unless they are measured conjointly (Luce & Tukey 1964). Such measurement involves the transformation of two ordinal variables into a third variable. If a number of mathematical conditions are met it is proven (Krantz et al. 1971 cited in Michell 1990:79) that the variables can be considered as quantitative and thus measured on an interval scale.

The Rasch model, named after the Danish mathematician Georg Rasch (1901 – 1980), is a theoretical stochastic model that meets the requirements of conjoint measurement. The Rasch model for ordered responses, shown in Equation 1, gives the probability of a student selecting a Likert category for a specific item. More specifically the probability that student *n* will answer the k^{th} category of item *i*, denoted P_{nik}, is given by:

$$P_{nik} = \frac{\exp(\beta_n - \delta_i - \tau_k)}{1 + \exp(\beta_n - \delta_i - \tau_k)}, \qquad \text{Equation (1)}$$

where: β_n is the interest level of student *n*, δ_i the interestingness of item *i*; and, τ_k the additional level of interestingness to δ_i if a person chooses the k^{th} category.

A student's interest, as measured by their total response to the items in the scale, and the interestingness of the items, as measured by the total student response to each item, will form a conjoint system if the empirical data is sufficiently close to the expectations of the theoretical model. Fit statistics, discussed below, that are based on the difference between estimated model parameters and actual observed values are used to assess this. Based on these fit statistics, items with poor fit are analysed, modified and in some cases removed, so that the observed data is as close as possible to a conjoint system.

Student responses for all items in this study were analysed using the Rasch modelling program Winsteps (Linacre 2006). This program produces a model fit statistic, termed the *infit*, which is an information weighted sum of squared standardised residuals (see Equation 2). More specifically, given that the actual response of student *n* to item *i* is X_{ni} and the Rasch model expected response is E_{ni} , the infit statistic for the item, v_i is defined as:

$$v_i = \frac{\sum_{n=1}^{N}(X_{ni} - E_{ni})^2}{\sum_{n=1}^{N} W_{ni}}. \qquad \text{Equation (2)}$$

where, W_{ni} is the expected variation associated with each response X_{ni} and *N* the number of students in the study. The expectation of this statistic is one, so that items whose fit statistics differ considerably from this can be regarded as being inconsistent with the model. Bond and Fox (2007:243) recommend that for rating scales $0.6 \le v_i \le 1.4$.

In this study an iterative approach was used to establish construct validity. Initially the model was applied to student responses to all items. Items whose fit statistic was inconsistent with the model were flagged for further analysis, modification or removal. The model was then applied to the student responses of progressively smaller subsets of the original items in order to obtain a subset with suitable fit statistics. In this process, however, particular attention was taken to ensure that content validity was maintained.

Results

The sample

There were 140 students in the pilot study taken from four Queensland schools. Of these students, 49 were male and 91 were female. Their mean age was 13.0 years. Forty eight students were sampled from year 7; 60 from year 8; and, 32 from year 9.

Statistical analysis of items

As described above, an iterative approach was used to identify the most suitable items for the proposed scale. As a result of this refinement process a measurement scale consisting of 20 items was created. This scale was able to explain 62% of the variation in student responses and produced a reliability estimate of $\alpha = 0.93$. Item fit statistics for the scale are shown in Table 3, and the actual items in Appendix 1. An item map, showing the placement of students and item thresholds is shown in Figure 1.

Using Rasch analysis, eight of the ten items in the MII were found to form a measure that explained 75% of the variance in student responses and produced a reliability estimate of $\alpha = 0.94$. The correlation between these measures of student interest in maths and interest in statistical literacy was moderate ($r = 0.52$) and significant ($p < 0.01$).

Discussion

As shown in Figure 1, the 20 items of SLIM constitute a measure of statistical literacy that spans a range of 3.6 logits[4]. Students' interest measures, however, extend below the easiest (most interesting) item, suggesting that further item development in this range of the interest scale is required.

The relatively small proportion of variance explained by SLIM is of concern and points to the highly idiosyncratic nature of individual interests and in particular their relation to the contexts used in this instrument. The use of more general items, similar to those used in the MII, may produce a measure that explains a greater proportion of variance but this will reduce its fidelity. Such reasoning suggests that a second measure, containing items that assess students' interest in statistics generally, be developed and used in conjunction with SLIM. This instrument should provide a more accurate benchmark figure for students' interest, while SLIM will provide richer details.

The moderate correlation between interest in mathematics and interest in statistical literacy supports the concurrent validity of the instrument. One would expect that as middle school statistical concepts are taught in the mathematics curriculum, a student's affect towards mathematics and towards statistics, should be positively associated. The strength of such a relationship has not been explored in the literature, although in a study of 166 undergraduate psychology students, Tremblay (2000) reported a moderately strong association ($r = 0.66$) between mathematics interest and statistics attitudes. The weaker association reported in this study may be due, in part, to the tendency of adolescents to have less stable emotions than adults (Larson et al. 2002).

[4] The logit is the unit of measure in Rasch scales and is the natural logarithm of the odds ratio.

Table 3: Items assessing statistical literacy together with relevant statistics

Item	Difficulty δ_i	Standard error $SE(\delta_i)$	Infit statistic v_i
1	0.09	0.08	1.23
2	0.23	0.11	0.9
6	0.06	0.1	1.2
7	0.22	0.09	1.02
9	0.32	0.09	1.02
10	0.22	0.09	1.13
11	0.29	0.09	1
12	0.08	0.1	0.88
16	0.03	0.08	1.06
17	0.16	0.08	0.87
19	0.61	0.09	0.8
20	-0.06	0.08	0.87
21	-0.06	0.08	1.01
22	-0.19	0.08	1.14
23	0.16	0.08	0.78
24	-0.45	0.08	0.91
26	-0.43	0.08	1.15
27	-0.54	0.08	1.17
28	-0.25	0.08	0.88
29	-0.47	0.08	1.03

Conclusion

The items described in Appendix 1 form a measure of statistical literacy that span the domain of generalization that is defined by the theoretical model shown in Table 1. While the statistical properties of the measure support its construct validity; they also suggest the need for further development. In particular the creation of a more general measure of students' interest in statistics and further items that are able to assess lower levels of interest. The development of SLIM, therefore, will be an ongoing process that extends beyond that which is reported in this paper.

Acknowledgements

Thanks are extended to Dr. Rosemary Callingham, Professor Ian Hay, and Professor Jane Watson for their kind assistance and advice. This research was funded by Australian Research Council grant number LP0669106, with support from the Australian Bureau of Statistics.

Difficulty (logits)	Person locations	Item threshold locations
2		
1.8		19.5,
1.6		9.5, 11.5
1.4		1.5, 7.5, 10.5, 12.5, 17.5, 23.5, 2.5, 6.5
1.2	.	16.5, 20.5, 21.5
1	#	19.4, 22.5, 28.5
0.8	##	24.5, 26.5. 29.5
0.6	#####	7.4, 9.4, 10.4, 11.4, 2.4, 27.5
0.4	###	1.4, 12.4, 16.4, 17.4, 23.4, 6.4
0.2	########	19.3, 20.4, 21.4, 22.4
0	######	9.3, 11.3, 26.4, 28.4
-0.2	######	1.3, 7.3, 10.3, 12.3, 17.3, 23.3, 2.3, 6.3, 24.4, 27.4, 29.4
-0.4	######	16.3, 20.3, 21.3
-0.6	#########	19.2, 22.3, 28.3
-0.8	####	24.3, 26.3, 27.3, 29.3
-1	#####	7.2, 9.2, 10.2, 11.2, 17.2, 23.2, 2.2
-1.2	####	1.2, 12.2, 16.2, 20.2, 21.2, 6.2
-1.4	#	22.2, 28.2
-1.6	#	24.2, 26.2
-1.8	#	27.2, 29.2
-2	#	
-2.2	#	
-2.4		

1. # denotes 2 students
2. 9.5 denotes location of the threshold for the 5th category (‘describe me well’) of item 9

Figure 1: Item map for SLIM

References

Bakker, A., Derry, J. & Konold, C. 2006, 'Using technology to support diagrammatic reasoning about center and variation', *Proceedings of the 7th International Conference on Teaching Statistics, IASE*. Retrieved 18 April 2008 from http://www.stat.auckland.ac.nz/~iase/

Bidgood, P. 2006, 'Creating statistical resources from real datasets: The STARS project', *Proceedings of the 7th International Conference on Teaching Statistics, IASE*. Retrieved 18 April 2008 from http://www.stat.auckland.ac.nz/~iase/

Bond, T.G. & Fox, C. M. 2007, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 2nd edn, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Deci, E.L. 1992, 'The relation of interest to the motivation of behavior: A self-determination theory perspective', in *The Role of Interest in Learning and Development*, eds K.A. Renninger, S. Hidi & A. Krapp, Lawrence Erlbaum Associates, New Jersey, pp.43-70.

Gal, I. 2003, 'Teaching for statistical literacy and services of statistics agencies', *The American Statistician*, vol.57, no.2, pp.80-84.

Haussler, 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*, vol.20, no.2, pp.223-238.

Hidi, S. & Renninger, K.A. 2006, 'The four-phase model of interest development', *Educational Psychologist*, vol.41, no.2, pp.111-127.

Hoffman, L. 2002, 'Promoting girls' interest and achievement in physics classes for beginners', *Learning and Instruction*, vol.12, pp.447-465.

Kane, M.T. 2006, 'Validation', in *Educational Measurement*, 4th edn, ed R.L. Brennan, American Council on Education and Praeger Publishers, Westport (CT), pp.17-64.

Larson, R.W., Moneta, G., Richards, M.H. & Wilson, S. 2002, 'Continuity, stability, and change in daily emotional experience across adolescence', *Child Development*, vol.73, no.4, pp.1151-1165.

Lee, C. & Famoye, F. 2006, 'Teaching statistics using a real time online database created by students' , *Proceedings of the 7th International Conference on Teaching Statistics, IASE*. Retrieved 18 April 2008 from http://www.stat.auckland.ac.nz/~iase/

Linacre, J.M. 2006, '*Winsteps 3.61.2*', (Computer Software), Winsteps.com, Chicago.

Litman, J.A., & Spielberger, C.D. 2003, 'Measuring epistemic curiosity and its diversive and specific components', *Journal of Personality Assessment*, vol.80, no.1, pp.75-86.

Lock, R.H. 2006, 'Teaching an introductory statistics class based on sports examples', *Proceedings of the 7th International Conference on Teaching Statistics, IASE*. Retrieved 18 April 2008 from http://www.stat.auckland.ac.nz/~iase/

Luce, R.D. & Tukey, J. W. 1964, 'Simultaneous conjoint measurement: A new type of fundamental measurement', *Journal of Mathematical Psychology*, vol.1, no.1, pp.1-27.

Michell, J. 1990, *An Introduction to the Logic of Psychological Measurement*, Lawrence Erlbaum Associates, Hillsdale (NJ).

Michell, J. 1997, 'Quantitative science and the definition of measurement in psychology', *British Journal of Psychology*, vol.88, pp.355-383.

Mitchell, M. 1993, 'Situational interest: Its multifaceted structure in the secondary school mathematics classroom', *Journal of Educational Psychology*, vol.85, no.3, pp.424-436.

Mitchell, M. 1997, 'Situational interest in the statistics classroom', *Proceedings of the Annual Meeting of the American Educational Research Association*, Retrieved on 18 April 2008 from ERIC database.

Schiefele, U. 1991, 'Interest, learning, and motivation', *Educational Psychologist*, vol.26, no.3 & 4, pp.299-323.

Schiefele, U. & Csikszentmihalyi, M. 1995, 'Motivation and ability as factors in mathematics experience and achievement', *Journal for Research in Mathematics Education*, vol.26, no.2, pp.163-181.

Schiefele, U., Krapp, A. & Winteler, A. 1992, 'Interest as a predictor of academic achievement: A meta-analysis of research', in *The Role of Interest in Learning and Development*, eds K. A. Renninger, S. Hidi & A. Krapp, Lawrence Erlbaum Associates , New Jersey, pp.183-212.

Stevens, T. & Olivarez, A. 2005, 'Development and evaluation of the mathematics interest inventory', *Measurement and Evaluation in Counseling and Development*, vol.38, no.3, pp.141-152.

The Macquarie Dictionary 1997, 3rd edn, Macquarie Library, Sydney

Tremblay, P.F., Gardner, R.C. & Heipel, G. 2000, 'A model of the relationship among measures of affect, aptitude and performance in introductory statistics', *Canadian Journal of Behavioural Science*, vol.32, no.1, pp.40-48.

Watson, J. & Callingham, R. 2003, 'Statistical Literacy: A complex hierarchical construct', *Statistics Education Research Journal*, vol.2, no.2, pp.3-46.

Watson, J.M. 2006, *Statistical Literacy at School: Growth and Goals*, Lawrence Erlbaum Associates, Mahwah, New Jersey.

Appendix 1: SLIM

No	Item
I'm interested in:	
1	Doing magazine or online surveys
2	Surveys that find out how people feel about things, e.g. school or climate change.
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 probabilities (or chances) for dice, coins and spinners.
12	Creating graphs for my data using computer programs.
I would 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.
It's important to me personally that I:	
23	Can understanding news reports that use averages.
24	Know how the chance of being injured from risky behavior (e.g. drink driving) is calculated.
26	Can believe scientific claims based on data, e.g. 'smoking is bad for health'.
27	Use the use correct graph when displaying my data.
28	Can understand graphs that appear on the internet or in newspapers.
29	Can arrange data into tables (using computers or by hand).