Mathematics Teacher Education and Development 2013, Vol. 15.2, 5–26

The Statistical Literacy Needed to Interpret School Assessment Data

Helen Chick
University of Tasmania

Robyn Pierce
University of Melbourne

State-wide and national testing in areas such as literacy and numeracy produces reports containing graphs and tables illustrating school and individual performance. These are intended to inform teachers, principals, and education organisations about student and school outcomes, to guide change and improvement. Given the complexity of the information, it is of interest to determine the critical statistical skills required to make sense of such data. This paper examines the statistical literacy necessary to interpret the graphical presentations of school assessment data for the Australian NAPLAN testing process. A framework for professional statistical literacy that acknowledges the importance of context is used to identify different levels of data interpretation. The implications for helping users make better use of such data and for teacher education more broadly are discussed.

Keywords: statistics education research • data interpretation • statistical literacy • workplace data use • school improvement

The Need for Statistical Literacy

In education—as in other workplace sectors—quality control, accountability, and forward planning may be informed by examination of statistical data. The technological revolution has facilitated the collection, analysis, and sharing of vast quantities of data. National literacy and numeracy testing has become an established part of the education profile, for which all teachers must be prepared, in many countries including the UK (e.g., Children, Schools and Family Committee, 2008), the USA (e.g., Baker, 2007), and Australia (e.g., Ministerial Council on Education, Employment, Training and Youth Affairs [MCEETYA], 2007). Such tests are advocated to identify students' educational needs, to support data-driven decision-making when planning teaching practice, and promote schools' accountability to their students and their funding authority. Australia, for example, has developed a *Measurement Framework for National Key Performance Measures* (MCEETYA, 2007) for monitoring and advancing quality outcomes from school education. Integral to this framework are cycles of assessment and data collection. Extensive databases, incorporating assessment results and socio-economic data from students across Australia, allow complex linking and analysis of this information. The government's claim is that

> literacy and numeracy assessments provide rich data about individual student performance and assist teachers to plan learning activities for students. They also enable schools to develop a more objective view about the performance of their students compared to those in other schools and in relation to state-wide standards. (MCEETYA, n.d.)

Governments expend significant resources on collecting such data from the education sector with the intention that this should inform planning and practice. The *National Assessment Program—Literacy and Numeracy* (NAPLAN), for example, involves students in Years 3, 5, 7, and 9 from all States and Territories. In Victoria, approximately 260,000 students from all Government, Catholic, and Independent schools participate in the yearly NAPLAN testing, coordinated by the Victorian Curriculum and Assessment Authority (VCAA, n.d.).

Planning and conducting tests, collecting and analysing data, and producing data reports is a time-consuming and expensive process. It is therefore important that the outcomes of the process are used by systems, schools, and teachers in ways that will benefit students. There are several levels at which these data have application, with consequential demands on the statistical knowledge of the relevant users. At the systemic level, when governments examine across-system or across-schools performance, it may be necessary to employ a statistical expert in order to conduct a detailed analysis of the data and highlight issues for which schools may be expected to be accountable. Principals of schools need to engage with the data to identify whole-school issues, and they too might employ an expert for in-depth analysis, but ideally they need to be able make sense of the data for their school and determine the implications for themselves. For teachers, the data can be examined at the levels of the individual student and also the class. Consideration of individual students' data—to identify those, for example, performing below national benchmarks—is normally a simple task; what is more complex is the analysis of class data in order to inform planning for teaching.

The fact that school data are provided in graphical formats seems based on an assumption that teachers can interpret them successfully. Less than 20 years ago, however, most Australian adults would have experienced schooling with little emphasis on statistics. It was only in 1991, with the publication of *The National Statement on Mathematics for Australian Schools* (Australian Education Council, 1991), that "Chance and Data" became significantly and formally recognised across the entire P-12 curriculum in Australia. In the ensuing few years the states produced their own curricula (e.g., Board of Studies, 1995) based, in varying degrees, upon this national document. This increased emphasis on statistics has remained in versions of curricula since that time (e.g., VCAA, 2008). The relative recency of this inclusion of statistics into the school curriculum means that teachers over the age of 35 at the time of writing (2011) may have had only limited experience with statistical thinking activities. The push for statistical literacy (defined below) is more recent still, implying that a significant proportion of the general teaching workforce may not be well equipped to undertake the sort of data interpretation necessary to make best use of school assessment data like NAPLAN. So although teachers form a well-educated work force, the majority do not have formal training in statistics beyond the secondary level (Pierce, Chick, & Gordon, 2013), and many have only limited experience with analysis or interpretation of quantitative data. This deficit—which may well affect capacity to interpret statistical data—needs to be addressed through a combination of pre-service or in-service education.

The purpose of this paper is to consider the statistical demands made of principals and teachers who examine such data. It presents an analysis of the statistical literacy required to read and interpret certain reports provided by the Victorian NAPLAN data service. It begins with some background from literature, and examines frameworks for describing data interpretation skills. An analysis of three of the graphical report formats provided to schools is then conducted, with some comments about the extent to which the findings might extend to data from sources wider than just Victoria. The paper concludes with a discussion of the implications of these findings. Our purpose in this paper is not to examine teachers' actual capacities to deal with representations of complex data—a goal that is part of the larger objectives of this work—but to identify the nature of the knowledge required.

Background

Reading and interpreting statistical reports requires more than conventional literacy: it requires statistical literacy. The scope of this term encompasses sufficient knowledge and understanding of numeracy, statistics, general literacy, and data presentation to make valuable use of quantitative data and summary reports in a personal or professional setting (Ben-Zvi & Garfield, 2004; Gal, 2002; Watson, 2006). It need not imply the capacity to conduct sophisticated statistical tests, but it does include the ability to question the sampling techniques used to collect the data, interpret possible explanations and consequences of the data, and identify limitations in the data and the conclusions.

In many contexts it is important to be able to convey the results of some quantitative analysis of data to both those with and those without statistical training. Statistical graphs are used commonly in data reports, with the assumption that graphs provide a picture of the data, and that this picture conveys key messages. Good graphical communication, however, requires good graph design. In 2001 Tufte, drawing on earlier work, summarised six principles of such design. These included using physical representations that are proportional to the actual numerical quantities represented, showing data variations rather than design variation, and including clear labels and explanation of the text. Conventions for simple designs, which keep the visual coding transparent, have been established over time. Details of the design of such "standard" graphs are included in most introductory statistics text books and even school mathematics texts. These standard graphs include bar graphs, box plots (or box-and-whisker plots), and stacked dot plots (see, for example, Aczel & Sounderpandian, 2006; Howell, 2002; Moore & McCabe, 1993).

The producers of school reporting data, perhaps reasonably, assume that it is sensible to use standard graphs in data reports. If a graph is constructed clearly, without distortion, then it might be expected to convey a message; however, reading such a message requires knowledge of graph interpretation. Some of the interpretation may be straightforward, but other aspects—particularly more serious "informal inference"—require some technical skills and understanding of the statistical context, along with familiarity with the potential critical influences of the real world context. This informal inference, beyond straightforward reading of data values, is vital if the data are to be used as a basis for decision-making.

Curcio's (1987) study of graph comprehension in Year 4 and Year 8 students highlighted the ideas of "reading the data" (the capacity to read literally the direct factual information on the graph), "reading between [or within] the data" (attend to two or more data points on the graph, often for comparison purposes), and "reading beyond the data" (extend, predict, and infer from the data). More recent work of Shaughnessy and colleagues (1996, 2007) suggests an additional category termed "reading behind the data" which pays particular attention to the context from which the data arise. In Shaughnessy's (2007) handbook chapter the four categories were given more detail, and expanded into eight, with the higher-level ones associated with deeper interpretation and appreciation of context and variation.

Watson (2006) also emphasised the place of context in the interpretative process. The first tier of her three-tiered statistical literacy hierarchy involves an understanding of basic terminology, and then the second tier requires "an understanding of probabilistic and statistical language and concepts when they are embedded in the context of wider social discussion" (p. 16). The third tier concerns the ability to challenge and question statistical claims. The statistical knowledge base posited by Gal (2002, p. 10) also indicates the importance of knowing why data are needed, having familiarity with basic terms, and understanding how statistical conclusions are reached.

A Framework for Professional Statistical Literacy

As they stand, these frameworks are not ideal for the situation of examining the statistical literacy required to interpret typical real-world and workplace data; in particular, they did not meet the needs of the authors when analysing the graphs that are a focus of this study. Curcio's (1987) framework deals with relatively straightforward data sets for Year 4 and 8 students, rather than more complicated data intended for adults in a particular context. Shaughnessy's extension (2007), while useful, has more categories than appeared necessary for the current situation. Watson's statistical literacy hierarchy (2006) highlights critical components that are required for interpreting data, but it is couched in reference to the broader construct of statistical literacy.

The Framework for Professional Statistical Literacy presented in Figure 1 (below) draws on the work of all of these authors, but allows a focus on the statistical literacy specifically required for interpreting data in tabular, graphical, and other forms. Three levels are proposed for technical data interpretation, with different degrees of sophistication, and these are then embedded in consideration of contextual issues. It should be noted at the outset that this hierarchy is not intended to indicate a teaching sequence, although successful performance at the higher levels is predicated on competence at the lower levels.

The lowest level of the framework shown in Figure 1, Reading values, involves being able to read directly accessible single items explicitly evident in the data. This is akin to Curcio's (1987) reading the data or Watson's first tier, knowledge of terminology. Examples of this capacity include reading a single data point on the graph or being able to identify axis labels. A practised graph reader, when first "getting oriented to a graph", is usually functioning at this level: reading the title, checking the axes for variable names and ranges of values, and doing a quick check of one or more specific data points. In order to start "telling stories" about the data, however, the graph reader must move beyond attending to single values.

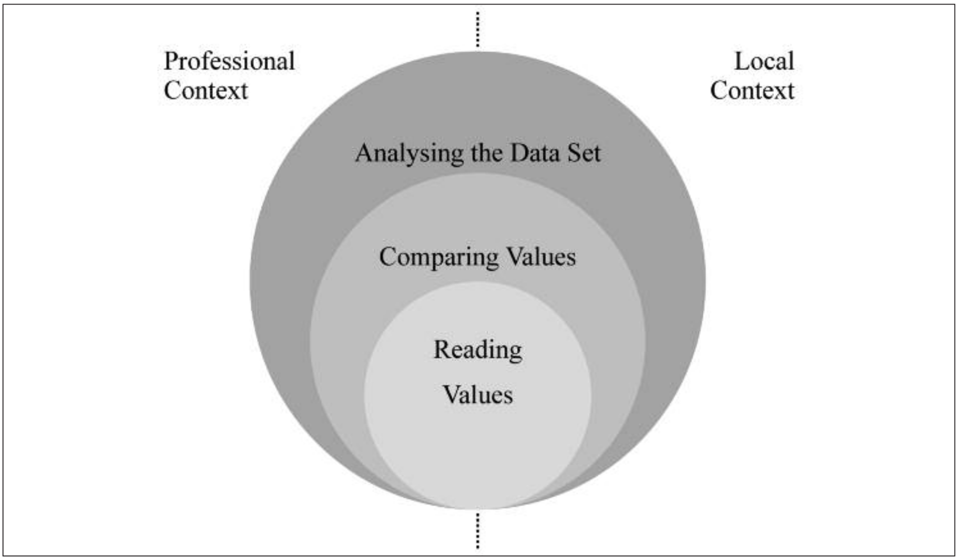


Figure 1. A framework for professional statistical literacy (Pierce & Chick, 2011, p. 633).

The *Comparing values* level involves attending to multiple aspects of the data to make direct comparisons. This is the level that applies when comparing data values, looking for trends, identifying skewed data from the shape of a boxplot, and so on. Drawing statistically valid conclusions about the data requires more than mere comparisons, however.

The highest level of technical data interpretation, Analysing the data set, involves applying relevant statistical tools to interpret the data as presented and imagining how representations might change with changes in the data. This level is characterized by attention to the data set as a whole rather than as individual values or components. This level of interpretation is required in order to understand, for example, that differences between data sets may not be statistically significant; or the effects of sample size on results.

These technical skills, however, are insufficient for making full sense of the data. Attention must also be given to the broader context from which the data arise and to querying the nature and causes of the conclusions. There are two contexts that need to be considered. The first is the professional context, which, in the case of NAPLAN data, is the knowledge that teachers should all have, as teachers, about the NAPLAN process, terminology (e.g., "bands"), the format of the tests, and the nature of the questions asked in tests, for example. Second, the local context may apply to only a subset of teachers, such as those in a particular school, and concerns their knowledge of particular characteristics of their school. So, for example, there may be a statistically significant difference between school performance data and the state's performance data that can be noted by anyone functioning at the *Analysing the data* set level. Proposing that this difference is explained by the in-school implementation of a targeted teaching program or by socio-economic differences is indicative of interpretation that attends to local context.

To illustrate these distinctions more clearly using the sort of school assessment data that are the focus of this study, consider Figure 2. This shows school performance against whole-state performance on a number of curriculum dimensions or "assessment areas".

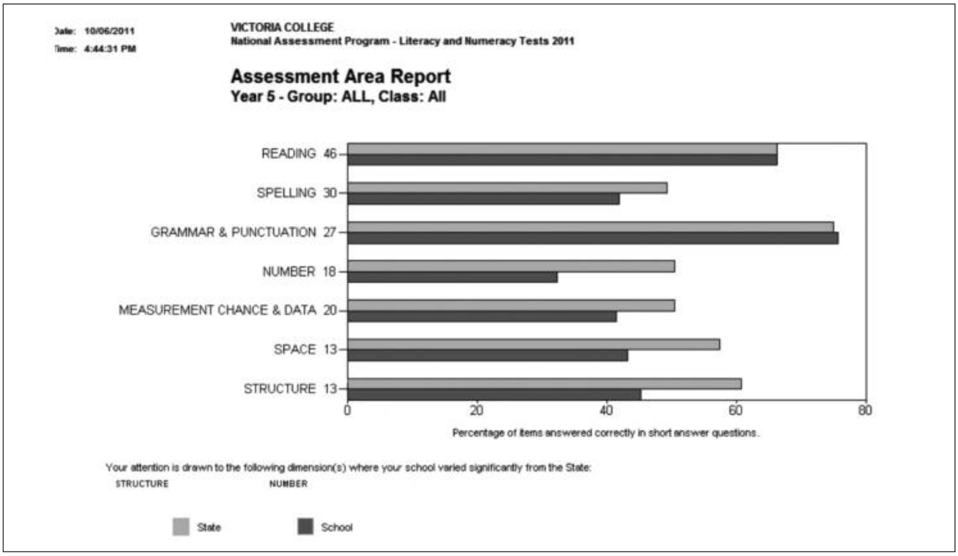


Figure 2. Demonstration school's Assessment Area Report from NAPLAN data service (similar to graphs in VCAA, 2009, p. 18).

Working at the attributes level, one can read that the school's grade 7 cohort answered about 43% of the Space items correctly. At the comparisons level, it might be noted that this was about 15 percentage points below the performance of the state cohort; and that, in fact, the school appears to have done worse than the state across many of the assessment areas, apart from "Reading" and

"Grammar & punctuation". If the reader cannot function at the analysing level it is likely that this poor performance will be interpreted inappropriately—imagine possible reactions among some parents on seeing this graph of school performance. At the analysing level, the question of statistical significance should be raised and addressed if possible. In this case, the supplied data actually point out that only the number and structure results are statistically significantly different from the state results. An outside observer may understand the impact on statistical significance of a small school cohort or a wide variation in results. Someone with knowledge and understanding at the setting and context level would recognise that, for instance, the school has many students with language backgrounds other than English, which may explain the relatively low Spelling outcome, but must also be aware that the observed difference is still not statistically significant. Such a person might wonder how the significantly poorer results in Number and Structure might combine to affect the overall construct of "numeracy" (used in other reports and graphs), and expect to see that the school might perform significantly differently from the state on this criterion.

It is also important to realize that it may be possible to have enough knowledge of the local context to try to function at the setting and context level—for example, by trying to explain observed differences in terms of socio-economic background—but that a lack of the requisite technical aspects may render such analysis incorrect if, for example, there is no understanding of whether differences are statistically significant or in keeping with the level of variation that may be expected in a small sample. Similarly, a lack of knowledge of context can actually hamper the application of technical skills, as illustrated by the fact that it is not apparent what is meant by the "13" in the assessment area category "Space 13". This makes it difficult to ascertain whether and how this information might affect any statistical interpretation. One of the standard statistical graphs used by the NAPLAN data service is the box plot (see Figure 3). Drawing on the work of others, Pfannkuch (2006) highlighted that, for box plots, attending to several critical components is essential for comparing values, prior to fully interpreting the data. She pointed out that the middle part of the data usefully characterises the group and that comparing box plots should incorporate comparison of these middle parts of the data. Components of the spread, including the interquartile range and the whiskers, allow a determination of the variation within and between box plots to give a sense of the shape of the data (e.g., if it is symmetrical or skewed). Furthermore, it may be appropriate to compare equivalent and non-equivalent key values; for example, a difference between data sets may be evident if the 25th percentile for Group 1 is above the 75th percentile of Group 2.

Watson (2012) has recently examined the history of the boxplot and its place in the curriculum. The work of Pfannkuch (2006) and Wild, Pfannkuch, Regan, and Horton (2011) provides a comprehensive analysis of the kind of reasoning that might generally be involved at the analysing data stage for box plots. Pfannkuch's report—part of a larger study—found that informal inference is a complex matter. As part of her study, five statisticians used their knowledge of formal inference in interpreting box plots to identify what is required for successful informal inference (i.e., to draw considered conclusions from the data, and understand the variation within it). Based on their experience and expertise they identified four key elements for informal inference: comparing centres (e.g., mean, median), comparing differences between centres while simultaneously taking into account variation, checking the distribution of the data (e.g., attending to its shape, presence of outliers, clusters of data values), and considering sample size effects. Each of these elements involves one or more levels of the framework for professional statistical literacy. Comparing centres, for example, involves knowing what part of the box plot represents the median (the reading data level) and then comparing centres is obviously at the comparing data level. Consideration of sample size effects requires the beginning of functioning at the analysing data level, since it is necessary to appreciate how the data may change with changes in sample size. Wild et al. give a more detailed discussion of factors that might be considered in "making the call" to identify differences between data sets depicted as box plots, and actually suggest a set of milestone tests for these differences. Pfannkuch also noted, however, that there had been no research on how teachers reason when comparing box plot distributions and no definitive accounts reporting on the process of going about drawing informal inferences.

Principals and teachers working with reports, such as those provided by the NAPLAN data service, need to be able to attend to the technical aspects and the setting and context of the data in order to make informal inferences about school assessment data. As seen in Pfannkuch's summary (2006) this is not a trivial task. Monteiro and Ainley (2007) refer to the idea of critical sense, where "a sophisticated reading of graphs involves mobilising a range of different kinds of knowledge and experience … [including] knowledge about the processes of data collection and analysis, and … knowledge of social context from which the data has been drawn" (pp. 202–203).

That such skills are difficult for teachers who work with student performance data has been confirmed by the preliminary results of the larger study, of which this work is part. Over 700 primary and secondary teachers responded to a survey made available to a random sample of approximately 2000 teachers across Victoria. Of these, the vast majority claimed they felt they could adequately interpret the data they receive, yet there were statistical skills and understanding items based on graphs like that in Figure 3 that were answered correctly by very few of these teachers (see Pierce & Chick, 2013). It is timely to examine more closely exactly what knowledge is required by teachers for making sense of the data they receive in their working context.

In the present study three demonstration reports were examined in order to identify more precisely the statistical literacy required at each level of the data interpretation hierarchy.

Methodology

The data reported in this paper were collected as the first part of a large study that considers the professional literacy required of principals and teachers, their statistical literacy skills, and factors affecting their intention to engage with the data.

The Victorian NAPLAN data service supplies assessment data for schools in Victoria. They also provide a data set for a fictitious demonstration school, "Victoria College", for demonstration purposes; and the reports on these fictitious data were analysed for this paper. Although we chose this Victorian source because it is of immediate interest to our local situation, we shall later make some reference to other types of data representation arising from other jurisdictions.

The data—summarising individual, class, school, state, and national NAPLAN results—are presented in graphs and tables, according to the type of report selected. The range of reports includes analysis of individual items on the NAPLAN tests, data on content areas (e.g., Reading; Spelling; Number; Measurement, Chance and data), and school summary data with state and national comparisons. For the purposes of this report, only the graphical representations were analysed. The most common representations shown in the Victorian NAPLAN data reports are box plots and bar graphs (see Figures 2 to 4, which use data for the demonstration school). Some of the other representations used elsewhere are discussed later in the paper.

The graphs were analysed to determine the statistical literacy skills required to interpret them. A first pass on the analysis was completed by the authors, in the company of an experienced senior mathematics teacher. All three have strong statistical literacy skills, with the second author having experience in teaching statistics and statistical consulting. We looked at each of the graphs in Figures 2 to 4, with which we were not initially familiar, and made explicit amongst ourselves the aspects that we were attending to and the skills we were using as we read and interpreted the graphs. We made notes about the skills that were required.

At the time of this initial analysis we did not have a particular framework in mind as an analytical tool, but instead merely recorded each component that we noticed. We were, however, mindful that we appeared to be seeing several levels of complexity involving statistical literacy in this process. The data interpretation hierarchy then arose out of our examination of the graphs and our later investigation of the alternative frameworks discussed earlier. At the basic level, there were the skills needed simply to read the directly accessible information on the graph (the reading values level), and then to be able to make comparisons across data values (the comparing values level). We noted that we could interpret some aspects of the graphs in light of our broader statistical knowledge about significance and sample size effects (the analysing data level), but that a full understanding of the data's meaning for this (fictitious) school relied on a knowledge of additional information about the context which we did not have but which we knew we would need (i.e., the professional and local context).

Having identified all the aspects of statistical literacy that we felt we had used in interpreting the graph, the second author, in consultation with the first author, then classified each skill under one or other levels of the data interpretation hierarchy. We were also interested to see which of the aspects identified by Pfannkuch (2006) as necessary for informal inference were required. Having categorised the statistical skills as described above, we then examined the informal inference requirements to determine if they corresponded to Pfannkuch's analysis. Our purpose in this analysis is not to do the informal inference or to teach how to do it, but rather to investigate the statistical literacy skills required.

Results

The results that follow present the statistical literacy required for interpreting the data in three different graphical reports produced by the NAPLAN data service. In each case, the results are classified according to levels of the professional statistical literacy framework hierarchy.

Graph 1: Assessment Area Report

The first graph, shown in Figure 2, is an Assessment Area Report that presents data about different components of literacy and numeracy. The skills required to read this graph are outlined in Table 2. Unlike box plots, where the box and whiskers capture several points of information, the only part of the "bar" of a bar graph that provides information is the endpoint. Although Pfannkuch's analysis of the skills necessary to interpret box plots is not applicable here, there are parallel skills required, such as comparison of endpoints (instead of centres), comparison of differences between endpoints (with a need to take into account variation), and a need to consider sample size effects.

In terms of analysing the data set, some understanding of statistical significance is also necessary, in order to know which differences "matter". A more complete understanding of this would be demonstrated by attending to the local and professional contexts, which requires the graph reader to understand features of the graph in context. This might involve explaining why a school "varied significantly from the State", and understanding how the assessment area data are generated (what kind of tests were used). A combination of knowledge of technical aspects and of context is necessary to be able to recognise that although the horizontal axis claims to show the "percentage of items answered correctly in short answer questions" it seems to mean the average percentage of items answered correctly by the students in the school. This use of percentage is different from the way it is used in some of the other NAPLAN tables. Finally, knowledge of context is also required to understand the meaning of the numbers on the end of the labels for the subjects. These numbers may or may not have a statistical impact. If they are merely labels, then knowing this is only knowledge at the reading data level—like knowing what the acronyms represent. On the other hand, if they indicate the numbers of questions involved in the tests, then this may influence the interpretation of the data at the higher levels.

Table 2
Data Interpretation Skills Required to Interpret Figure 2

Level	Skills
Reading Values	Read the scale (e.g., note that it does not go to 100)
	Read individual data points as indicated by bar values
	Read labels
	Read the legend
	Understand that the left axis is categorical, not numerical
Comparing values	Compare the magnitude of numbers
	Know how to compare values (proportional/absolute comparisons)
Level	Skills
Analysing data	Consider the number of questions when comparing values
	Know which numbers to compare for interpretative purposes Reconcile how big the difference is between the state and the class in terms of real numbers, and hence
	Determine when differences are numerically important (perhaps not "statistically significant" in a formal sense)
	Understand that it is not appropriate to draw conclusions about differences between small groups and state or national values
	Understand about variation and its implications
Local and professional context	Understand features of this graph in context (such as the structure of NAPLAN tests and how they provide data for each of the bars)
	Determine when differences are educationally significant

Graph 2: School Summary Report

The second graph, shown in Figure 3, is the School Summary Report. This presents data about the distribution of the performances of students from a school on different components of literacy and numeracy and compares these with state and national outcomes.

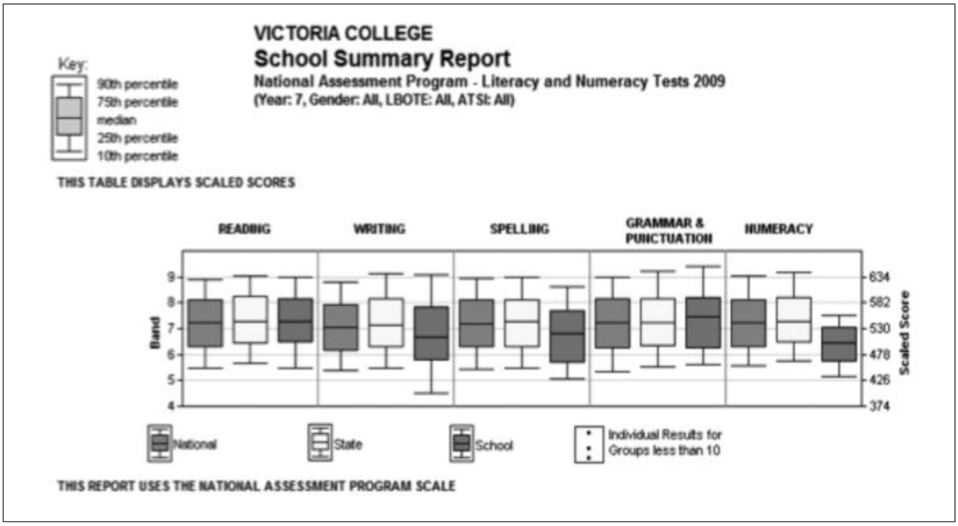


Figure 3. Demonstration school's School Summary Report from NAPLAN data service (similar to VCAA, 2009, p. 16).

The skills required to read this graph are outlined in Table 3. The use of a box plot provides more information to the user but adds to the complexity of interpretation. The graph reader now has to attend to multiple facets of the boxes in order to make meaningful comparisons and understand the data. These include the skills mentioned by Pfannkuch (2006): comparing centres (medians); attending to distribution or shape (i.e., the components of the box and the whiskers); and considering sample size effects (e.g., to determine whether differences between school and state data are significant).

In this case, the demands of the comparisons level are heavier, because of the complex information presented in a box. The school's Writing data in Figure 3, for example, is slightly skewed, with the top half of the students' score more widely spread than the lower half.

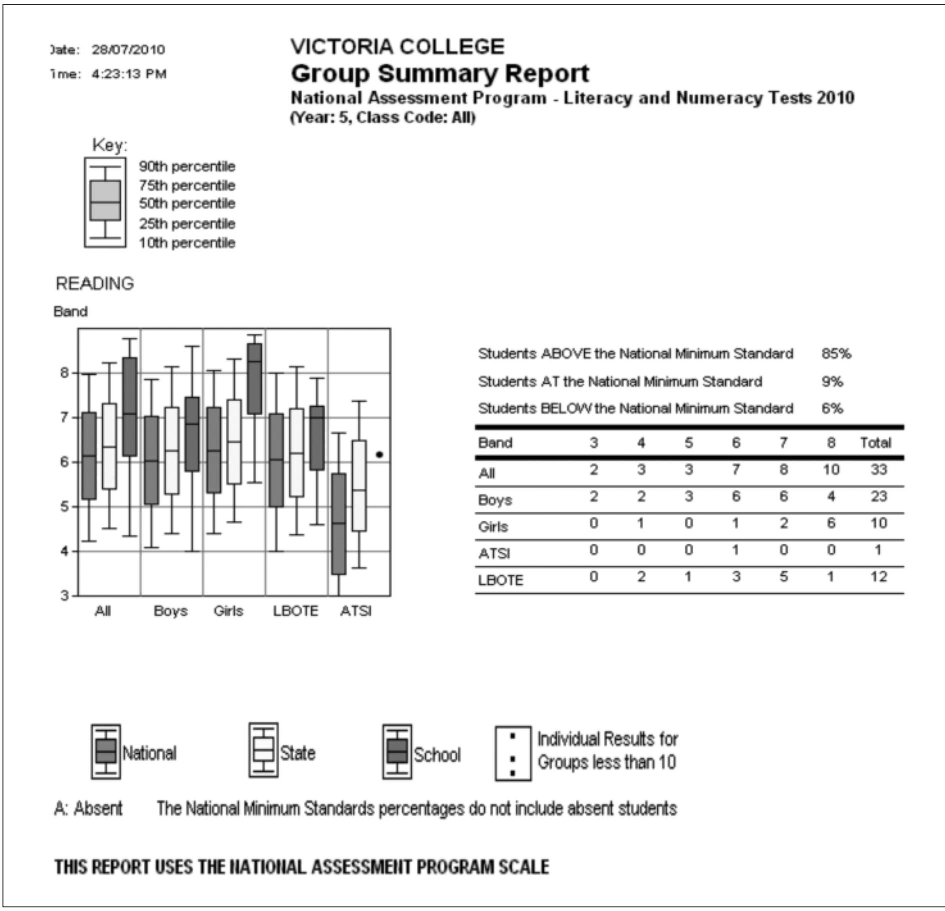
Further, given that Figure 2 has Number, Measurement Chance & Data, and Space as separate components, it is important to understand how the "numeracy" construct arises from these. Fully interpreting all the implications of this, both statistically and for the school, requires functioning at the level of analysing the data set with significant attention to the context.

Table 3
Data Interpretation Skills Required to Interpret Figure 3

Level	Skills
Reading Values	Read the key (to know this box plot differs from textbook plots)
	Read values against the band-axis
	Read labels
	Read median value, including when the median and one or more of the quartile values coincide (see Spelling)
	Read percentile values
	Read the value for a whisker including when one when falls on the boundary of the graph (see Writing)
	Understand that the horizontal width of the boxes is irrelevant
	Understand that results for groups less than 10 are displayed as points
Comparing values	Understand the meaning of median
	Understand the meaning of percentiles
	Describe the shape of distributions i.e., identify a distribution as skewed or symmetrical
	Understand the implication for the shape of the distribution of the data when the median is very close to or at one end of the box
	Understand how box plot features indicate the shape of the data
	Compare shapes of distributions
	Know how to compare both within and between box plots (involving multiple comparisons)
	Comparing the magnitude of numbers
Analysing data	Understand and interpret the absence of outliers
	Determine when differences are numerically significant (perhaps not "statistically significant" in a formal sense)
	Understand that it is not appropriate to draw conclusions about differences between small groups and state or national values
	Implicit understanding about variation and its implications
Local and professional context	Know how many children at their school are represented by each component of the box plot
	Determine when differences are educationally significant
	Understand features of this graph in context (e.g., derivation of the "numeracy" construct)

Graph 3: Group Summary Report

The final graph to be considered in detail, Figure 4, is a Group Summary Report. This graph allows comparisons among subgroups of students from a school (e.g., by gender or language background) with state and national results.



Band	3	4	5	6	7	8	Total
All	2	3	3	7	8	10	33
Boys	2	2	3	6	6	4	23
Girls	0	1	0	1	2	6	10
ATSI	0	0	0	1	0	0	1
LBOTE	0	2	1	3	5	1	12

Figure 4. Demonstration school's *Group Summary Report* from NAPLAN data service (similar to VCAA, 2009, p. 17).

The skills required to read this graph are outlined in Table 4. The comments made for Graph 2 about interpreting box plots are relevant here. Interpreting this graphic requires the same skills and understanding as Figure 3, together with the additional abilities described in Table 4. That is, the skills listed in Table 4, required to interpret Figure 4, are in addition to the relevant box plot skills described in Table 3 above.

Table 4
Data Interpretation Skills Required to Interpret Figure 4

Level	Skills
Reading Values	Understand the acronyms (e.g., ATSI, LBOTE)*
	Understand what the subsets mean*
	Understand symbols used (e.g., A is used for absent)*
	Understand that with small numbers individual students may actually be revealed (see ATSI)
Comparing values	Know how to link information in the table to the box plot
	Other skills as listed in Table 3
Analysing data	Understand that you cannot "add" the boys' and girls' box plots
	Read and interpret the table using skills as listed in Table 3
Local and professional context	Understand how the size of the subsets will affect the overall results (e.g., could the good performance of the class be due to a small number of high-performing girls—this case is not shown)
	Other skills as listed in Table 3

* This is contextual information but at a factual level: no statistical interpretation or critical analysis is involved.

Discussion

The graphs provided by the NAPLAN data service largely comply with the guidelines for good graphics suggested above by Tufte (2001). They are presented simply, without unnecessary distracting ornamentation. It should be noted, however, that box plots with whiskers extended to the 10th and 90th percentiles are not standard (see, for example, Aczel & Sounderpandian, 2006; Howell, 2002; Moore & McCabe, 1993; noting that there is also some variation in these references about how to treat extreme outliers). The reader must also be aware of the range of the scale used in each report. Nevertheless, despite this variation —which, in any case, is clearly indicated by the key supplied with the data—the graphs are good examples of appropriate summary data representation.

The statistical literacy required to work at the reading data level for these graphics has three main components. First, the reader needs to be aware of and understand the key components and labels on the graph: title, scale labels, and scale. In Figure 2, for example, the left hand side scale is categorical whereas the horizontal scale is numeric. Second, the reader needs to note the legend or key in order to know what values are being read. For example, in Figure 3 a reader must note which bar refers to the State and which to the school. In the same figure the key to the box plots indicates that these are not drawn in the way that is typically presented in school and statistics text books. As noted above, in the NAPLAN data service box plots the whiskers extend to the 10th and 90th percentiles respectively, and outliers are not shown. This leads to the third element of statistical literacy required by the reader: knowledge and understanding of the values that may be read from each type of graph. For example, in Figure 1 the only data values that can be read from this graph are the values associated with the endpoint of each bar. In Figure 3 there are five key values that may be read from each box plot, associated with the 10th, 25th, 50th, 75th, and 90th percentiles.

Comparing data requires readers to have further understanding of the graphs' components. For the bar graphs they need to consider the absolute and proportional differences, and for box plots they need to be able to recognise and understand the implications of the shape of the distributions. Importantly, making comparisons involves paying attention to not only the location of the centre of the distribution, indicated on the box plot by the median (50th percentile), but also the spread, indicating the variability in results within the school group. In Figure 4, for instance, comparing the data for boys and girls shows that the centre of the distribution of girls' results was above that of the boys', but the variation in the boys' results was greater (ignoring the invisible outliers).

Fully interpreting the data requires more than just an understanding of the key components and structure of the various graphs. It demands an understanding of both the types of claims appropriate to make based on informal inference, and also the context within which to interpret these claims—such as the school, the students, and the tests used to collect the data. This is where the critical thinking that is at the heart of statistical literacy must be applied. The graph reader, at this stage, is making inferences—albeit informal—based on the data. In so doing, functioning at the higher analysis level is necessary.

Furthermore, if such data reports are to be of value for decision making and planning, then the school principal or teacher must consider the implications of these numbers within the context of their school, thus highlighting the importance of the local and professional contexts. For example, in the Assessment Area Report (Figure 2) the reader might think "We scored below the state average in Measurement Chance & Data. Is this something we need to address in our teaching or is there some other explanation for this result?" Other explanations may relate to group size and natural variation or to local events that impacted on the testing.

Interpreting the box plots of Figure 4, as a second example, means paying attention to the differences between groups relative to the spread or variability within each group, leading to questions about whether the difference in performance of the LBOTE students (with a Language Background Other Than English) in comparison with the whole school cohort is significant. It also requires recognition that few, if any, conclusions can be drawn about the ATSI (Aboriginal and Torres Strait Islander) students as a school group because only one ATSI student was involved in the testing.

The four key elements identified by Pfannkuch (2006) and her statistician colleagues were, indeed, quite critical for a full understanding of the data, although in our analysis we applied the milestone heuristic suggested by Wild et al. (2011) only informally. In order to determine how the school's results compared with state and national results—or to conduct any of the other important comparisons among the results represented as box plots in Figures 3 and 4 above—it was necessary to attend not only to the centres (medians) of the box plots, but also to the distribution indicated by the components of the boxes and the whiskers. Further consideration of variation was necessary in order to determine the magnitude and significance of any differences.

As Wild et al. point out, the influence of sample size on the significance of differences is particularly critical for informal inference, and hence for making valid conclusions about when the results indicate causes for concern or satisfaction. Without these skills it is likely that teachers will focus incorrectly on differences that are not statistically significant or dismiss those that are.

Although we have only considered some of the representations that are used in Victoria, where there is a strong focus on box plots, it should be noted that there are many other kinds of representations used in different states, together with those used in data from sources like PISA (e.g., Organisation for Economic Co-operation and Development, n.d.). Taking just two examples from New South Wales documents (Figures 5 and 6), it is clear that all levels of the framework for professional statistical literacy still apply.

In Figure 5 there are comparisons that can be made among school, state, and national values, but in order to understand the data set fully there is a need to get a more holistic sense of each year's data, and also the trend data.

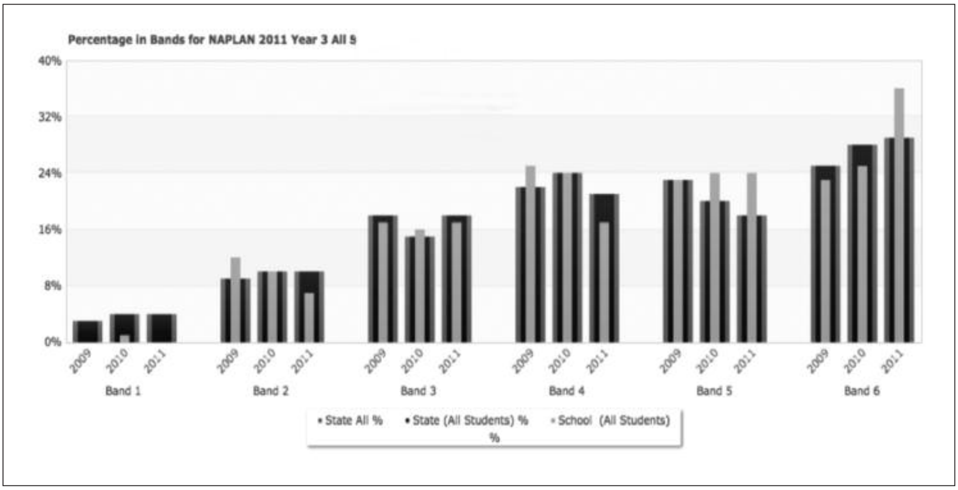


Figure 5. NSW NAPLAN data graph (NSWE&C, 2011, p. 1).

The scatter plot in Figure 6 (below) allows the reading of individual student values, but comparisons are now made more complex because two variables are involved. More detailed analysis is needed to get a sense of class performance overall, in comparison with the State, and to identify any outlying student values.

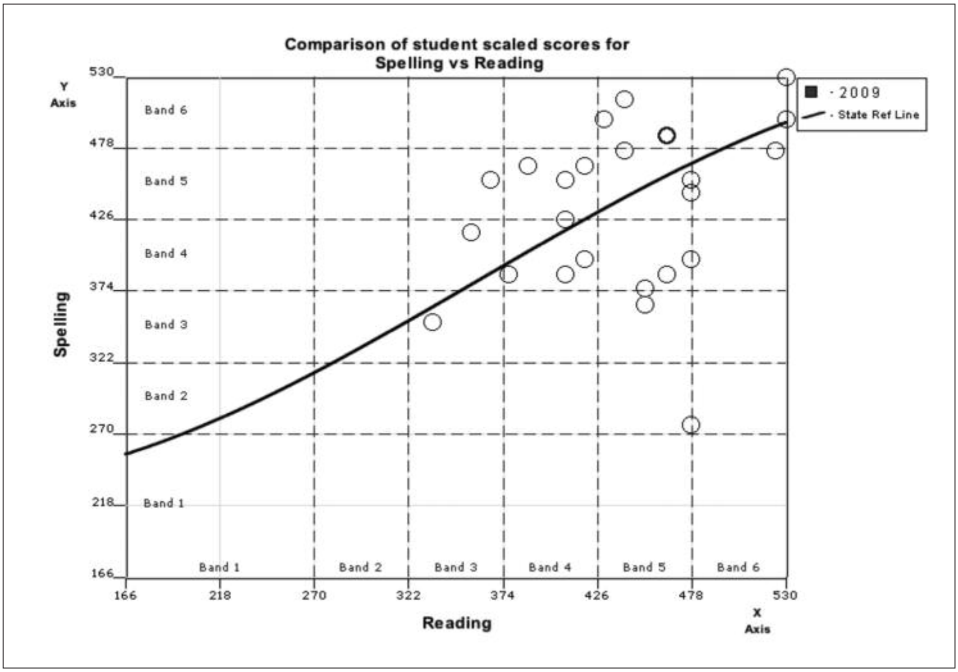


Figure 6. NSW NAPLAN data scatter plot (NSWDET, 2009, p. 45)

Conclusion

The analysis of the graphs from the NAPLAN data service has highlighted that complex critical thinking is required to make sense of the data, particularly in the light of their context. It should not be assumed that everything that the data might reveal is evident just through the graphics alone, nor that someone with limited statistical literacy could read and make sense of these reports. The picture may, indeed, be worth a thousand words, but its subtle nuances require a careful reading to get the full story. Having said this, neither is it the case that the skills required to read the graphs are particularly difficult. The user needs to be "fluent" in reading the elementary aspects, which is essentially the scope of the reading and comparing values levels: the capacity to attend to scale and labels, for example, together with reading data values as represented on the graph and making comparisons among them. These skills, if not already understood, are relatively straightforward and can be taught at a simple, hands-on, professional development session (see, for example, Pierce & Chick, 2012). It is important that a basic level of professional statistical literacy is included in pre-service education for all teachers but especially for mathematics teachers who reasonably may be expected to provide statistical literacy support within their schools.

More complex are those skills needed to analyse the data as a whole and attend to the implications of context. This requires some understanding of informal inference, and the statistical literacy to know what level of importance to place on the various results. Principals and teachers need to be able to determine which differences should be given attention, and then use knowledge of their own context to identify appropriate responses and strategies, or provide alternative explanations for the outcomes. This demands a capacity to question the data, an awareness of sampling issues, and the kind of relational thinking that makes it possible to keep track of how one variable may affect another in a variety of circumstances. These skills are likely to be more difficult to develop, but teachers' own familiarity with their school contexts might mean that it is possible to develop scenarios and associated graphs that allow such a level of thinking to be fostered.

Finally, it is important to consider the graphs themselves and to ask whether or not they present data in the best possible format. The graphical reports from the VCAA are constrained to two formats—compared bar graphs, and compared box plots—and are consistently presented. (For instance, apart from width there is no change in the way the box plots are presented between Figures 3 and 4). Although the box plots are not standard, a key is provided and, as mentioned, the key values on the box and whiskers do not change between graphs. Missing from the representation, which may be of interest to teachers, are the outlying values, although teachers also have access to individual student data. One of the critical requirements for functioning at the "analysing the data set" level is to appreciate the distribution of the data values among the box and whisker components of the box plot: that there are, for example, the same number of values in both the 25th to 50th and the 50th to 75th quartiles regardless of the fact that the boxes representing each quartile may be of different physical lengths.

In conclusion, the graphical representations convey complex information and there is a need for statistical literacy and critical thinking in order to understand and make best use of them. Interpreting graphics also requires attention to information about the professional and local context. To what extent teachers currently have sufficient statistical literacy to make sense of these data is an important question. It is also vital to examine how to help them gain such skills, otherwise there is limited likelihood of these reports producing educational improvements anticipated by governments.

This also has implications for teacher education programs. Since the understanding of such data has potential for improving teaching, it is essential that the specific statistical literacy requirements are covered in pre-service teacher training. This has been addressed at some institutions (e.g., Watson, 2011), but with a greater knowledge of what, exactly, is required it should be possible to target appropriate statistical skills and high-level critical thinking about data. The framework for professional statistical literacy allows a hierarchical consideration of the kinds of analyses that can be made and offers scope for teachers to build up their understanding of data sets by working up through the levels in turn.

Acknowledgements

The authors would like to thank Roger Wander for input into identifying the statistical literacy skills required to read the graphs; the Melbourne Graduate School of Education for a small grant funding this research; and Jane Watson for helpful discussions about the framework and this manuscript. During the life of this paper the first author worked at the University of Melbourne and the University of Tasmania. An earlier version of this paper was a technical report published as a working paper:
http://www.edfac.unimelb.edu.au/sme/research/workingpapers.html. This work was part of a larger project funded by the Australian Research Council (LP100100388), the Victorian Curriculum Assessment Authority and the Victorian Department of Education and Early Childhood Development.

References

Aczel, A. D., & Sounderpandian, J. (2006). *Complete business statistics* (6th ed.). New York: McGraw Hill.

Australian Education Council (1991). *A national statement on mathematics for Australian schools*. Carlton, VIC: Curriculum Corporation.

Baker, E. (2007). Presidential address: The end(s) of testing. *Educational Researcher, 36*(6), 309–317.

Ben-Zvi, D., & Garfield, J. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Board of Studies (1995). *Curriculum and standards framework: Mathematics*. Melbourne: Author.

Children, Schools and Family Committee (2008). *House of Commons: Select Committee on Children, Schools and Families. Third Report*. Retrieved: April 25, 2009, from http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/169/16902.htm

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.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury/Thomson Learning.

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA, 2007). *Measurement framework for national key performance measures*. Retrieved April 25, 2009, from http://www.mceetya.edu.au/verve/_resources/2007_National_Measurement_Framework.pdf

Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA, n.d.). *Benefits of participating in national assessments*. Retrieved May 8, 2009, from http://www.mceetya.edu.au/verve/_resources/Benefits_of_participation_in_national_ assessments1.pdf

Monteiro, C., & Ainley, J. (2007). Investigating the interpretation of media graphs among student teachers. *International Electronic Journal of Mathematics Education, 2*(1), 187–207

Moore, D. M., & McCabe, G. P. (1993). *Introduction to the practice of statistics* (2nd ed.). New York: W. H. Freeman.

New South Wales Department of Education and Training (NSWDET, 2009). *SMART NAPLAN data software help manual*. Retrieved September 14, 2011, from http://www.schools.nsw.edu.au/media/downloads/schoolsweb/learning/yr7_12 assessments/smart/help/installation/helpmanual.doc

New South Wales Education and Community (NSWE&C, 2011). *What's new in 2011 NAPLAN data?* Retrieved September 14, 2011, from http://www.schools.nsw.edu.au/media/downloads/schoolsweb/learning/yr7_12assessments/smart/what s_new_2011nap.pdf

Organisation for Economic Co-operation and Development. (n.d.) *PISA 2009 country profiles*. Retrieved September 16, 2011, from http://stats.oecd.org/PISA2009Profiles/

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* (pp. 631–639). Adelaide, SA: AAMT.

Pierce, R., & Chick, H. (2012). Improving teachers' statistical literacy. *The 8th Australian Conference on Teaching Statistics*, OZCOTS 2012. Retrieved September 14, 2011, from http://opax.swin.edu.au/~3420701/OZCOTS2012/OZCOTS2012_PierceR_Final_paper.pdf

Pierce, R., & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal. 25*, 189-205. DOI 10.1007/s13394-012-0046-3.

Pierce, R., Chick, H. L., & Gordon, I. (2013). Teachers' perceptions of the factors influencing their engagement with statistical reports on student achievement data. *Australian Journal of Education, 57*, 237–255.

Pierce, R., Chick, H, Watson, J., Dalton, M., & Les, M. (2012). Trialling a professional statistical literacy hierarchy for teachers. 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* (pp. 602–609). Adelaide, SA: 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.

Tufte E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press.

Victorian Curriculum and Assessment Authority (VCAA, 2008). *Victorian essential learning standards. Discipline-based learning strand: Mathematics*. Melbourne: Author.

Victorian Curriculum and Assessment Authority (VCAA, 2009). *NAPLAN 2009 Reporting guide*. Melbourne: Author. Retrieved 25 Aug 2013 from https://naplands.vcaa.vic.edu.au/Tutorial/WebHelp/Guides/reportingguide09_3579.pdf

Victorian Curriculum and Assessment Authority (VCAA, n.d.). *National Assessment Program Literacy and Numeracy Testing* http://www.vcaa.vic.edu.au/prep10/naplan/index.html

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

Watson, J. M. (2011). *Personal and professional numeracy*. Subject notes. Hobart, TAS: University of Tasmania.

Watson, J. M. (2012). Boxplots in the Australian Curriculum. *Australian Mathematics Teacher, 63*(3), 3–11.

Wild, C. J., Pfannkuch, M., Regan, M., & Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society, 174* (Part 2), 247–295.

Authors

Helen Chick, University of Tasmania, Private Bag 66, Hobart, TAS 7001, Australia. Email: helen.chick@utas.edu.au

Robyn Pierce, University of Melbourne, 234 Queensberry St, Melbourne, VIC 3010, Australia. Email: r.pierce@unimelb.edu.au