THE CHANGING BRAVE NEW WORLD OF STATISTICS ASSESSMENT

JOLLIFFE, Flavia
University of Kent
United Kingdom

There have been many changes over the last thirty to forty years in the way that statistics is taught and assessed. From hand calculations and assessment tasks that tested whether students had learnt how to answer questions of a type they had seen before correctly, we have moved to assessments that use computers and resources available on the internet and to examinations where sheets with formulae are provided or that are open-book. These newer forms of assessment present their own challenges. Examples of tasks are given in this paper, with discussion of their implementation and appropriate references to publications in statistical education.

INTRODUCTION

"Learning emphasises what the student understands": Holmes (2002). Learning in statistics might be defined as learning how to apply statistical methods, when to apply them, and what methods to apply, and underpinning these is the understanding of what statistics is all about. Learning in general can take the form of imitation, practice, and experiment, and statistics is no different in this respect. Assessment in statistics needs to match the learning and teaching (see Holmes, 2002). Garfield and Chance (2000) list some learning goals in statistics; Hubbard (2003) lists some specific purposes of assessment questions in statistics such as how to make decisions about suitable models, and how to apply a result in a new context.

The forms an assessment might take include tests and quizzes, oral assessments, take-home assessments, projects, and formal examinations. Depending on the task, an attempt at an assessment might be hand-written or done on a computer, and might be required to be completed by an individual student or by a group of students working together. Different kinds of assessments are appropriate for different purposes, and a mix of these might be desirable (Gal & Garfield, 1997b).

The emphasis in this paper is on changes that have occurred in the assessment of learning of applied statistics taught at the university level. Some examples of writing and computing assessments that the author has used, mainly of the take-home variety, are presented and discussed.

THE WAY IT USED TO BE

Thirty to forty years ago assessment questions in statistics fell into two main categories. One type tended to involve substituting numbers in formulae and used artificial data with easy numbers where the arithmetic could be done "by hand". The other type involved fairly standard algebraic derivations or manipulations. This was in line with the teaching of statistics at that time which, even to students whose main area of study was an application area such as engineering or the social sciences, was sometimes fairly mathematical in its presentation. Texts such as Spiegel (1961) and Ractliffe (1967) illustrate the approach and the types of questions set. To be fair, the constraints placed on computation greatly limited what could be done. For instance, multiple linear regression examples rarely had more than two explanatory variables. In teaching descriptive statistics, emphasis on changing the scale and origin in order to calculate the mean and variance of a grouped distribution, although illustrating an important property of these measures, often resulted in students thinking more about the steps involved in the computation than in interpreting and understanding their answers.

The questions set did little more than test the ability of students to learn how to do questions of a type they had seen already, or to memorise derivations they could reproduce in examinations. Thus it could be said that the questions assessed learning by imitation and by practice, but they did not really assess the understanding of statistical ideas or the ability to apply statistical methods. They were not assessing student learning of statistics as we define it now. Statistics was taught mainly at the university level, and although some courses paid lip service to

practical work, teaching tended to be in a traditional lecture format with little interaction between students and the lecturer, and assessment was almost entirely by timed closed-book examinations. Examination questions were similar to homework questions. Parts of questions might ask for discussion of results, explanation of reasoning, or what other data would be useful. On the whole, most of the questions in the more applied papers consisted of asking for calculations of specified quantities. On theoretical papers the questions might be described as exercises in algebra. It is in applied statistics that the main changes in assessment have taken place.

CHANGES IN THE TEACHING OF STATISTICS

Many of the changes in both teaching and assessment that have occurred are due to such influences as the journal *Teaching Statistics* (first published in 1979), the International Conferences on Teaching Statistics (held every four years since 1982), and the online journals *Journal of Statistics Education* and the *Statistics Education Research Journal.* This sharing of ideas, often backed up by research studies evaluating the effect of what has been done, is invaluable.

Probably the greatest changes in the way that statistics is taught, that in turn affect the way it is assessed, are a result of the so-called technological revolution and the massive expansion in the availability of hand-held electronic calculators, computers, and Internet access. These have freed teachers from computational restrictions on examples and opened up the types of statistical methods that can be covered. Asking students to do real statistics on real data, and to report on the results, is now feasible in a way that it was not in the past. Educators believe that the use of real data on topics of interest to students, whose main area of study is not necessarily statistics, helps to motivate them to learn statistics and to enjoy doing so. There are many sources of data on the Internet, including some where searches can be done by topic. Two useful sources are the *Journal of Statistics Education* at http://www.amstat.org/publications/jse/jse_data_archive.html, which has links to associated papers in the journal, and the StatLib datasets archive at http://lib.stat.cmu.edu/datasets/. The instructor could ask students to input data, download data from the Internet, or give them access to an electronic file of data.

Statistics is now taught in schools, including primary schools, and the approach taken and the assessments set depend to some extent on the age of the children or students involved. The expansion of teaching has given rise to an expansion in publication of statistics texts and of other teaching aids such as dedicated teaching software. The text-books already cited (Spiegel 1961, Ratcliffe 1967), and many others published at that time, look distinctly old-fashioned now. Today's texts often have accompanying CDs and associated web sites, might contain quizzes, traditional type exercises, case studies, and suggestions for longer investigations, and tend to use real examples both in the text and in the questions left for the reader to do. However, older texts sometimes contain useful material that is not covered routinely in more recent statistics text-books.

WRITING ABOUT STATISTICS

Students whose main discipline is neither statistics nor mathematics might resent having to take statistics courses, and some of them will be fearful of anything numerical and perhaps of using computers, so it is necessary to win their confidence and introduce more technical aspects slowly. In assessing their learning, tasks appropriate to their abilities and on which they have a reasonable chance of success are needed. Assessment tasks which involve writing about statistics are ideal, especially early in the course. There are also arguments for requiring future statisticians to write (Rudke-Sharpe, 1991), not the least of which is the ability to communicate with non-statisticians. More mathematically minded students are sometimes very poor at the writing parts of questions, but it is in writing assignments that we can test their understanding, and writing is in itself a way of learning. Such instructions as "Comment on your results" in assessment questions do not really go far enough.

Students in health related subjects and environmental health might be taught some statistics under the heading of epidemiology. A useful introductory assessment question that can be posed after an overview of what the course covers is to ask students to give examples of real studies that use the concepts and/or methods of epidemiology and to identify which of these occur

in the examples. They should be asked to state the source of each example and to give a description of it in their own words. They might also be asked to write about why epidemiology is important in their discipline. The teacher can learn about the application area from the students here, and the examples could be a useful source of questions in later repetitions of the course. A question of this kind can be adapted for students in other application areas and could also be set to students specialising in statistics.

In general, quantitative essays (see Jolliffe, 1976), where the emphasis is on the numbers rather than on the words, can be used to assess statistics learning. These are not dissimilar to projects, but they are considerably shorter and less ambitious in scope. Students would be expected to find publications of relevance, include numerical information in their essays (possibly in the form of summary tables), represent information as diagrams, and suggest further analysis. To do all of this successfully they would need to demonstrate some statistical thinking and show that they had learnt some statistical techniques. For example, social science students could be asked to write essays on topics such as "The Measurement of Poverty" or "The Effect of Opinion Polls on the Results of Elections".

The Chance newsletters (http://www.dartmouth.edu/~chance/chance_news/news.html) are a fruitful choice of fairly brief reports of results of studies and also contain discussion questions. Extracts from these reports followed by short questions can be posed as assessments. For example, a report of a study suggesting that children who live near overhead power lines are more at risk of developing cancer than those who do not might lead to questions asking students to suggest a study design that might be suitable for assessing the risk, and to list confounding factors that should be considered.

A similar assignment is asking students to read an article from a journal and to ask several short questions about the methods and results reported and on issues raised. As an example, the weekly *British Medical Journal* is online at http://www.bmj.com/ and issues published from January 1994 to the present may be downloaded free of charge from there. A paper reporting the results of a randomised controlled trial could, for example, lead to questions asking students to comment on any inclusion and exclusion criteria used in recruiting subjects and why these were needed, and how an analysis on an intention to treat basis was interpreted in the trial. Students specialising in statistics could be asked to explain how to interpret the results of, say, a Cox regression analysis given in table form. The journal's site has a checklist in question form for statisticians who review papers, and students could also be asked to answer these questions for a specific paper and to give a brief justification for each answer. This last is good training for future statisticians, but also assesses statistics learning in a way that standard questions can not. For example, answers to questions such as "Were the statistical analyses used appropriate?" and "Pre-study calculation of sample size reported?" on the checklist, backed up with either the reference to what was done or an explanation of what should or might have been done, forces students to think in depth about the issues involved, and, to some extent, to take on the role of statistical consultants.

Since student responses to assessment questions such as these vary so much, and students often make points that did not occur to the lecturer, there might not be a standard response. Marking will then need to be of a qualitative nature. As a guideline to an appropriate mark to award, marks might be considered in terms of the percentage bands corresponding to different classes of degree. With practice the marking does not take any longer than marking a more traditional type of question, and is considerably more enjoyable. Some statistics educators might feel they would be happier grading questions where marks can be allocated to the different steps of a calculation or of an algebraic question, but they could be pleasantly surprised when they experiment with setting and marking questions requiring written responses.

USING THE COMPUTER IN ASSESSMENTS

Computer software is increasingly becoming an integral part of statistics teaching at university level. This means that even students with a weaker mathematical background can be asked to implement more advanced techniques such as cluster analysis and logistic regression, and can be taught when such techniques are appropriate and how to interpret the results. In examinations or in other situations where it is not practical to ask students to produce output,

questions can involve the interpretation of given computer output and a discussion of its appropriateness.

One of the problems posed by assessments involving the production of computer output that are not done under examination conditions is that it is perhaps more difficult to detect who did the work than in assessments done "by hand" because the output produced by different students from the same package is likely to be identical. It is therefore essential that students are expected to produce individual answers in a substantial part of an assessment. One way of doing this is to ask for an interpretation of the output and for comments on the results. In the case of assessments where a spreadsheet such as Excel is used, students could be asked to describe, in the language of the spreadsheet and in such a way that someone else could reproduce the results, how they did calculations. Calculating index numbers (Jolliffe, 2003), and finding death rates are particularly suitable for this, but so is finding a standard deviation by using a spreadsheet as a calculator rather than by a built-in function. Students have to understand what they are trying to calculate in order to do this, so this type of question assesses their learning.

Supervised computer tests where electronic data files are provided are also a possibility. If the main task in the test is to produce output students could be asked to save their attempts to a floppy disk and the instructor could mark on the screen. Alternatively, or as well, students could be asked to copy and paste a selection of their output to a Word file in which they also give answers to questions on, or related to, their output. A pedagogic argument for tests of this kind is that students are being assessed on their ability to work under a time constraint. It is not unreasonable to make such tests open book, but allowing students to access material on the web makes it easier for them to exchange files. Careful invigilation is needed, but checking what students are looking at on screen is impossible if the language is not one the invigilator understands.

Students in relatively advanced statistics courses might be asked open-ended questions on real data sets either as a timed test or as a take-home assignment. For example, they might be asked to compare two or more methods of treatment in a given data set and asked to give a general summary of their findings supported by results of their statistical analysis. They could also be asked to comment on what other information might be useful to throw light on the investigation under consideration. In the case of a data set containing observations on several variables they could be asked to choose an interval scale variable they think can be explained by other variables in the set and to develop a regression model that explains their chosen variable and that might describe an interesting "real-world" relationship. Further, they could be asked to justify their choice of variables and to comment on both the statistical analysis and the relevance of their results.

Asking students to perform simulations ensures that each student has a unique data set. Another way to achieve this is to ask them to take a random sample from a very large data set and to do some statistics on the resulting sample. The randomness involved here also serves to teach students something about the concept of randomness, especially if they compare their results with those of their friends (Jolliffe, 2003). Although it might be difficult to detect whether the random process has been performed correctly, results very different from the norm will suggest that it has not, and we can mention this to the students concerned and point out that the chance of obtaining results such as theirs is very small. Pipelers, Thas, de Vleeschauwer and Ottoy (2006) give some technical details of randomising data sets which are available with feedback on responses to questions for students to self–test themselves. Randomly generated questions are another possibility. Simonite, Ells and Turner (1998) made use of an electronic list of students to produce individualised questions and solutions. McKenzie and Goldman (2006) mention that some publishers make randomised questions available electronically.

With any assessment it is similar unusual errors on more than one script that alert an assessor to the possibility that students have worked together or copied one another, but on occasion assessors also recognise that they have seen similar wording, even if correct, on other scripts. Assessments using the computer are no different.

CONCLUDING COMMENTS

It is important that teachers do not rely completely on assessment questions of types that have been used previously, either by themselves or by others. They should be open to change, and they should review how successful alternative approaches have been, just as they probably do in their teaching. However, Hubbard (2003) warns that non-standard questions should be introduced gradually into a course so that students become accustomed to the way they think about statistics. If performance on such questions counts towards a final mark on a course, a gradual introduction is important in case the new type of question is not successful. Students should certainly not be asked such questions for the first time in an examination. As regards examinations, the provision of formulae and open book examinations are developments worth considering as they remove dependence on memory and mimic the world of work where access to sources of information is available. Such help will not be of much use to students who have minimal learning in statistics.

Grading open-ended questions that test understanding is time consuming, and some of the types of assessment discussed in this paper would be difficult to implement in large classes, though it is worth noting that Simonite, Ells and Turner (1998) developed and used their method of generating individualised coursework in a large class in the UK as a means of reducing opportunities for the students to copy one another. Wild, Triggs, and Pfannkuch et al (1997) describe how they adapted traditional methods to assess students' statistical thinking on a course in New Zealand with 2,600 students on it.

There are many other types of task that can be set for different levels of student in this changing brave new world of statistics assessment. There are, for instance, many suggestions in Gal and Garfield (1997a) and some innovative assessments in Garfield and Chance (2000).

REFERENCES

Gal, I. and Garfield, J. (1997a). (Eds.) *The Assessment Challenge in Statistics Education.* Amsterdam, Netherlands, IOS Press.

Gal, I. and Garfield, J. (1997b). Curricular goals and assessment challenges. In I. Gal and J. B. Garfield (Eds.): *The Assessment Challenge in Statistics Education,* 1-13. Amsterdam, Netherlands, IOS Press.

Garfield, J. and Chance, B. (2000). Assessment in statistics education: issues and challenges. *Mathematics Thinking and Learning,* 2, 99-125.

Holmes, P. (2002). Teaching, learning and assessment: complementary or conflicting categories for school statistics. *Proceedings of the Sixth International Conference on Teaching Statistics.* CD-ROM and online at http://www.stat.auckland.ac.nz/~iase/publications.php, Hawthorn, Australia, International Statistical Institute.

Hubbard, R. (2003). Questions for formative and summative assessment that encourage deep rather than surface approaches to learning basic statistics in a computer environment. Paper presented at the Joint Statistical Meetings. Available online at https://app.gen.umn.edu/artist/articlesauthor.htm

Jolliffe, F. R. (1976). A continuous assessment scheme for statistics courses for social scientists. *International Journal of Mathematical Education in Science and Technology,* 7, 97-103.
Jolliffe, F. (2003). The same but different. *Teaching Statistics,* 25,1, 14-16.

McKenzie, J. and Goldman, R. (2006). Questions to assess the understanding of statistical concepts. *Proceedings of the Seventh International Conference on Teaching Statistics.* CD-ROM and online at http://www.stat.auckland.ac.nz/~iase/publications.php

Pipelers, P., Thas, O., de Vleeschauwer, D., and Ottoy, J-P. (2006). A web-based exercise tool using randomized datasets for statistical education. *Proceedings of the Seventh International Conference on Teaching Statistics.* CD-ROM and online at http://www.stat.auckland.ac.nz/~iase/publications.php

Ractliffe, J.F. (1967). *Elements of Mathematical Statistics, 2nd edition.* Oxford, Great Britain, Oxford University Press.

Rudke-Sharpe, N. (1991). Writing as a component of statistics education. *The American Statistician,* 45, 292-293.

Simonite, V., Ells, P. and Turner, W. (1998). Using IT to generate individualised coursework questions and solutions. *Math&Stats*, 9, 1, 16-18.

Spiegel, M.R. (1961). *Theory and Problems of Statistics.* New York, USA, Schaum Publishing Company.

Wild, C., Triggs, T. and Pfannkuch, M. (1997). Assessment on a budget: using traditional methods imaginatively. In I. Gal and J. B. Garfield (Eds.): *The Assessment Challenge in Statistics Education*, 205-220. Amsterdam, Netherlands, IOS Press.

Home Â» Browse Â» Magazines Â» News and Opinion Magazines Â» Monthly Review Â» Article details, "Brave New Corporate World: An Assessment of..." Magazine article Monthly Review.Â They emphasize the need to reorganize corporate management in the United States, with a view to enabling U.S.-based corporations to accommodate rationally to the changing structure of international trade. They argue that macroeconomic demand management is insufficient unless accompanied by mechanisms for dealing with change in industries, firms, and regions. The disturbing employment, profit, and growth statistics of the past fifteen years are seen as the consequence of corporate management's emphasis on short-run profits rather than long-run growth. and production of this report. The Brave New World of M&A. 5. Preface.Â he M&A market has entered a new and much more competitive phase. The stakes are much higher, reflected in the mounting size of deals, and the growing involvement of PE firms has injected a fresh rigor and professionalism into the entire process. Consolidation Is the Name of the Game in the Sixth M&A Wave. Sample summary statistics and balance check. Exposure's impact on beliefs regarding media. Impact of exposure to uncensored information.Â Does the acquisition of politically sensitive information change citizensâ€™ beliefs, attitudes, and behaviors? Answers to these questions are far from clear. Citizens with access to uncensored Internet may not seek out politically sensitive information, due to lack of interest in politics, fear of government reprisal, and unawareness or distrust of foreign news outlets.Â 1Freedom Houseâ€™s Freedom of the Press Report shows that 86% of the worldâ€™s population does not enjoy media free from censorship. In particular, states with â€œunfreeâ€ media are concentrated among regimes that are undemocratic and grant limited political rights for their citizens.