TOWARDS STATISTICAL LITERACY - RELATING ASSESSMENT TO THE REAL WORLD

Penelope Bidgood
Kingston University, UK
Penelope.Bidgood@kingston.ac.uk

Students have different strengths and different approaches to learning so the assessment process should give opportunities for them to demonstrate their abilities and achieve the relevant learning outcomes. Reforms in statistical education at all levels place increasing emphasis on students' abilities to think and reason statistically using real data in appropriate contexts. The huge expansion in technology has given access to various data sources and advances in statistical software have greatly expanded the range of analyses that can be conducted almost instantaneously. Hence there are varied assessment strategies in statistics, whether in specialist or service courses, in which students can be set realistic problems to solve. This talk will discuss some issues in the assessment of statistics, drawing upon the author's experience in various projects over the years.

INTRODUCTION

The increase in attention to developing statistical literacy and the consequent importance of statistical thinking and reasoning has led to changes in the teaching of statistics at the tertiary level. This reflects, in part, requirements of employers who want graduates with a balance of technical statistical skills, including analysis and interpretation, with the ability to communicate their findings. They want graduates who can carry out a statistical investigation, choosing appropriate methods to analyse their data effectively, including building statistical models where appropriate, and present their results successfully, whilst appreciating the limitations of their methods.

Recent developments in statistical education, at all levels, emphasises statistical literacy (Ben Zvi & Garfield, 2004, Rossman & Chance, 2002), with a growing importance given to helping students to think and reason statistically using real data set in appropriate contexts, rather than focusing on statistical skills, procedures and computations.

The terms "statistical literacy", "statistical reasoning" and "statistical thinking" are often used interchangeably. Jolliffe (2010) summarising Garfield et al defines statistical literacy as the understanding, use and interpretation of basic statistics. Statistical reasoning involves understanding at a deeper level and is the way people reason to make sense of statistical ideas and information. At the top of the hierarchy is statistical thinking which includes knowing how and why particular statistical methods and models are used. Any, or all of these, as well as statistical knowledge, might be assessed but assessment of statistical reasoning and thinking presents its own challenges. Similarly, Wild and Pfannkuch (1999) provided five components of statistical thinking to be: recognition of a need for data; ability to "transnumerate" the data; recognition of variation; being able to reason from models; and being able to integrate statistical and contextual knowledge.

Garfield et al (2010) suggest that an assessment on statistical literacy might typically use words such as identify, describe, translate, interpret, read, compute, whereas statistical reasoning would use explain how/why and statistical thinking the words evaluate, critique and generalise. Of course these are neither exhaustive nor exclusive lists but give some indication of the types of assessment that might be used.

Assessments can be set where students gather and investigate real and relevant data, analyse them appropriately and communicate their findings effectively. Such tasks reflect the "Increasing emphasis and research in the statistical and statistical education community on data driven educational strategies; on the nature of statistical thinking and reasoning and on the roles of technology in statistical education" (MacGillivray, 2004).

A report on the state of mathematics, statistics and operational research (MSOR) at 71 tertiary institutions in England and Northern Ireland in 2000 reported that "Student engagement and performance has often been greatest when dealing with well-focused problems of a practical nature" (MSOR Overview Report, 2000). It also noted that assessment emerged as one of the most problematic areas in MSOR provision; in a follow-up study it was found that good practice

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

involved a wide range of assessment instruments to be used to address learning objectives. (Bidgood and Cox, 2002).

The American Statistical Association funded the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project, which focused on introductory college courses. The project report (GAISE College Report, 2005) recommended that the teaching of introductory statistics should emphasise statistical literacy and develop statistical thinking; use real data; stress conceptual understanding rather than mere knowledge of procedures; foster active learning in the classroom; use technology for developing conceptual understanding and analyzing data; use assessments to improve and evaluate student learning.

Aspects of real world assessment are enhanced by the huge advances in technology over the last few years. The internet allows access to a vast range of data from many fields, thus creating a valuable resource for lecturers and students. Statistical software has greatly expanded the range of analyses that students can conduct and this affects the assessment process (Bidgood et al., 2008b). Using software that allows students to visualise and interact with data appears to improve students' understanding of random phenomena and their learning of data analysis (Garfield, 1995).

ASSESSMENT WITH REAL DATA

Statistics is incorporated into many disciplines, in the sciences, social sciences and medicine for example. Whatever the main subject area, it has been found that, in such courses, students respond better to statistics if the data are genuine and relevant to the main subject area and examples are applicable in the discipline (Jolliffe, 2007; Garfield, 1995). Budgett and Pfannkuch (2010) describe statistics assessment with budding journalists, lawyers, politicians and sociologists.

Consequently, the teaching, learning and assessment resources produced should use authentic data and relevant scenarios. For example, Rossman and Chance (2002) found that data from scientific studies, popular media or student-collected motivate the students and they are introduced to statistical concepts, methods, and theory through a data-oriented, active learning pedagogical approach. Tyrrell (2010) describes assessment in a class with both specialist statistics and business students. Mashhoudy (2010) describes how both elementary investigations, through descriptive statistics and graphical methods, or more complex modelling can be applied to real and constantly changing data which students collect themselves from the internet.

There are many web sites that, taken together, allow thousands of data sets to be downloaded, so that in theory, lecturers and students have access to a rich source of information. However, many of these datasets lack descriptions of the contexts in which the data were obtained and some of the databases are difficult to access. Also, there are very few websites that provide help with teaching and learning activities using the data set and not every lecturer has sufficient time to create examples in context. For example, Barnett (2004) found that there was not a vast amount of UK-based university-level statistics teaching material freely available on the internet for general use, as popularly supposed.

The STARS (Statistical Resources from Real Datasets) project was one initiative to address this problem. Its main aims were to make available real datasets and associated scenarios applicable in the subjects of psychology, health and business and to develop learning and assessment materials to accompany these datasets for use with various packages (stars.ac.uk). The worksheets were designed to be used in introductory courses and cater for a wide range of student backgrounds, abilities and needs. Typically each dataset had three or four associated worksheets, each produced in a version using MINITAB, SPSS and Excel. Examples of worksheet subjects used in the 5 Health Sciences scenarios are shown in Table 1.

Table 1 Statistical Areas covered in the Health Sciences

Dataset	**Typical question**	**Statistical topic**
Obesity	*Did patients receiving the new drug lose significantly more weight than those who received the placebo?*	Charts
		Descriptive Statistics
		Tests of Means
		Tests of Association
Triglyceride	*Can we predict change in triglyceride level from change in weight?*	Charts
		Descriptive Statistics
		Correlation and Regression
		Further Regression
Breakfast	*How does adding sugar to low GI breakfast affect children's hunger later in the day?*	Charts
		Descriptive Statistics
		Tests of Means
		Non-parametric Tests
Overdose	*Do gender, age or marital status affect the risk of taking an overdose?*	Charts
		Descriptive Statistics
		Tests of Association
IVF	*Are larger clinics more successful than smaller ones at IVF treatment?*	Descriptive Statistics
		Tests of Means
		Correlation and Regression

THE PROBLEM OF PLAGIARISM

Assessments that require students to collect their own data, or to analyse given data using a statistical package have their own problems, not least collusion between and within cohorts of students or the risk that students might find data for which the analysis has already been done and is published. In service teaching for non-specialists, classes are often very large, with more than 500 students being quite typical in the business field. Large numbers of students, with the associated plagiarism issues, has an effect on the types of assessment used, the way in which they are marked and feedback to students.

The Plagiarism in Statistics Assessment (PiSA) project report (Bidgood et al., 2008a) gives several examples of ways to overcome these difficulties. For example it was found that there was an increasing move from take-home assignments towards "hands-on" practical assessment to counteract the plagiarism problem; however, lecturers were keen to use real and relevant data. Students can be given some data in advance and answer questions, either by hand or on a computer, in a later, timed, supervised session. This allows students to discuss the data with each other, but ensures that what is handed in and marked is their own work. A common technique is to ask students to gather their own data, either on themselves, from experiments or through general or specialised websites, such as Economagic for economists, with possible restrictions on the type and amount of data collected.

In particular the project found that there was much innovative work taking place on individualising assessment data, but also that here was much duplication of effort across the UK.

Two assessment tools were developed as part of the STARS project – Individualised Student Coursework using Spreadsheets (ISCUS), which is based on Excel (Hunt, 2005) and Dynamic Resources Using Interesting Data (DRUID), which is not tied to any particular package, but uses specific datasets (Davies & Payne, 2001). These each allow individualised datasets to be produced, together with assignments and solutions.

VARIETY IN ASSESSMENT

It is widely recognised that it is beneficial to employ a variety of methods when assessing students. Different topics in statistics can require different assessment regimes; for example knowledge of the laws of probability would not usually be assessed in the same way as the ideas of questionnaire design. However, most statistics teaching is in the applications or applied area and so most statistics assessment is focussed on how what students learn relates to the real world.

Statistics is ideally suited to providing a wide variety of assessment opportunities, whether in a specialist or service course. For example, students can be asked to collect and analyse their own data; they can be set realistic problems to solve either individually or in groups; they can carry out simple experiments and simulations; they can focus on how to communicate the results of statistical analysis to a non-specialist audience either graphically, verbally or in writing; they can critique the study designs and analyses of others. Different methods of assessment are appropriate for different elements of the curriculum.

It was with these issues in mind that the Royal Statistical Society Centre for Statistical Education (RSSCSE) under the auspices of the UK Higher Education MSOR Network funded the Variety in Statistics Assessment (ViSA) project to gather accounts of recent successful experiences in assessment of statistical learning at tertiary level, from around the world. This variety in what is assessed and how it may be assessed is discussed more widely in publications on the Variety in Statistics Assessment project (Bidgood et al., 2008b). Gal and Garfield's book (1997) addressed similar issues, although it covered all levels of education, whilst the ViSA project focussed on the tertiary level, which has seen many changes in the student intake and assessment methods in the last few years.

Statistical software has greatly expanded the range of analyses that students can conduct almost instantaneously and this has affected the assessment process. The massive expansion in technology means that large datasets, when available, can be used and stored easily. The use of technology has allowed earlier accessibility to, for example, complex investigations, exploratory data analysis and visualization, simulation, and re-sampling. (Begg et al., 2004).

Holmes (1997), using Bloom's taxonomy developed the following to represent some of this variety (Table 2), both in the terms used to pose the problems and the methods which might be appropriate to solve them.

Table 2 Methods of Assessment

Form of learning	**Verbs**	**Methods**
Knowledge	Define. State. List. Recognise Show. Label. Name. Identify	quiz part exam question
Comprehension	Explain. Clarify. Discuss Review. Describe. Recognise	Multi choice, assignment exam
Application	Demonstrate. Use. Calculate Estimate. Fit. Implement	Coursework. Apply specific techniques to particular problem
Analysis	Investigate. Solve. Interpret. Explore. Analyse. Explain Compare. Contrast.	Analyse a data set Oral presentation. Case studies Analyse computer output
Synthesis	Design. Formulate. Model. Improve. Adapt. Develop. Construct. Devise. Combine.	Design and carry out an investigation Prepare a report & present it. Essay. Open ended questions.
Evaluation	Assess. Criticise. Contrast. Review. Distinguish	Critically assess others' projects Defend own presentation

At one time, statistics was usually assessed mainly by formal examination; however, as the subject has become more applied, with an increasing emphasis on statistical literacy and reasoning, and with technological advances, the subject has lent itself to a more varied assessment regime. There are many modes of assessment, for example computer-based assignments, investigations, modelling assignments, presentations and group work. Students can be set more realistic problems; they can complete weekly on-line quizzes; they can carry out simple experiments and simulations; they can keep portfolios of their work; they are often required to communicate the results of their analysis graphically, verbally or in writing, including poster presentations; they can be asked to critique the study designs and analysis of others.

An example of the latter kind of assessment is given by McNiece (2010) who explains a research-based assessment in a medical statistics, where students have to find their own example of a medical case study. There are also arguments for requiring graduates to be able to write about their findings, not least of which is the ability to communicate with non-statisticians. Forster and Wild (2010) describe assessment of students' writing about statistics.

CONCLUDING REMARKS

Snee (1993) argued that students learn statistical thinking and methods by focusing the content and delivery of statistical education on how people use statistical thinking in real life situations.

There are two aspects in relating assessment to the real world –first, to use genuine data in context and secondly, to imitate some part of a professional or academic statistician's working life. The use of real data helps undergraduate students in service courses as they do not always see the immediate relevance of statistics course in their own discipline. Although each aspect can be found in either service or specialist courses, the former is perhaps more relevant to service courses where it is important to choose examples of interest, applied in the main subject area and the latter more pertinent in specialist courses which include the modelling process, case studies and projects. "At any level, statistics requires at least some synthesis in both thinking and skills. At the workplace, user and professional levels, good data analysis practice and thinking require a balance of technical statistical skills , quantitative skills, judgement, the ability to comprehend and model (in the mathematical sense) in contexts that may be unfamiliar, and analysis and interpretation that include balance, synthesis and communication" (MacGillivray, 2004).

This is greatly enhanced by developments in technology, which enable students to access various datasets, or to store data which they have collected efficiently. Typically, service courses in statistics are in the first year of a degree programme, although students often require statistical skills and reasoning in later parts of their programme, particularly when they have to analyse data in their own final year projects. The aim should be to produce graduates, whether specialist statisticians or not, who are statistically literate and are familiar with the different phases of a statistical investigation , are able to choose correct and methods and models in order to analyse data effectively and can present their results to both statisticians and non-statisticians.

Assessment that relates to the real world covers many aspects of what is required of the modern undergraduate - to access and work with genuine data; to work collaboratively, to be engaged in active learning, to have conceptual understanding, to be able to use technology appropriately, and to be able to communicate statistics effectively.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of various people to the STARS, ViSA and PiSA projects, most notably Neville Hunt, Colin James, Flavia Jolliffe, Bradley Payne and Vanessa Simonite. I also thank the RSSCSE for financial support for the ViSA and PiSA projects.

REFERENCES

Barnett, V. (2004). Review of online statistics teaching material. *MSOR Connections, 4*(2), 43-45. Birmingham, UK: Learning and Teaching Support Network, Maths, Stats & OR Network.

Begg, A., Erickson, T., Macgillivray, H., & Matis, T. (2004). Working group report on statistics curriculum: Content and framing. In G. Burrill & M. Camden (Eds.), *Curricular developments in statistical education.* Voorburg, The Netherlands: International Statistical Institute

Ben-Zvi, D., & Garfield, J. B. (2004). Statistical literacy, reasoning and thinking: Goals, definitions and challenges. In *The challenges of developing statistical literacy, reasoning and thinking* (pp. 3-16). Dordrecht, the Netherlands: Kluwer.

Bidgood, P., & Cox, W. (2002). Student assessment in MSOR. *MSOR Connections, 2*(4), 9-13. Birmingham, UK: Learning and Teaching Support Network, Maths, Stats & OR Network.

Bidgood, P., Hunt, N., Payne, B., & Simonite, V. (2008a). The PiSA (Plagiarism in Statistics Assessment) Project. *MSOR Connections, 8*, 1, 15-19.

Bidgood, P., Hunt, N., & Jolliffe, F. (2008b). ViSA - Variety in Statistics Assessment *MSOR Connections*, *8*, 1, 13-14.

Budgett, S., & Pfannkuch, M. (2010). Assessing students' statistical literacy. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 103-121). Chichester, UK: Wiley.

Davies, N., & Payne, B. (2001). Web-created real data worksheets. *MSOR Connections*, *1*(4), 15-17. Birmingham, UK: Learning and Teaching Support Network, Maths, Stats & OR Network.

Forster, M., & Wild, C. (2010). Writing about findings: Integrating teaching and assessment. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 87-102). Chichester, UK: Wiley.

GAISE group (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report.* Accessed via www.amstat.org/education/gaise/GAISECollege.htm

Gal, I., & Garfield, J. (1997). *The assessment challenge in statistical education.* Amsterdam, Netherlands: IOS Press.

Garfield, J. (1995). How students learn statistics. *International Statistical Review*, *63*, 1, 25-34.

Garfield, J., delMas, R., & Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary statistics courses. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 75-86). Chichester, UK: Wiley.

Holmes, P. (1997). Assessing project work by external examiners. In I. Gal and J. Garfield (Eds.), *The assessment challenge in statistics education*. Amsterdam, Netherlands: IOS Press.

Hunt, N. (2005). Individualised statistics coursework using spreadsheets. *Teaching Statistics*. Nottingham UK: Teaching Statistics Trust.

Jolliffe, F. (2007). *The changing brave new world of statistics assessment*. Paper presented at the IASE/ISI Satellite. http://www.stat.auckland.ac.nz/~iase/publications/sat07/Jolliffe.pdf

Jolliffe, F. (2010). Assessing statistical thinking. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 71-74). Chichester, UK: Wiley.

MacGillivray, H. (2004). Coherent and purposeful development in statistics across the educational spectrum. In G. Burrill and M. Camden (Eds), *Curricular Developments in Statistical Education*. Voorburg, The Netherlands, International Statistical Institute.

Mashhoudy, H. (2010). Individualised assignments on modelling car prices using data from the Internet. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 247-27). Chichester, UK: Wiley.

Mathematics, Statistics and Operational Research Overview Report (2000). *Quality Assurance Agency for Higher Education*. Gloucester, UK: QAA.

McNiece, R. (2010). An assessment strategy to promote judgement and understanding of statistics in medical applications. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 123-131). Chichester, UK: Wiley.

Rossman, A. & Chance, B. (2002) A data-oriented, active learning, post-calculus introduction to statistical concepts, methods and theory. In B Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town*. Voorburg, the Netherlands, International Statistical Institute.

Snee, R. (1993). What's missing in statistical education? *The American Statistician*, *47*.

STARS: Creation of Statistical Resources from Real Datasets (Available from http://stars.ac.uk)

Tyrrell, S. (2010). Staged assessment: A small-scale sample survey. In P. Bidgood, N. Hunt and F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 163-171). Chichester, UK: Wiley.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.