Trialling a Professional Statistical Literacy Hierarchy for Teachers

Robyn Pierce
University of Melbourne
<r.pierce@unimelb.edu.au>

Helen Chick
University of Tasmania
<Helen.Chick@utas.edu.au>

Jane Watson
University of Tasmania
<Jane.Watson@utas.edu.au>

Michael Dalton
Victorian Curriculum and Assessment Authority
<dalton.michael.m@edumail.vic.gov.au>

Magdalena Les
Victorian Curriculum and Assessment Authority
<les.magdalena.m@edumail.vic.gov.au>

Teachers now require high levels of statistical literacy in order to take advantage of the many statistical reports analysing assessment data that are provided by system authorities. In this report 16 items from the Attitudes and Statistical Literacy Instrument (ASLI) are used with 704 teachers to provide a hierarchical scale of teacher ability to interpret these assessment data. Using Rasch analysis, three levels of ability are identified, related to reading values, comparing values, and analysing a data set. Implications are drawn for professional learning for teachers and for further research.

Introduction

Evidence-based practice is expected of today's principals and teachers. Schools must provide education system authorities with data on a wide variety of variables, including student assessment data. These are analysed and then returned to schools and teachers in the form of statistical reports. Schools are expected to consider these data for planning and decision-making, requiring commensurate statistical skills. This is just one example of the need for statistical literacy in today's society generally, and workplaces more specifically. In this paper we report on research trialling a proposed framework (Figure 1) for professional statistical literacy by considering the Rasch analysis of the responses to survey items based on this framework. First some background is provided and the details of the study explained. Next the results of the Rasch analysis are discussed and finally some conclusions drawn. Our purpose is to determine whether or not there is a hierarchy of understanding for professional statistical literacy, and what areas are difficult for teachers.

Professional Statistical Literacy

A Framework for Professional Statistical Literacy

A number of authors have proposed hierarchies for statistical understanding. Curcio's graph comprehension study (1987) involved relatively elementary data sets, and identified three levels: "reading the data" (directly reading factual information on the graph), "reading between [or within] the data" (attend to multiple points, often for comparison), and "reading beyond the data" (predict and infer from the data). Shaughnessy and others (1996, 2007) posit a further category focusing on context, named "reading behind the data". Watson's three-tiered statistical literacy hierarchy (2006) also emphasised the place of context. Her first tier involves understanding basic terminology, the second tier focuses on understanding language and concepts within the context of wider social issues, and the third tier focuses on challenging and questioning statistical claims. Finally, Gal's proposed statistical knowledge base (2002, p. 10) emphasises knowing why data are needed, having familiarity with basic terms, and understanding how statistical conclusions are reached.

Pierce and Chick's (2011) framework for considering professional statistical literacy draws on and synthesises this background (Figure 1). The teacher must be able to examine data at multiple levels, each more complex than and dependent on the lower levels (as

In J. Dindyal, L. P. Cheng & S. F. Ng (Eds.), Mathematics education: Expanding horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia). Singapore: MERGA.

indicated by the nested circles in Figure 1). The lowest level, *reading values*, involves understanding features such as keys, scale, and graph type, together with the capacity to read specific data points on the graph. The second level, *comparing values*, requires awareness of relative and absolute differences, early informal inference, and low-level statistical tools. Finally, *analysing the data set* involves being able to consider the data as a whole: observing and interpreting variation, trends, and changes with time or other variables; and attending to the significance of results. The framework also acknowledges the role of context, with two key contexts identified as being significant (the surrounding split rectangle in Figure 1). *Professional context* concerns information relevant to the profession and needed to interpret the data set (e.g., meaning of specialist terms such as “band”, “VELS level”); and the *local context* comprises knowledge about the specific data set that is not evident in the data set alone (e.g., knowledge of local school situation). These two context components may overlap, hence the dashed line between them in Figure 1. This framework underpinned both the structure of and items created for the study, in order to assess school principals’ and teachers’ statistical literacy.

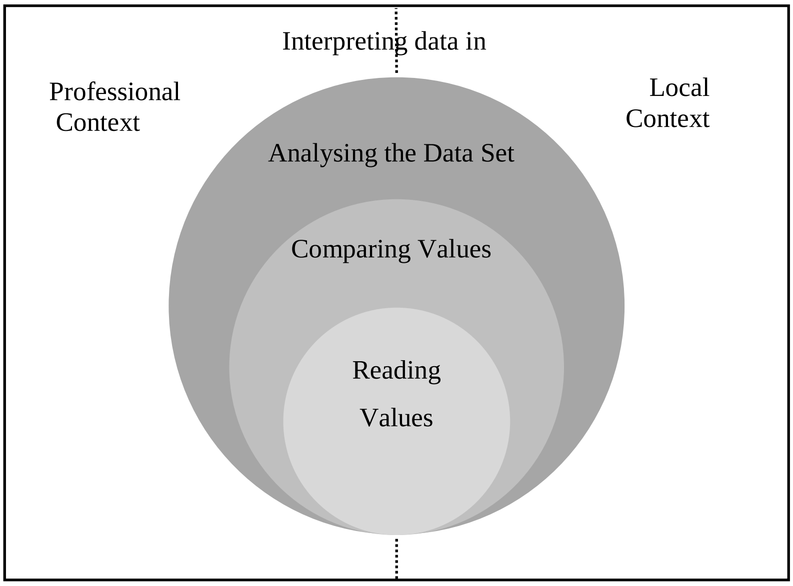


Figure 1. A framework for considering professional statistical literacy (Pierce & Chick, 2011, p. 633).

Tools for Professional Statistical Literacy

The statistical tools required to analyse data are dependent on the form in which the data are supplied. In the present case, the main graphical tool that is used for interpreting data sets provided by systems is the box plot, although some data are supplied in tabular form, with numerical summary statistics. Appropriate interpretation of box plots is hence essential for high-level statistical literacy. Although straightforward definitions of the box plot are found, for example, in the *Australian Curriculum: Mathematics* (Australian Curriculum, Assessment and Reporting Authority, 2011, p. 53), many variations in format exist, as summarised by Wickham and Stryjewski (2011) in their discussion of the 40-year history of the box plot. It is thus important to read carefully the keys associated with any presentation of box plots. In addition, there is a subtlety associated with the frequency versus density representation of a box plot compared, say, to a histogram. In a histogram area represents frequency of values, whereas in a box plot area represents density of values. Bakker, Biehler, and Konold (2005) presented data on this confusion from students as a reason for postponing the introduction of box plots to the later middle school years. Teachers’ ability to interpret box plots is hence a major component of assessing their statistical literacy understanding.

The Study

Instrument

This study was conducted with teachers from Victorian government schools, and was Stage 2 of a larger project. Stage 1 had involved the use of pen-and-paper questionnaires and face-to-face focus group discussions with 152 teachers (see Pierce & Chick, 2011). This led to the development of an online questionnaire for Stage 2, the focus of this study. The online Attitudes and Statistical Literacy Instrument (ASLI) took respondents about 20 minutes to complete and included items used in Stage 1 but refined to use the Stage 1 teachers' own language in the case of some items, and to focus on key issues identified in Stage 1. The number of items was reduced from the original pen-and-paper questionnaires to target misconceptions and knowledge gaps identified in Stage 1.

This paper reports data from the 16 items that formed the statistical literacy section of the online ASLI (sample items are shown in Figures 2 and 3). The first 9 items (numbered 1-9 for the purposes of the Rasch analysis and corresponding to items numbered 13-18, including sub-questions, in the actual ASLI) were based on typical reports regarding student achievement that are returned to Victorian schools following analysis of NAPLAN data. These reports were chosen because NAPLAN reports are sent to most schools and the reports of Victorian Certificate of Education data follow a similar format. Some further items (numbered 10-16 in the Rasch analysis, or 19-25 in the ASLI) targeted teachers' understanding of distributions represented by histograms and boxplots. Items numbered 8, 9, 11, and 16 in the Rasch analysis were open response items, whereas all others were multiple-choice items using distractors based on data from the Stage 1 questionnaires and interviews. [For this paper, the Rasch item numbering will predominate, but the ASLI numbering is included to allow cross-referencing with other papers from this data set.]

Data Analysis

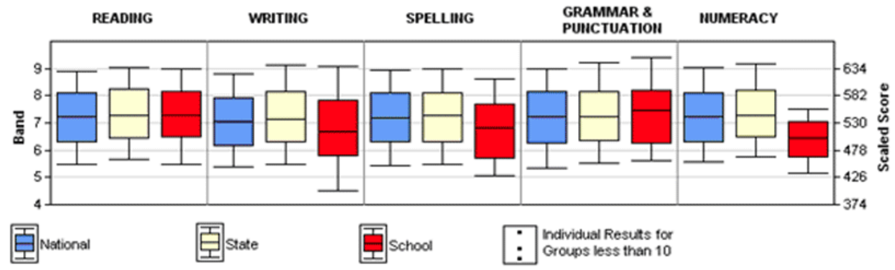
For multiple response items the options were coded on a numerical scale where the best or correct answer was allocated the highest score. Open response items were numerically coded by comparing them with scoring rubrics, which allowed partial credit for responses exhibiting partial understanding. Scoring was done primarily by one researcher with a second independently scoring a systematic sample of at least 10% of the responses. Agreement between scorers was at least 80%. Once all of the scoring and coding was complete the data underwent Rasch analysis using *ConQuest* (Australian Council of Educational Research, n.d.) to determine whether, as hypothesised, professional statistical literacy is a unidimensional hierarchical construct.

Participants

The goal for Stage 2 was to gather data from a random sample of 1000 Victorian primary and secondary school teachers (excluding those who had participated in Stage 1 of the project). In anticipation of a 60% school participation rate and then a participation rate of 50% from teachers within schools, 104 randomly selected schools were approached. Sixty percent of principals agreed to their school's participation but in the end only 704 teacher responses were received (217 male, 487 female; 314 primary and 390 secondary teachers). This still provides a sufficiently large sample from which to draw tentative conclusions about areas of success and difficulty for teachers with respect to professional statistical literacy, and for examining the hierarchical nature of this construct.

SPELLING	National	251824	448.8	493.9	540.7	587.2	631.4	540.0	
	State	62225	450.8	494.9	513.6	587.4	632.8	541.1	
	School	45	428.1	453.6	520.0	587.0	613.0	518.0	10.5
GRAMMAR & PUNCTUATION	National	251824	443.1	491.3	540.9	589.1	633.8	539.5	
	State	62225	453.1	497.0	542.6	591.0	644.7	545.0	
	School	45	457.1	492.5	553.0	531.8	684.2	553.1	13.0
NUMERACY	National	250466	455.0	494.7	540.8	588.9	635.7	543.6	
	State	62003	484.8	503.1	545.0	593.2	643.3	550.8	
	School	43	435.0	454.2	502.4	533.2	556.4	499.1	7.1

THIS TABLE DISPLAYS SCALED SCORES



THIS REPORT USES THE NATIONAL ASSESSMENT PROGRAM SCALE

***13. How many students in the State sat the 2009 Year 7 SPELLING Test?**

- ○ 251353
- ○ 62149
- ○ 45
- ○ 62225
- ○ 251824
- ○ I don't know

Figure 2. Item 1 (ASLI item 13) from the on-line ASLI.

***15. Look at the 'Writing' and the 'Spelling' results for Victoria College.**

Choose the best response for each of the following statements:

	DEFINITELY TRUE	DEFINITELY FALSE	NOT ENOUGH INFORMATION	I DON'T KNOW
For Writing the median result is higher than the mean result	○	○	○	○
The spread of the 50th – 90th percentile results is wider in Writing than in Spelling	○	○	○	○
In the Writing results, fewer students were between the 25th percentile and the median than were between the median and the 75th percentile	○	○	○	○
Victoria College's Writing results have a greater range than the State results	○	○	○	○

***16. From the results displayed what would you identify as the school's weakest area? (Choose one only)**

- ○ Reading
- ○ Writing
- ○ Spelling
- ○ Grammar & Punctuation
- ○ Numeracy

Figure 3. Items 3-7 (ASLI items 15—with sub-items— and 16) from the on-line ASLI.

Results

The Rasch analysis of data from the ASLI showed that all items had individual weighted mean square values between the commonly accepted values of 0.75 and 1.30 (Bond & Fox,

2007) and that the item infit mean square was 1.00 (S.D. 0.124), suggesting that the items form a hierarchical unidimensional scale. An alpha reliability value of 0.81 indicates that the scale also meets the requirements of classical psychometric theory. The Rasch analysis produced the graph of respondent abilities (on the left) and item difficulty (on the right) in Figure 4. The existence of many respondents above logit 0 suggests that the ASLI included many that the teachers found to be easy. This was, in fact, a deliberate part of the design, in order to counter the anxieties with respect to quantitative data exhibited by many teachers in the pilot study.

The right side of Figure 4 provides some details of the easiest and hardest four items, as indicated by the Rasch analysis. These results show that the easiest items were focused on the reading and comparing levels of the framework (the inner circles of Figure 1), and the hardest items focused on the analysing and interpreting data in context levels (the outermost circle, together with the framing context rectangle).

The easiest item, as determined by the Rasch analysis, was an item that required teachers to locate a single datum from a table based on the locating information supplied in the item. The next easiest items involved some form of comparison between or among values. Two of these, Rasch items 3 and 13, involved comparison of two things: in the first case, of two readable data points, and, in the second, of the spreads of two graphs. Since spread involves attention to the extreme values of the graph, this comparison necessarily involves more than two data points, which suggests this item should have been the harder of the two. That this was not the case may be due to the complexity of locating, in a large complex table, the two specific values needed for the comparison in item 3. The remaining easy item, Rasch item 7, involved the comparison of box plots, and asked teachers to identify the school's weakest area from the data presented in Figure 2. The work of Pfannkuch (2006) highlights that a full comparison of box plots involves consideration of the five number summaries of the relevant data sets (i.e., the critical components of the box and whiskers). This suggests that item 7 had the potential to be a difficult item, because of the number of comparisons to be made, both between any two box plots, and across the collection of box plots. However, the school's numeracy box plot (the right-most box plot in Figure 2) is so markedly different from and lower than the others at the macro scale that even a simplistic reading and comparison of the plots could yield the correct answer. It is hypothesised that if the differences had been less "gross" then this item would have been more difficult.

At the other extreme, most of the harder items involved not only analysis of the data set as a whole entity, but attention to context as well. Item 5 addressed the "density vs frequency" misconception associated with box plots, in which the size of a region is taken to be proportional to the number of data points in it (which is true for histograms but not for box plots). To answer item 5 correctly teachers had to understand that the two components of the box in a box plot—determined by the median and the 25th and 75th percentiles—actually contain the same number of data points. This involves understanding how a box plot depicts the whole data set, and knowing the technical meanings of median and percentile.

Items 6 and 9 addressed particular peculiarities of the reports that are supplied to teachers by government agencies. This brings professional context into play, because these characteristics are a consequence of how the agencies choose to depict the data in their reports. Although the key for the box plots in Figure 2 clearly states that the whiskers end at the 10th and 90th percentiles, the results indicate that teachers were misled by the visual image of the box plot and had a tendency to assume that it showed the full range of the data, rather than hiding the results of the top and bottom ten percent of students. Similarly, item 9 asked teachers to explain the differences between two kinds of box plot representation that they receive, neither of which show the full data set. The difficulties with these two items highlight the power of a graphical image to give rise to misleading convictions, in this case associated with ignoring what is not seen.

```
                |                |
  3             |                |
                |                |
                |                |
                |                |
               X|                |
               X|                |
               X|                |
               X|                |
               X|                |
  2           XX|                |
              XX|                |
             XXX|                |
            XXXX|                |
           XXXXX|16              |
        XXXXXXXX|                |
          XXXXXX|                |
          XXXXXX|9               |
  1      XXXXXXX|                |
        XXXXXXXX|                |
      XXXXXXXXXX|6               |
      XXXXXXXXXX|5               |
        XXXXXXXX|                |
       XXXXXXXXX|2 15            |
        XXXXXXXX|8               |
       XXXXXXXXX|                |
       XXXXXXXXX|                |
  0     XXXXXXXX|                |
           XXXXX|11              |
          XXXXXX|10 12 14        |
            XXXX|                |
             XXX|                |
             XXX|4               |
             XXX|                |
              XX|3               |
               X|                |
 -1            X|13              |
               X|                |
               X|                |
               X|                |
               X|1 7             |
                |                |
               X|                |
               X|                |
                |                |
 -2            X|                |
                |                |
================================
Each 'X' represents   4.0 cases
```

Rasch Item (ASLI Item)	**Main aspect of statistical literacy being assessed**	**Level of Framework**
16 (25)	Relationship between histogram and boxplot with respect to skewness, range and location of key percentiles	Analysing the data set
9 (18)	Reading and interpreting the key of a graph	Analysing with Professional context knowledge
6 (15iv)	Knowledge of subset relationship between School and State; reading key to NAPLAN-style boxplots	Analysing with Professional context knowledge
5 (15iii)	Knowledge of meaning of “percentile”; density vs frequency in a boxplot	Analysing the data set
3 (15i)	Reading a table	Compare (low level)
13 (22)	Comparison of horizontal aspects of histogram	Compare (low level)
7 (16)	Comparison of relative positions of School boxplots	Compare
1 (13)	Reading a table	Read

Figure 4. Left: Rasch analysis map of latent distributions and response model parameter estimates. Right: Focus of 4 easiest and 4 hardest items as identified by Rasch analysis

The most difficult item was an open-response item asking teachers to explain why a particular box plot represented some data that had originally been presented in a frequency histogram. In this case, the context was irrelevant (although one was supplied), but, in order to receive full credit, teachers had to identify and link key points on the box plot with the original histogram data. This required extensive analysis of the full data set in order to locate the appropriate percentiles and median. The multiple choice item that preceded this on the ASLI, item 15, simply required the selection of the appropriate box plot rather than justification for the choice, but was also the equal fifth-hardest item overall.

The items of middle difficulty, according to the Rasch analysis, generally involved comparisons, or relatively straightforward analyses of the data set as a whole. Two of these are worth examining more closely. Item 12 required teachers to determine which of two distributions, displayed as histograms, had the larger mean; however, in order to design data sets that led to an unambiguous distinction in the means, the resulting distributions presented to teachers were quite noticeably distinct. This made it possible to guess correctly based on a superficial examination of the graphs rather than a deep analysis of the range of possible values for the means (which is what the authors needed to consider when designing the item). This may explain why this item was not among the harder items; it seems plausible that a similarly focused item could be "designed" to be quite difficult. Item 2, in contrast, appeared to involve the straightforward reading of data from a graph, but in this case professional contextual knowledge was needed in order to realise that the graph datum could only take whole number values, rather than the apparent continuum of values that were shown on the vertical axis. Most teachers who got this wrong were successful in reading an appropriate fractional value, but did not know that the corresponding variable was discrete. Here the impact of professional context knowledge on professional statistical literacy is particularly evident.

Implications and Conclusions

The results of the Rasch analysis suggest that there is, indeed, a hierarchy for professional statistical literacy that reflects, to a reasonable degree, the different levels posited in the framework in Figure 1. Teachers appeared to have little difficulty with direct data reading of single values and simple comparisons of data values or comparisons among whole data sets where the distinctions were grossly evident. The items that caused greater difficulty were those that required deeper analysis of the whole data set, and attention to the technicalities of the data presentation in this particular professional context. Because of the scope of the survey, which was intended to be used with teachers across a range of school contexts, and hence had to use generic data rather than data from their own specific schools, the role of local context could not be explored.

Although it appears that we can safely assume that teachers can read straightforward data values, the results indicate that teachers do not always have a good understanding of the technicalities of certain kinds of representations. In this case, the extensive use of box plots for presenting data causes difficulties, particularly since the ones used in the reports have certain idiosyncrasies, such as the cut-off percentiles for the whiskers. Although a key was supplied, the basic graph-reading principles of actually reading the key of a graph, seemed sometimes to be neglected by data users. Alternatively, the implications of this information were not fully understood, perhaps because of an assumption about already knowing what was being depicted. In follow-up work with Stage 1 teachers, participants showed they were not fully cognisant of the implications of the fact that the box plots do not show the top and bottom 10% of the data, being noticeably surprised on realising that their strongest and weakest students were, in effect, invisible in the data representation. Similarly, the prevalence of the density versus frequency misconception, observed by Bakker et al. (2005) for school students, also suggests that the value of box plots is being undermined by lack of familiarity or fluency with this graph type. It is not clear that alternative representations might be better, because in any representation there is a need to understand how it depicts the whole data set, and to consider for what is displayed or hidden by a particular representation. Regardless of the representation type, it seems that professional learning for teachers should focus on gaining fluency with holistic analysis and how whole data sets are depicted.

The professional statistical literacy framework provides a reminder of key issues to be addressed in the teaching of statistics, to ensure that high-level statistical literacy skills, applicable beyond school, are developed. In particular, the results and framework highlight the impact of and need for paying attention to statistical context (e.g., key and scales), along with professional and local contexts in all work with data. De-contextualised skills may result in incomplete interpretations of data, limiting the statistical literacy of the user.

Acknowledgements

We thank the schools and teachers for their involvement in this project. We also acknowledge the contributions of other members of the research team: Sue Buckley (Victorian Department of Education and Early Childhood Development); Ian Gordon, Sue Helme, Roger Wander (University of Melbourne). This research is funded by the Australian Research Council (LP100100388), the VCAA, and the DEECD.

References

Australian Council for Educational Research. (n.d.). *ConQuest: Generalized item response modelling software*. [Computer software]. Melbourne: Author.

Australian Curriculum, Assessment and Reporting Authority. (2011). *The Australian Curriculum: Mathematics, Version 1.2, 8 March 2011*. Sydney, NSW: Author.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education (IASE) Roundtable, Lund, Sweden, 28 June-3 July 2004* (pp. 163-173). Voorburg, The Netherlands: International Statistical Institute.

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

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382-393.

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

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal, 5*(2), 27-45.

Pierce, R. & Chick, H. (2011). Reacting to quantitative data: Teachers' perceptions of student achievement reports. In J. Clark, B. Kissane, J. Mousley, T. Spencer, & S. Thornton (Eds.), *Mathematics: Traditions and [new] practices* (Proceedings of the 34th Annual Conference of the Mathematics Education Research Group of Australasia, Alice Springs, pp. 631-639). Adelaide: AAMT/MERGA.

Shaughnessy, J. M. (2007). Research on statistical learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: Information Age Publishing.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics in education* (pp. 205-237). Dordrecht, The Netherlands: Kluwer.

Watson, J. M. (2006). *Statistical literacy at school*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wickham, H., & Stryjewski, L. (2011). *40 years of box plots*. Manuscript submitted for publication. Retrieved from http://vita.had.co.nz/papers/boxplots.html