ISSUES FOR THE ASSESSMENT AND MEASUREMENT OF STATISTICAL UNDERSTANDING IN A TECHNOLOGY-RICH ENVIRONMENT

Rosemary Callingham
University of Tasmania, Australia
Rosemary.Callingham@utas.edu.au

As diverse technology use increases in education a number of issues are raised for assessment. This is true in all subjects, but may be particularly pertinent to statistics because students can now use large data sets and deal with multiple variables as part of their learning experiences. The issues are of two kinds. In classrooms, how do teachers assess work that has been produced through the use of technology? Using criteria that were developed for pencil-and-paper assessment may not be sufficient to capture the nature of students' thinking when the burden of computational data analysis and data display are removed. Outside assessment processes are also challenged with the advent of tools such as Computer Adaptive Testing, and complex interactive data displays. The implications of using technology as part of assessment processes, both inside and outside the classroom, are explored.

INTRODUCTION

"New goals for statistics learners", described by Garfield and Gal (1999), include the need to: Understand the purpose and logic of statistical investigations; Understand the process of statistical investigations; Master important procedural skills; Understand probability and chance; Develop interpretive skills and statistical literacy; Develop ability to communicate statistically; and Develop useful statistical dispositions. Progress towards many of these new goals can be facilitated by the use of technology. It is expected that students in school today have access to sophisticated technology in the form of hand-held or portable devices, as well as personal computers. Using these tools students can manipulate and "play" with data more quickly and in ways that are not possible when the data are in hard copy formats, and which provide access to powerful ways of understanding statistics in line with the goals of Garfield and Gal.

Jolliffe (1997) indicated that the ability to use computers effectively was one aspect of assessing statistics in the classroom that needed to be considered. Lajoie (1997) in the same volume described the use of computers to enhance teaching and assessment, including tracking students' actual interactions with the computer using specialist software to provide information to teachers about the ways in which their students were using technology tools. The period since the publication of these papers has seen technology use in schools increase significantly, with new kinds of software, more interactivity, miniaturization in the form of graphing calculators, smartphones and MP3 players, use of data loggers and many other applications. Lesh (2000, p. 193) describes how technology has produced an "explosion of representational media" that, although reducing the computational load, has "radically increased the interpretation and communication demands". He argues that technology has changed the nature of the conceptual systems that are created both within education settings and in the wider world. If the nature of the cognitive constructs developed in classrooms has changed, this must have implications for assessment generally as well as assessment of statistics specifically. In particular, the changed interpretive and communication demands must be addressed in statistics assessment.

In the classroom, teachers must recognize the affordances offered by the technology and adjust their assessment expectations accordingly. This adjustment might include the nature of the assessment task and the criteria used to make judgements about students' achievement. Outside the classroom, external agencies, whether planning assessment for placement at the end of schooling or measuring educational achievement for monitoring processes, must consider both the nature of the learning that has occurred in technology-rich environments and also developing approaches to using technology for assessment in large-scale testing programs. Issues associated with these two perspectives are considered in this paper.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php [© 2010 ISI/IASE]

ISSUES FOR CLASSROOM ASSESSMENT

In the classroom, statistical understanding can be enhanced by the use of appropriate technology (Lipson, Francis & Kokonis, 2006). As one example, in the middle-years of schooling, Tinkerplots (Konold & Miller, 2005) provides an environment that allows students to "tinker" with the data using drag-and-drop approaches to creating data representations. Use of colour lets students explore multiple variables in one representation, and both discrete and continuous variables can be considered. The Tinkerplots interface uses "data cards" to display an individual record but then provides a graphical interactive interface that provides opportunities for students to explore, display, and summarise the data contained in the cards. By asking students to produce a report and applying some suitable framework for assessment criteria, judgements can be made about the quality of students' work. Watson's (1997) tier model of statistical literacy is one such framework, or the statistical literacy hierarchy identified by Watson and Callingham (2003) is another possibility. Both these frameworks recognize the interaction between statistics and social settings and explicitly address the communication and interpretation issues identified by Lesh (2000).

Fitzallen (2008) took a different approach to assessment, using a traditional pen-and-paper instrument to identify students' understanding of graphing and data representation developed in a Tinkerplots environment. In a transition period while students are less familiar with software packages, it may be appropriate to use traditional assessment tools, but it does not address the call by Joliffe (1997) to assess computer use itself, nor Lesh's (2000) observations about the changes to conceptual systems that technology use creates.

An aim for statistics education (Garfield & Gal, 1999) is to move students to higher levels of thinking where they integrate the mathematical concepts with an understanding of the context in which the problem is situated. Teachers in a current research project, *StatSmart*, (Watson, Callingham & Donne, 2008) indicated that their students had the greatest difficulty interpreting the data and the results that they produced, rather than in manipulating the statistical computations. Figure 1 shows an innovative approach to assessing students' understanding of outliers from one teacher in this project. The students, all Grade 9 girls, were asked to collect data relevant to their peer group and to present an article for a teenage magazine as a biography of "Miss Outlier". They used technology to produce various displays of data and then interpreted these creatively in the context of a magazine article. Lucy Outlier (on the right) explores data related to anorexia and Amy's House (on the left) examined data about house size and energy.

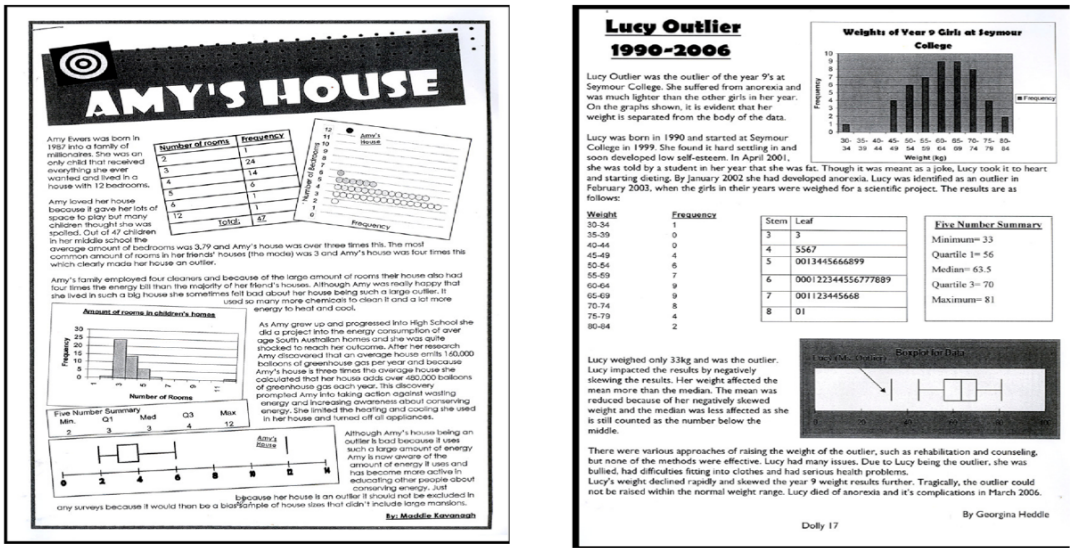

Figure 1. Using technology to support assessment tasks that address understanding of statistical concepts

Arguably the students could have produced more sophisticated displays but their articles demonstrate an attempt to grapple with the statistical notion of an outlier in the context of a social

setting of immediate interest to themselves. This attempt is supported by the use of technology which has allowed them to produce graphical representations of the data with minimal cognitive demands, so that they could consider the meanings behind the statistics. In this way, classroom assessment is using technology as a tool to scaffold higher levels of thinking.

Technology use, however, cannot compensate for poorly conceived tasks. The same principles that underpin all assessment must also apply when technology is used in classrooms either to produce work for assessment or as an integral part of the assessment itself. The challenge, however, is greater when the assessment is external and high stakes.

ISSUES FOR EXTERNAL ASSESSMENT AND MEASUREMENT

The impact of technology on large-scale assessment that is imposed by agencies outside the school is growing as computing facilities can handle larger amounts of data and have faster processing speeds. In many respects, however, the nature of the questions used has not changed.

Figure 2 shows an example of a Grade 8 item from the 2003 Trends in International Mathematics and Science Study (TIMSS) (TIMSS & PIRLS International Study Centre, 2009). It addresses students' understanding of average but can be answered correctly by any student who can apply the algorithm. In terms of the goals identified by Garfield and Gal (1999), it assesses only "Master important procedural skills". It is a machine-scored item that is cheap and efficient to administer on a large-scale but does not address higher-order thinking or interpretive capabilities.

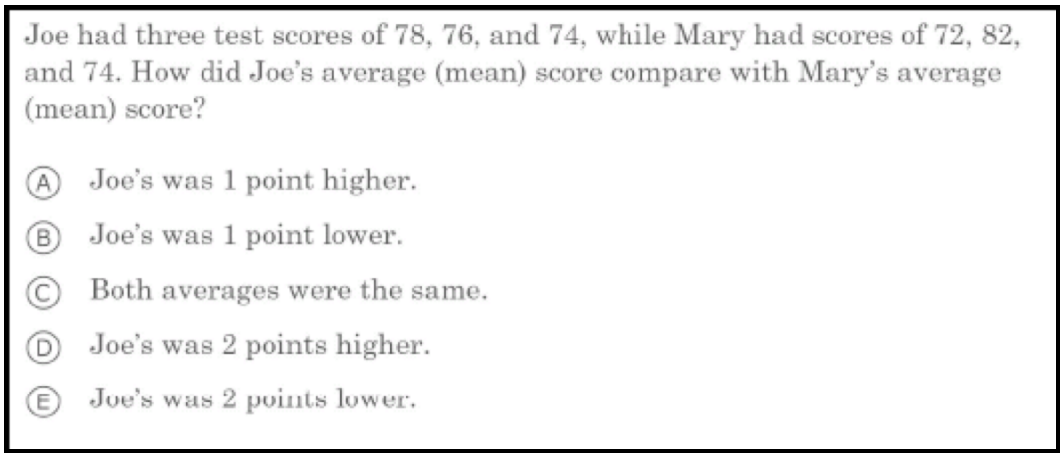

Figure 2. Grade 8 item from TIMSS 2003

In contrast, consider the item called "Oxygen" from World Class Tests (World Class Arena, 2009) shown in Figure 3 and aimed at the same age group. To answer this correctly, students interact with the graph using a slider to manipulate two different variables to identify their effect on the rate of oxygen production. They answer a sequence of carefully designed questions designed to test their ability to make inferences based on data. Responses are collected in hard copy and scored using a scoring scheme.

This approach provides students the opportunity to manipulate multivariate data in a context that is relevant to them in light of discussions about climate change. It makes use of the power of technology to provide questions that are challenging and that address higher-order thinking. These tests are not constrained by curriculum expectations and, hence, are more able to draw on technology to identify what is possible for students to do rather than what is expected of students at that age.

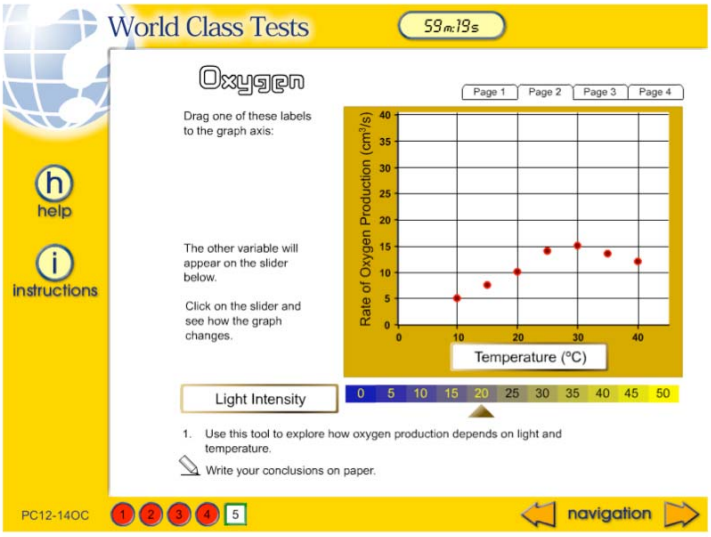

Figure 3. Oxygen item from World Class Arena text for 12-14 year-old students

One issue about the very different nature of these two tests is whether they are measuring the same construct, different dimensions of understanding statistics or two totally unrelated constructs. Wilhelm and Schroeders (2008) indicated that equivalence testing, to determine whether tests delivered using different modes are measuring in the same way, is important in high stakes examinations where part or all of the examination is delivered in different formats. Often the format affected the results suggesting that one or other of the formats, technology-based or paper-and-pencil, was likely to measuring a different construct. Their studies, however, were using tools such as smartphones. Few studies have been undertaken of the equivalence of instruments such as the examples indicated, and these are urgently needed.

Further complexity is added where computer adaptive testing is used. In this system, students answer questions and, on the basis of their responses, the computer provides the next item according to the ability of the student. These systems rely of having a large bank of validated items and are usually underpinned by Rasch measurement (Bond & Fox, 2007). One such system is the Lexile or Quantile Frameworks (Stenner, 2009) which automatically produces appropriate items for students and then places them onto a validated scale. At present, however, these systems rely on multiple choice questions similar to the TIMSS item shown in Figure 2. Systems are being developed, however, to automatically score open-ended written works, such as essays. These systems rely on sentence length and complexity, rather than a true analysis of the ideas presented, and, as such, would appear to have limited use at present for statistics education, where subtle differences in interpretation may indicate widely differing understanding. For example, when responding to a question “what does random mean”, two responses made by students are “Picked without order or any distinct pattern” and “To just pick anything”. The first of these answers is more sophisticated but uses very much the same kind of language. At the present state of development most computer software would not have enough information on the basis of these two answers to distinguish them, but a trained rater can identify such fine distinctions.

A somewhat different way of using the power of technology is that taken by the ARTIST project. ARTIST (Assessment Resource Tools for Improving Statistical Thinking) provides a database of validated items organized by topic and learning outcome. Teachers may use the data base to build tests appropriate to their context. Impressively, there are no purely computational items and both open-ended and multiple choice or forced choice item types are available (Garfield, del Mas & Chance, 2006). The topics are scaled and teachers are provided with these scales to use to make judgements about their students. This approach uses the power of technology to create meaningful assessments tailored by individual instructors to their specific needs.

DISCUSSION

Ben-Zvi (2000) indicated ways in which technology impacts specifically on statistics learning, echoing Lesh's (2000) view that technology is an agent for cognitive change. Garfield and Chance (2000) listed some of the challenges in assessment for statistic educators and this list included the need to embrace the use of new technological tools. The increasing impact of technology on teaching and learning statistics, however, has not been matched by developments in assessment to the same degree. Many of the exemplary approaches, including World Class Arena and ARTIST tools, are used by limited numbers of educators for specific purposes but have not made a serious impact on mainstream assessment.

Assessment must provide effective feedback to students and teachers to improve learning outcomes (Black & Wiliam, 1998). Technology holds the promise of doing this in immediate and effective ways. To realize this potential, however, requires creative thinking on the part of teachers and systems. In addition, statistics teaching itself needs to have more emphasis not only in the mathematics curriculum, but in cross-curriculum contexts as well.

In addition, serious thought needs to be given to Lesh's (2000) and Ben Zvi's (2000) contentions that using technology fundamentally changes the nature of the cognitive constructs developed. If this is so, then using traditional assessment approaches could disadvantage students who are increasingly learning via technology use. The nature of the changes to cognition needs to be identified and this is another area that needs further research. Given the claim that communication and interpretation demands are much greater in a technology-rich environment (Lesh, 2000), there is an urgent need for statistics educators at the school level to consider the increased cognitive load and change teaching and assessment accordingly.

Perhaps the greatest challenge for statistics education with respect to technology is to gain an understanding of the importance of statistics in a technological world, and the subsequent need to be able to assess students' understanding in meaningful ways using the technological tools with which students' are becoming familiar. Clearly, there is still much work to be undertaken.

REFERENCES

Ben-Zvi, D. (2000). Toward understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1), 127-155.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, March, 7-74.

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

Fitzallen, N. (2008). Validation of an assessment instrument developed for eliciting student prior learning in graphing and data analysis. In M. Goos, R. Brown, & K. Makar (Eds.), *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australasia*. Brisbane: MERGA.

Garfield, J., & Chance, B. (2000) Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*(1), 99-125.

Garfield, J., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review, 67,* 1–12.

Garfield, J., del Mas R., & Chance, B. (2006). *Assessment resource tools for improving statistical thinking (ARTIST)* [website]. Accessed 29 October 2009 from https://app.gen.umn.edu/artist/index.html

Jolliffe, F. (1997). Issues in constructing instruments. In I. Gal & J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 191-204). Amsterdam: IOS Press & International Statistical Institute.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration.* [Computer software] Emeryville, CA: Key Curriculum Press.

Lajoie, S. P. (1997). Technologies for assessing and extending statistical learning. In I. Gal & J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 179-190). Amsterdam: IOS Press & International Statistical Institute.

Lesh, R. (2000). Beyond constructivism: Identifying mathematical abilities that are most needed for success beyond school in the age of information. *Mathematics Education Research Journal, 12*, 3, 177-195.

Lipson, K., Francis, G., & Kokonis S. (2006). Developing a computer interaction to enhance student understanding in statistical inference. In A. Rossman & B. Chance (Eds.) *Working cooperatively in statistics education. Proceedings of the 7th international conference on teaching statistics*. (CD ROM). Salvador, Bahia, Brazil: International Association for Statistical Education and International Statistical Institute.

Stenner, A. J. (2009, July). *Lexile framework*. Plenary address to the Pacific Rim Objective Measurement Symposium (PROMS), July 28 2009. Institute of Education, Hong Kong.

TIMSS and PIRLS International Study Centre (2009). *TIMSS publically released items*. Boston: Author. Accessed 29 October 2009 from http://timssandpirls.bc.edu/

Watson, J.M., & Callingham, R.A. (1997). Data cards: An introduction to higher order processes in data handling. *Teaching Statistics 19*, (1) 12–16.

Watson, J.M. (1997). Assessing statistical thinking. In I. Gal & J.B. Garfield (Eds.) *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press & International Statistical Institute.

Watson, J.M., & Callingham, R.A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3-46.

Watson, J.M., Callingham, R. & Donne, J. (2008). Establishing pedagogical content knowledge for teaching statistics. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.) *Challenges for Teaching and Teacher Education.* (Proceedings of the ICMI Study 18 and IASE 2008 Round Table Conference). Monterrey, Mexico: ICMI, IASE, ISI.

Wilhelm, O., & Schroeders, U. (2008, September). *Traditional and computerized ability measurement: Stressing equivalence versus exploiting opportunities*. International Research Workshop on the transition to computer-based assessment. Lessons learned from the PISA 2006 computer based assessment of science (CBAS) and implications for large scale testing. Reykjavik, 29 September–1 October.

World Class Arena (2009). *Oxygen, interactive problem solving item*. London: GL Assessment. Accessed 29 October 2009 from: http://testwise.worldclassarena.org/testwise/ondemand/welcome.html.