SUFFICIENTLY ASSESSING TEACHERS' STATISTICAL LITERACY

Helen Chick[1], Robyn Pierce[2] and Roger Wander[2]
[1]University of Tasmania, Tasmania, Australia
[2]University of Melbourne, Victoria, Australia
Helen.Chick@utas.edu.au

Teachers are inundated with data, including reports from mandated testing. Interpreting these requires professional statistical literacy, involving technical statistical knowledge and the capacity to interpret data meaningfully in the context of teachers' professional work. For a study investigating teachers' data use, the authors developed instruments for assessing statistical literacy based on typical reports and the critical components of statistical literacy required to understand them. Designing an instrument that assessed workplace-relevant statistical knowledge, and did not take too long to complete, involved design choices through a process of beginning with an initial pen-and-paper design and then designing an online version with automated marking. This paper will examine some of these design issues and their impact on assessing teachers' statistical literacy prior to providing appropriate professional learning.

BACKGROUND

Schools in many countries collect extensive data on their students' learning, including from national testing, for use as an evidence base to inform decisions about planning and teaching. Interpreting the statistical reports prepared by education authorities requires professional statistical literacy, involving technical statistical knowledge and the capacity to interpret data meaningfully. The authors were part of a research team that designed instruments for assessing statistical literacy to inform training on school data interpretation. They used a hierarchy for professional statistical literacy to examine typical school system reports, identify key knowledge and skills required to understand this data, and then construct items that focused on critical components of statistical literacy. Designing the instrument involved a number of considerations, both practical and statistical; this paper examines some of these design issues and their impact on assessing teachers' statistical literacy prior to providing appropriate professional learning.

KNOWLEDGE REQUIRED TO UNDERSTAND TYPICAL REPORTS

Teachers require statistical literacy to make good use of quantitative data, which means understanding key statistical ideas along with a positive disposition towards the use of such data (Ben-Zvi & Garfield, 2004; Gal, 2002; Watson, 2006). Professional statistical literacy requires understanding the professional context as well, which for Australian teachers includes knowing about the testing scales used and the curriculum. In addition, teachers need to interpret data in light of their local context, since individual school and classroom factors may affect the data.

A study of common key report formats for teachers in the Australian state of Victoria reinforced the elements of statistical literacy hierarchies proposed by Curcio (1987), Shaughnessy (2007) and Watson, (2006), and synthesized in a framework for professional statistical literacy (e.g., Pierce & Chick, 2013). This hierarchy provided a framework for designing appropriate assessment items. Interpreting reports involves consideration of data at multiple levels (as indicated by the nested circles in Figure 1). The lowest level, *reading values*, involves the capacity to read specific data points on the graph or table, together with graph features such as key and scale. The second level, *comparing values*, requires attention across multiple data values or across representations, and may involve determining differences, early informal inference, and simple statistical tools. The third level, *analysing the data set*, interprets the data as a whole, acknowledging variation, trends over time or other variables, and significance. In addition to applying these statistical skills, full interpretation requires knowledge of professional and local contexts, such as national testing terminology and socio-economic status of the students involved.

ASSESSMENT OF STATISTICAL LITERACY

In the project reported in this paper we needed to assess teachers' statistical literacy in order to develop an evidence-based professional learning program. Data collection—and hence the

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

design of instruments for assessing statistical literacy—had two phases: first, face-to-face with small numbers of teachers and a paper-based instrument, and, second, online with a large number of teachers. In this paper we present the issues and constraints encountered in the design process, and highlight the complex interactions arising between our research purpose of assessing statistical literacy, and the realities of actually undertaking this research in different environments. Our focus will be on design issues and their implications for assessing the statistical literacy of teachers.

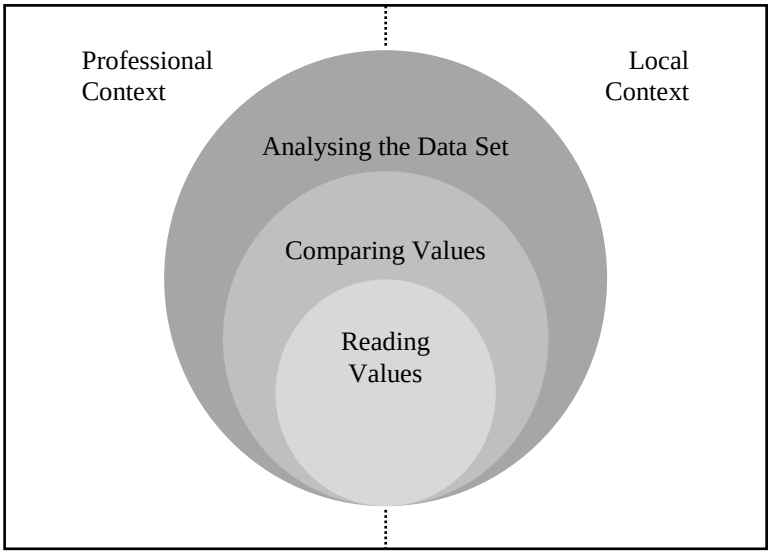

Figure 1. A framework for professional statistical literacy (Pierce & Chick, 2013, p. 193)

Background for the Two Phases of the Study

With regard to assessing statistical literacy, the study involved two phases. The first phase involved a paper-based survey followed by a focus group discussion. It was conducted with groups of seven or eight teachers, from each of ten primary and ten secondary schools selected through cluster sampling from five Victorian education department regions representing a range of demographic and school types. A significant purpose of this first phase was to collect data to inform the design of the instrument for the second phase, and to provide an indication of areas of concern for teachers' statistical literacy. The second phase used an online instrument to measure teachers' professional statistical literacy, to confirm that we had identified aspects with which the teacher workforce struggled, and that could be targeted in future professional learning programs. The instrument was used by over 700 teachers (volunteers from the invited pool of all teachers in a set of randomly selected government schools; see Pierce et al, 2013), and was designed so it could be computer marked, at least in part.

In addition to examining teachers' statistical knowledge and skills as required for interpreting student assessment data, the surveys also addressed the teachers' attitudes and predispositions to engage with systemic school statistical reports, based on the Theory of Planned Behavior (Ajzen, 1991). This was targeted using Likert items in both the paper and online instruments, but was also made evident through some open response items. This paper's main focus, however, is on assessing aspects of statistical skills and knowledge. We will examine factors that affected the design of the instruments, before considering a set of items in more detail.

Factors Associated with Teacher Attitudes

The assessment instruments were intended for all types of teachers, not just school data specialists or principals, on the assumption that there is a governmental expectation that all teachers need to be concerned about the results of literacy and numeracy testing that give rise to the reports that are our focus. We knew that some teachers would have extensive experience in interpreting the statistical representations, others very little or no experience in engaging with their school's data, and that some teachers whose core teaching or age-level focus was removed from the numeracy and literacy testing regime may not be as engaged in the assessment. These issues affected what

could be assessed and careful choices had to be made regarding the number, type, and focus of the items presented. Despite this, we encountered individual teachers in the focus groups whose disengagement from the data was obvious. This added to our impetus to make the online assessment quite short (discussed later). We also endeavoured to make the items as relevant as possible given that we were using generic examples; the ideal would have been to use their school's own data, although these may not have exemplified all key statistical literacy issues.

In addition to this, we anticipated—and confirmed in focus group discussions—that some teachers would experience "test anxiety" or "mathematics anxiety" associated with responding to a statistical knowledge assessment instrument. To address this, we included a few items that—after completing the first phase—we knew would be easy to answer, such as reading a value from a table or graph. There were teachers who did not complete the online instrument, perhaps because it was undertaken in the absence of the supervision that occurred in the paper testing, although we also hoped that the online environment gave some the freedom to work more slowly, to look at items many times, and to work without the stress of someone observing.

Factors Associated with Physical and/or External Requirements

The paper-based instrument was used in an extended session with teachers that included the focus group discussion and so teachers had about an hour to complete it. This meant that we could include eight extended items on statistical literacy. In addition, we could ask open-response questions, including some that allowed teachers to highlight aspects of the representations and comment upon them (see later). Some of these open responses gave us access to the teachers' own statistical language for describing and interpreting data, which provided useful information about their understanding and informed activities in later professional learning sessions. The presence of the researchers in the room as teachers completed the instrument meant that most teachers attempted most questions, although in some cases the selected teachers were called away during the session to deal with other school matters and did not complete the task.

The online version was intended to be used in the teachers' own time, and we knew that a shorter collection of items would increase the likelihood of teachers choosing to complete them. Prior experience with online surveys suggested that respondents would be unlikely to spend more than 20-30 minutes on the task. The selection of questions for the shorter instrument is discussed in the next section. The online format had implications for the presentation of tasks, owing to the large size of some of the required graphics and the available screen space. Answering some of the questions required scrolling up and down to relate the questions to various graphic features before teachers could enter their response. As the online version was intended for a larger sample most questions were designed to be multiple choice or similar for ease of marking, although there were some open response prompts requesting explanations for choices. The wording of some of the options in the closed response items was informed by the responses in phase one. The need for mainly closed items, and the shorter structure of the online assessment decreased the scope and richness of the questions, but made it feasible to collect and assess evidence from a large cohort.

Factors Associated with the Statistical Literacy Construct

Most major items on both forms of the instrument involved presenting teachers with a statistical representation, usually graphical, like those they receive from data providers. The framework for professional statistical literacy informed the structure of the associated prompts, often starting with a request to read graphed or tabulated values, and subsequently requiring increasing comparison and analysis, and interpretation of context-specific information.

Of the eight main items on the paper instrument for assessing teachers' statistical literacy, two related to general statistical ideas (measures of central tendency, interpreting a bar graph). Each of the remaining six questions used a statistical representation (a graphic often accompanied by a table of data) as seen in the data reports sent to schools after national testing had been completed. To make the items relevant to both primary and secondary teachers, the national testing questions were focused on Years 5 and 7. The associated questions depended on the kind of representation, and targeted known misconceptions, capacity to make sense of the data, and teachers' views about the usefulness and clarity of the reports (see next section for more details).

In reducing the number of statistical literacy questions for the online instrument, there were three sets of related items associated with three main graphical prompts: two based on representations used in the data reports, and one set of graphs intended to address more general statistical literacy skills. The data report items were developed from corresponding items from the paper instrument, adapted to suit the online environment and to take into account the data already received from the first phase of the study. So, for example, a couple of straightforward items to give teachers confidence to tackle the remainder of the instrument were retained, but we eliminated some questions for which initial data showed that teachers had almost universal understanding. In addition, a new item was devised based on the language used by teachers (see next section).

The box-plot is the main graphic by which results are conveyed in Victoria, because it allows comparison of distributions, and so our questions centred on teachers' ability to interpret data in this particular representation. One question focused on the frequency/density misconception (where it is thought that the number of students in a quartile is proportional to its "length"), while another explored teachers' capacity to interpret the outlying results, influenced by a significant feature of Victoria's box-plots in which the whiskers terminate at the 10^{th} and 90^{th} percentiles. Though this was explicit in the key given for this graphic, we wanted to see if teachers grasped the consequence that 20% of any cohort's assessment results are missing from the representation.

One of the tensions experienced in designing the assessment items arose from the quite specific statistical knowledge requirements of professional statistical literacy. Because the instrument was designed for Victorian teachers it needed to incorporate the particular representations that were prevalent in Victoria, which meant a focus on box-plots. In other states, even though the professional context is the same and teachers also receive reports about student assessment data, the representations and form of the data are often different, which makes the specific statistical literacy demands different (e.g., understanding of confidence intervals may be required). In designing an instrument to assess the statistical literacy of teachers, especially in the case of the shorter online instrument, it was impossible to address the broader construct of statistical literacy in anything other than a brief way; it was even difficult to design an instrument that assessed all the aspects of professional statistical literacy used within this particular jurisdiction's set of reports. We devised an item that required teachers to determine which of two frequency histograms might represent a school's data, and also required them to match a histogram with an associated box-plot. This required them to attend to the shape of a distribution and how that might arise in practice and appear in different representations. The results of the first phase had revealed teacher confusion surrounding the role of a box-plot in describing a distribution. Some teachers had difficulty with the often asymmetric features of a smaller sample, such as their school results in comparison to the more normally distributed state and national results. The extra item allowed the more general statistical literacy issue of understanding distributions to be explored.

ILLUSTRATIVE EXAMPLE

A sample graph and table from a typical school data report (shown in part in Figure 2) was used as a stimulus for items on both the paper and the online versions of the instruments. In the paper version teachers were asked to identify what was represented by the yellow/cream graphs, which required them to read the key. This was readily answered; in the online version this initial question was changed to a multiple choice question associated with reading what number of students sat the spelling test, thus testing teachers' capacity to operate at the reading data level of the professional statistical literacy framework. Both versions of the instrument then required teachers to identify the band level above which 50% of the students had performed in numeracy. Again, this required more reading of the data, but with a need to understand that the box-plot shows the median, and, furthermore, the need to apply some professional contextual knowledge that the bands lie on an ordinal scale that only takes whole number values.

The paper instrument continued with a question that asked teachers to consider the writing and spelling results and identify a similarity and a difference in the results. The responses to this pair of open questions contributed to the development of a set of four closed questions for the online version, shown below, to which the teachers could respond "Definitely true", "Definitely false", "Not enough information", or "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

Both versions of the instrument concluded by asking teachers to identify the school's weakest area (an open question on the paper instrument, and a multiple choice item on the online instrument), and provided space to give reasons for the choice. Allowing an open response enabled us to see what statistical tools and interpretation were brought to bear in making their judgment.

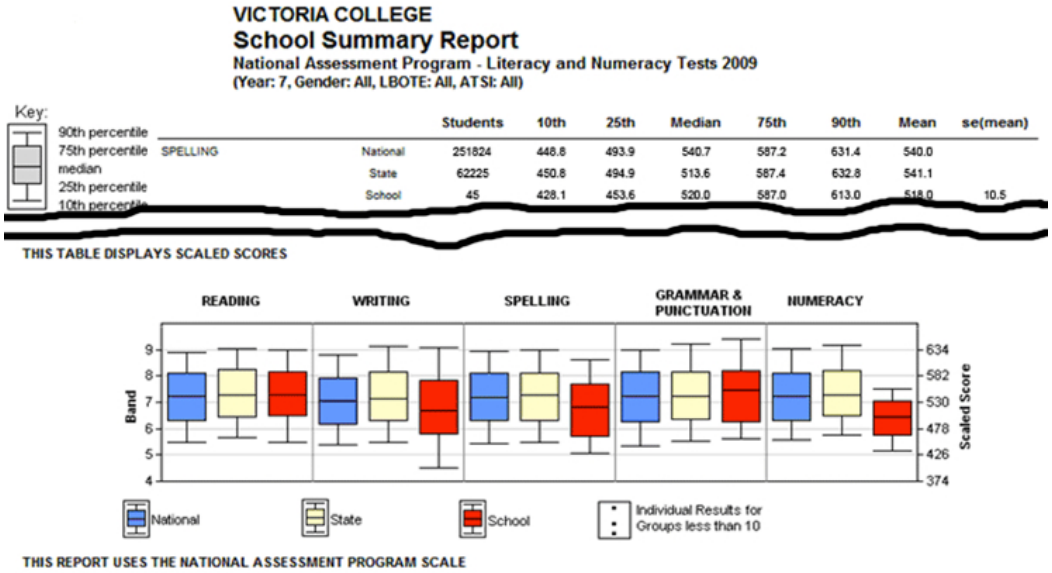

Figure 2. The box-plots and part of the associated table used within a school data report. The full graphic was used as a stimulus for assessing teachers' statistical literacy.

The paper version of the instrument also asked teachers to highlight aspects of the report that they felt were of use to them as a teacher and to indicate reasons for this; similarly, they were asked to identify aspects of the report which were viewed as being hard to make sense of and to annotate these to indicate what may cause difficulties and why. These two questions gave insight into teachers' attitudes towards the data types, but it was deemed too difficult to adapt these questions for the online instrument. both in terms of practical implementation in the online format and the amount of time it might demand of teachers when responding.

DISCUSSION, IMPLICATIONS, AND CONCLUSION

The paper instrument allowed us to have more open responses, which provided more nuanced data and access to the teachers' language for expressing statistical phenomena. In addition it provided a basis for deciding on the narrower range of items to be answered by many more teachers in the online version of the instrument. The responses to the open questions provided insight into teacher preferences for the form of reports, with teachers generally preferring graphs over tables. The online assessment appeared to meet its aim of being able to be completed in a short time frame, and could be undertaken in the teachers' own time, although we do not know who actually attempted the instrument in any teacher's name. The limited use of open response questions gave little opportunity to seek clarification or justification for some of the teachers' responses. Nevertheless the use of closed response items, using response choices arising from previous research and the phase one teachers' own language, made it possible to assign scores semi-automatically for most of the items, thus generating an overall statistical literacy "score" for much of the assessment (Pierce, Chick, Watson, Dalton, & Les, 2012). It must be noted that it was hard to construct an assessment instrument that did not "feel" like a mathematics test.

In both forms of the instrument there was, of necessity, a limited range of items, and we certainly cannot claim that either version provides a full test for statistical literacy. Moreover, we

were unable to fully evaluate the complete range of understanding of even the very specific statistical tools needed for the Victorian school data context. Indeed, as we designed both versions of the instrument, we became more aware of the complexity of professional statistical literacy: that to be professionally statistically literate requires functioning at the highest level of the framework—understanding the data set as a whole, including the nature of its distribution—as well as consideration of both professional and local contexts. In particular, the teachers appeared to have little difficulty with the "reading data" level or interpreting data for individual students, but making sense of the distribution of a cohort of students was much more difficult. It is likely that cognitive load—the need to keep track of so much information to get the "big picture"—plays a role here.

The frameworks—both the Theory of Planned Behavior (probing disposition but not discussed here; see Pierce, Chick, & Gordon, in press) and the professional statistical literacy framework (Pierce & Chick, 2013) proved helpful in structuring the assessment instrument. However, we must point out that disposition was addressed separately in the online version, whereas the paper instrument had allowed some elements of disposition to be embedded in open response items (e.g., the questions asking teachers to indicate aspects of the report deemed useful and difficult in Figure 2).

Finally, it was observed that some teachers have heuristics for interpreting these data sets that may well be applied without understanding the underlying properties of the distribution. These may permit a superficial interpretation of straightforward data, and may even allow some measure of success in a test of statistical literacy; but deeper understanding may not be present, and may be difficult to probe. It is also hard to assess teachers' capacity to identify what things *cannot* be deduced from data, and to measure their ability to interpret the statistical and practical significance of the reports they receive. Given the potential importance of such student data for teaching it is essential, however, that teachers have the statistical literacy to put it to good use.

ACKNOWLEDGEMENTS

This research has been funded by the Australian Research Council (LP100100388), the Victorian Department of Education and Early Childhood Development, and the Victorian Curriculum and Assessment Authority. We wish to acknowledge the contribution of other members of the research team: Ian Gordon, Sue Helme (University of Melbourne), Jane Watson (University of Tasmania), Michael Dalton, Magdalena Les (VCAA), and Sue Buckley (DEECD).

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50,* 197-211.

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

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.

Pierce, R., & Chick, H. L. (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*(3), 237-255.

Pierce, R., Chick, H. L., Watson, J. M., 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, eBook, pp. 602-609). Singapore: 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.

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