® WEAVING ASSESSMENT FOR STUDENT LEARNING IN PROBABILISTIC REASONING AT THE INTRODUCTORY TERTIARY LEVEL

MACGILLIVRAY, Helen
Queensland University of Technology
Australia

For any course in a student's degree program, the assessment should be part of an integrated assessment and learning package, with the components of the package combining to meet the learning objectives in a steady development of skills and operational knowledge that take account of the students' various prior and future learnings. This paper considers such a package for an introductory course in probability and distributional modelling, including its objectives with reference to the nature of statistical thinking in probabilistic and distributional modelling, and general assessment principles. A new component of assessment to strengthen the problem-solving environment and to better address some of the objectives is described, together with student and tutor feedback and student data.

INTRODUCTION

Tertiary educators' complaints of previous eras that students learn only for assessment are fading with the growing understanding that assessment should be designed for, and aligned with, student learning. Assessing for student learning requires identification of the purpose of the learning, of what the students are bringing to their learning, of how they learn and manage their learning, and of their perception of the roles of this particular learning in their courses and their futures. For the tertiary teacher, the variety and extent of demands and pressures on assessment packages can appear overwhelming and sometimes even contradictory. Amidst the balancing of formative, summative, flexible, continuous, rich and authentic assessment with demands for developing generic graduate capabilities such as teamwork, problem-solving and communication skills, lurk the problems of over-assessment and the politics of pass rates and attrition. The many dimensions of the assessment challenge are complicated in introductory courses by the diversity of student cohorts in which the wide range of backgrounds, programs, motivations and study skills need consideration in designing appropriate assessment and learning packages. Perhaps it is not surprising that Garfield et al (2002), in a survey of US statistics educators, report that of all areas of statistics education, assessment practices have undergone the least reform.

Calls for statistics educators to assess what they value (Chance, 2002) are reflected in general higher education literature emphasizing the role of assessment in learning (Angelo, 1999). Explicit aligning of assessment with objectives also features in both the general higher education (James et al, 2002) and statistics education literature (Gal and Garfield, 1998). Such alignment requires identification of course objectives in order to weave an integrated learning and assessment package appropriate for the student cohort and crafted to meet both statistics education goals and tertiary demands, particularly in diverse introductory course cohorts.

Contrary to the fears of tertiary staff who have been exposed to only the verbal descriptors/no marks methods of criteria and standards referenced assessment, the key messages in leading research in this general area are consistent with what is regarded as best practice by staff, and most desirable by students, in assessment in statistics (and mathematics). After a summary of these key messages, the paper briefly discusses how the spirit of the statistics education reforms can guide and link with objectives in the development of probabilistic reasoning and modelling at the tertiary level. A set of objectives for an introductory course are given, and the components of the assessment for learning package are described, with selected items, criteria and student reactions outlined or given where possible. A new component of the package was introduced in 2006 to strengthen the problem-solving environment and the learning at key stages, and the paper concludes with a brief analysis of the similarities and contrasts between the reactions of the 2005 and the 2006 cohorts.

KEY MESSAGES FROM CRITERIA AND STANDARDS RESEARCH

The general higher education literature emphasizes that the main objective of assessment should be an integrated package of learning and assessment, structured for student development, and based on the rich expertise, knowledge and experience of teachers. Students also need information on the roles of the components of assessment, formative and summative, within the package. *Students study more effectively when they know what they are working towards…Students are anxious to compare their performance against others*. (James et al, 2002).

Ongoing feedback to students is essential to learning, but it must be feedback that helps students identify their strengths and weaknesses and provides practical, efficient and effective guidance on how students can progress. Effective feedback can consist of a mixture of whole group/class guidance, with specific comments for individuals. Assessment that helps students to develop understanding of themselves as individual learners is clearly of prime importance. Such assessment also facilitates active engagement by students, and helps them understand and value the criteria, standards and methods of assessment and learning (Angelo, 1999).

Good learning and assessment packages in statistics/mathematics courses are integrated, balanced, developmental, purposeful packages with well-structured facilitation of student learning across the cohort diversity. Such packages possess an inbuilt *configuration or pattern of performance* for standards referencing as required by Sadler (1987). The configuration comes about through a combination of the construct of formative and summative assessment (aligned with student learning across the spectrum appropriate for the purpose and cohort), and the construct of timing, types and weights of assessment tasks. Sadler emphasizes the importance of *exemplars* (such as marked past student work, representative assessment tasks and model solutions) in identifying the characteristics (or criteria) of each component of assessment, with verbal descriptors to draw attention to salient criteria at different points. Indeed, interviews with students indicate they tend to ask for both criteria and exemplars for open-ended projects and reports, but request only exemplars for tests or smaller assignments. Students value exemplars for all tasks, including information on relationships of previous tasks or student work to the current learning situation. Students tend to regard the "ticks in criteria boxes" method as very poor feedback, and condemn a combination of grades over components of assessment without an explicit weighting schema.

DEVELOPMENT OF PROBABILISTIC REASONING AND DISTRIBUTION MODELLING

Recommendations for the "reform" of introductory statistics courses (Cobb, 1992) emphasized "more data and concepts, less theory, fewer recipes", and more active learning opportunities. Although it was recognised that the range of introductory statistics courses is substantial, the reform, and much of the associated research effort, has been oriented to the skills and thinking of statistical data investigations, both at school and tertiary levels, and the use of statistical data analysis methods and technology in real contexts at the tertiary level. Moore's (1997) observation that students need only an informal grasp of probability to follow the basic reasoning of statistical inference, is both correct and astute, and any probability considered in an introductory tertiary course whose focus is data analysis should be purposeful and minimalist (MacGillivray, 2006). If an introductory course is required to include probability and/or introductory distributions, the purpose, placement and structure should be clear, both within the course objectives and within the students' programs.

As is apparent in extensive research such as reported in (Metz, 1997), contexts such as coins, dice, spinners etc, play a valuable role in developing concepts of chance in children, especially as the research indicates that the time for children to become reasonably comfortable with the basic concepts is highly variable. However such contexts become increasingly abstract and remote from reality as students progress, with the overwhelming majority of tertiary students telling of their increasing boredom with years of school examples of coins, dice, cards etc.

The development of probability at school level beyond the realms of coins and dice needs urgent attention, and solutions are most likely to be found through connecting with data in real contexts. But what of the tertiary level? It is not only future statisticians who need development of probabilistic thinking in the same way as they need development of statistical thinking in data analysis. The increase in computing power plus the importance of areas such as risk analysis lead to demand for stochastic and distributional thinking in other disciplines. In addition, there is increasing user access to more sophisticated statistical data analysis techniques, and underlying all sound statistical data analysis techniques and the statistical thinking of practising professional statisticians are the concepts of statistical models and distributions. Certainly the term "distribution" features in statistical education reform, but more in terms of "distribution of data", (for example as in Cohen and Chechile, 1997), than in statistical modelling.

This general challenge of the nature and development of probabilistic thinking cannot be pursued here. However the objectives, and assessment and learning package considered here for an introductory tertiary course are not only aligned with the principles of assessment and those of statistical education reform (in terms of more data and concepts, less theory, fewer recipes and real contexts), but also aim to develop the probabilistic and distributional aspects of statistical thinking. These are relevant to applications and problem-solving in statistics, mathematical modelling, business, information technology, engineering and science, to name just some.

THE COURSE OBJECTIVES

The course is one of two first year statistics courses compulsory for all students taking a maths major or co-major, many of which are combined with other programs in business, information technology, engineering or science, including secondary mathematics education students. It is also available as an elective in many areas. The other first year statistics course compulsory for maths majors and co-majors is a data analysis course with minimalist probability, built around real data investigations (MacGillivray, 1998, 2005) and taken by large numbers of students in a variety of programs. The two courses may be taken in either order or together, and there are always differing opinions amongst students as to the best arrangement!

The course content is outlined in MacGillivray (2006). The content provides the vehicle for the objectives, and is oriented to the students' programs. It includes emphasis on estimating probabilities, problem-solving with conditional probabilities, choice and use of distributions, setting up simple Markov chains, collecting and evaluating data from queueing and other possibly Poisson processes, introductory bivariate concepts and combinations of correlated normals.

The course objectives are:

(a) To foster a sound understanding of the basic concepts and application of introductory stochastic and statistical modelling in building and analysing models of a wide range of processes.

(b) To help students unpack, analyse and extend their existing knowledge and understanding of probability and its many roles in statistics, mathematics, other disciplines and the real world.

(c) To develop skills in problem-solving and identification and building of models involving probability and random variables.

(d) To develop understanding and skills in linking data and stochastic models, particularly in estimation of probabilities and parameters and comparisons of models and data.

(e) To develop generic skills in group and individual work, written and oral communication.

(f) To consolidate foundation mathematical skills and develop students' abilities to use these in applications in probability and stochastic/statistical models

(g) To develop skills in discerning and applying relevant information and using the synergies between mathematical and statistical thinking

(h) To enable students to synthesize their skills and operational knowledge in introductory stochastic and statistical modelling for applications and transference to new situations.

THE ASSESSMENT PACKAGE

The course objectives and assessment are the result of an ongoing cycle in the "field" research of learning and teaching development, namely: analyse and develop; trial and implement; research, measure and reflect.

- FORMATIVE COMPONENTS

The course commences with a probability reasoning questionnaire (PRQ) whose purpose is to seed thought and contribute to a class forum later in the semester. It is based on a combination of the chance and data strands of the school syllabi and everyday contexts. Questions cover interpreting % chance and conditional % chance, combining % chances, assuming and not assuming independence, estimating conditional probabilities from a 2-way table of data, proportional chance reasoning, using probabilities of at least/no more than, and application of equally-likely probability in a simple network.

Each topic throughout the course commences with preliminaries which are designed as foundations, seeds and outlines for the concepts and thinking of the topic. For the first half of the course, these aim to help show students the operational understanding they are bringing to the course, enabling us to unpack, analyse, generalise, extend and apply. Hence the students are also able to self-assess their current skills in interesting and informal ways. In addition, this approach helps to demonstrate an essential element of mathematical thinking: that the analysis of solutions to simple problems enables generalisation and transferability to more complex problems and different applications. For the second half of the course, the preliminaries outline the key thinking of the next topic, linking with other topics. The process and some of the preliminaries are outlined in MacGillivray (2006).

As well as the PRQ and the preliminaries, class activities throughout the semester include quick data collections for estimating probabilities, class problem-solving, viewing and discussion of simulations of stochastic processes, use of selected computer modules specifically designed to support learning in distributional modelling, and problem-tackling on worksheets with unlimited individual help available. Collaborative work, discussion and lively debate are encouraged in all class activities.

- SUMMATIVE AND FORMATIVE/SUMMATIVE COMPONENTS

Formative/summative assessment contributes to some extent to overall grading, but its emphasis is learning and feedback that enables students to improve their performance on subsequent assessment of similar skills and operational knowledge.

The formative/summative assessment had two components prior to 2006 and three components in 2006.

(i) Four assignments based on class activities, examples and worksheets, with problems in authentic contexts contributed 20% to overall assessment prior to 2006 and 16% in 2006.

(ii) A group project in which two everyday processes that could be Poisson are chosen freely by each group, with access to advice, and data are collected in a manner that permits investigating their Poisson-ness. This involves thinking of everyday situations as statistical processes; deciding which processes are best observed through discrete variables and which continuous; the optimal way of collecting data from each process to permit evaluation as well as practicalities of the problem; and synthesizing the use of formal statistical inference and graphs to check for trends or dependence over time. Students consistently comment on its value as both a learning experience and an enjoyable activity. Its contribution to overall assessment is 10%.

(iii) A new formative/summative component was added in 2006 to strengthen learning opportunities for objectives (c), (e), (g) and (h). This is described in the next section.

The summative-only assessment consists of a problem-solving examination based on class activities, worksheets and hence assignments, ranging from simple to more complex in life-related authentic contexts. Students design and bring to the examination their own free summaries in any form they wish. Sufficient time is given for problem-tackling and the focus is on their processes. Its contribution to overall assessment was 70% prior to 2006 and 66% in 2006.

- CRITERIA AND STANDARDS

The criteria and standards for the assignments and examination are given in terms of exemplars with verbal descriptors on salient points in accordance with Sadler (1987). The criteria for the group project are given as:

a) Identify discrete variable that could be Poisson, and continuous variable that could be exponential. Collect sufficient consecutive data on each. Describe practicalities of collection.
b) Correctly use chi-square and Kolmogorov-Smirnov tests and interpret results
c) Graphs to assess absence of trend or consecutive dependence
d) Summarise conclusions and discuss any limitations

Each group receives a written report, with comments and level of achievement against each of the above, and an overall mark.

A NEW ASSESSMENT COMPONENT IN A PROBLEM-SOLVING ENVIRONMENT

Although the nature and role of each item of assessment must be clear, course objectives should be addressed by the overall assessment package, rather than individual assessment items. As part of the research, measure and reflect component of the "field" research cycle, and using observation of the most and least successful learning styles of the students, it was decided that more active learning was needed to develop skills in problem-solving, oral communication, discerning and applying relevant information, and transference to new situations. A number of topics were identified as most needful of immediate involvement of students in active problem-tackling in a problem-solving environment as described by Gal et al (1997), namely *an emotionally and cognitively supportive atmosphere where students feel safe to explore, comfortable with temporary confusion, belief in their ability and motivation to navigate stages*.

The tutorial/practical hours in weeks 4, 6, 9 and 12 were completely given to tutorial group exercises, which were carefully structured for immediate "hands-on" learning in areas of most need of this. The groups were allocated by the lecturer with different groups for each exercise. There was no compulsion to complete the exercise, as credit was for participation. Assistance was available as required, and each group handed in their group's work with the members' signatures. Full collaborative work was required, with particular emphasis on oral communication, with groups ensuring that explanations were shared within the group. Participation in each of these four special tutorials contributed 2% to the overall assessment.

Tutors and students voted the experiment an outstanding success, fulfilling all the above criteria for emotionally and cognitively supportive problem-solving environments and facilitating meeting and sharing across a class of 120 diverse first year students. The tutorials were buzzing with discussion verging on excitement, and early departures were practically non-existent. In the class forum and in written feedback, student opinion was almost unanimous that four was the ideal number. They did not want further encroachment on the more standard type of tutorial/practical, nor on the other components of the assessment. But the other tutorials benefited significantly from the influence of the four group-exercise tutorials, with increased participation, collaboration and active learning.

The first tutorial group exercise in week 4 was to use conditional probabilities and Bayes' theorem in detecting genetic diseases. This is challenging because there are three genotypes (diseased, carrier, and normal) and two probabilities of interest: the probability that the true DNA is 'no disease' given that the test says it is, and that the true DNA is 'has disease' given that the test says it is not. Student and tutor feedback agreed that this exercise was "good but a bit long".

The second tutorial group exercise in week 6 was setting up the matrix of a Markov chain. The problem was:

A particular industry needs to have technical services on call at night. It is assumed in the industry that the number of callouts per night, Y, follows a model with probabilities given by $Pr(Y = y) = (1 - p)p^y$, $y = 0, 1, 2, \ldots$ *..In one area of the industry, the value of p is assumed to be 0.2. Technicians in this area, when assigned to be on call at night are not permitted to do more than 3 callouts total over 2 nights. Once the limit of 3 is reached, any callout is directed to a backup procedure. The number of callouts done on night n for a technician on night call duty is a Markov chain. Obtain the matrix of transition probabilities for this Markov chain.*

Student and tutor feedback on this exercise was "challenging but very good".

The third group exercise was on using special distributions and included the following two questions.

(i) The service time for customers in a retail outlet is exponential with an expected length of five minutes, and service times are assumed to be independent of each other. Find the probability that

- *a service takes longer than ten minutes;*
- *a server completes more than 12 consecutive services in 30 minutes;*
- *in 12 services, no more than 1 takes longer than 10 minutes.*

(ii) Consider the data we collected in class on the number of children in your families (you + your siblings). What distribution might you consider for this variable and how might you estimate its parameter(s)?

The last group exercise was on applying the two-step conditioning result E(X) = E[E(X|Y)] in real contexts, and a simple bivariate discrete situation. Because the third and fourth group exercises were in the later stages of the semester, they were intentionally composed of smaller questions closer to worksheets to help combat the end of semester withdrawal tendencies often seen in first year students. Student and tutor feedback agreed in their approval of these.

THE QUANTITATIVE EFFECTS OF THE NEW ASSESSMENT COMPONENT

To gain some insight into the complex inter-relationships and contributions of the components of assessment, general linear modelling is used to analyse the relationships between, and effects of, the various components in the presence of the others. The importance of this type of approach, and avoiding the trap of quoting lists of two-way correlations, must be emphasized for such data in which all variables are highly inter-related. The variables considered are the total of the assignment marks, the group project mark, an indicator variable for participation in the project, the PRQ score, the exam mark, and, in 2006, the tutorial group exercise participation score. In the analysis of the results of all components of assessment, the exam mark is chosen as the response variable, as it is the largest assessment component, provides the greatest coverage of the objectives, and represents the range of operational skills and knowledge. Because major characteristics in the design of the assignments are to facilitate learning of the operational skills and knowledge, and to provide exemplars for the exam, the assignment total mark is then used as the response variable with the other components, except the exam mark, as predictors.

In 2005, significant predictors, after allowing for (that is, in the presence of) all other predictors, in explaining the exam mark were the assignments ($p=0.000$) and the PRQ score ($p=0.014$) with neither the group project mark nor participation in the project significant in the presence of the others. In considering the effects of the predictors other than the exam mark, on the assignments mark, both the group project ($p=0.06$) and the PRQ score ($p=0.019$) were significant in explaining the assignments mark.

In 2006, significant predictors in explaining the exam mark were the assignments ($p=0.000$), the PRQ score ($p=0.01$) and the group project ($p=0.032$) but not the tutorial group exercise participation score in the presence of the others. However, in considering the effects of the predictors other than the exam mark, on the assignments mark, the tutorial group exercise contribution was the *only* significant predictor ($p=0.004$) of the assignments mark, in the presence of the group project (mark and indicator) and PRQ score.

In all analyses above, the residuals indicated no problems with assumptions, and that the models were appropriate.

Because the assignments are designed to facilitate the learning and as exemplars (as described above), they should be the best predictors of the exam mark in the presence of all components of assessment. However the above shows that the tutorial group exercises are fulfilling their intended roles as significant enablers for all students across diverse prior understanding, and as facilitators of collaborative learning.

CONCLUSION

This paper demonstrates how identification and articulation of course objectives can provide genuine and practical assistance in weaving an integrated assessment and learning package consistent with general assessment principles and statistical education objectives, within the pressures of tertiary demands and diverse first year cohorts. Consideration of the learning and assessment package and the objectives also assisted in identifying a needful learning aspect and designing a highly successful new component of assessment to address this.

REFERENCES

Angelo, T. (1999). Doing assessment as if learning matters most. *Bulletin of the American Association for Higher Education.* http://aahebulletin.com/public/archive/angelomay99

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3). www.amstat.org/publications/jse/secure/v10n3/chance.html

Cobb, G.W. (1992) Teaching statistics. In L.Steen (Ed.) *Heeding the call for change: Suggestions for curricular action*. Washington, DC: The Mathematical Association of America.

Cohen, S. And Chechile, R.A.(1997). Probability distributions, assessment and instructional software: lessons learned from an evaluation of curricular software. In *The Assessment Challenge in Statistics Education*, eds. I. Gal and J. Garfield, Amsterdam: IOS Press, 253-262.

Gal, I., & Garfield, J. (1998). Aligning assessment with instructional goals and visions. *Proc. 5th Inter Conf on Teaching Statistics*. International Statistical Institute, 773-779

Gal, I., Ginsberg, L. and Schau, C. (1997). Monitoring Attitudes and Beliefs in Statistics Education. In *The Assessment Challenge in Statistics Education*, eds. I. Gal and J. Garfield, Amsterdam: IOS Press, 37-51.

Garfield, J., Hogg, R. V., Schau, C. and Whittinghill, D. (2002). First Courses in Statistical Science: The Status of Educational Reform Efforts. *Journal of Statistics Education* 10(2). www.amstat.org/publications/jse/v10n2/garfield.html

James, R., McInnis, C., & Devlin, M. (2002). Assessing learning in Australian universities. Melbourne: The University of Melbourne Centre for the Study of Higher Education.

MacGillivray, H.L. (1998). Developing and synthesizing statistical skills for real situations through student projects. *Proc. 5th International Conference on Teaching Statistics*, International Statistical Institute, 1149-1155.

MacGillivray, H.L. (2005). Helping Students Find Their Statistical Voices. *Proceedings of the ISI/IASE Satellite on Statistics Education and the Communication of Statistics*, Sydney, Australia: ISI, Voorburg, The Netherlands.

MacGillivray, H.L. (2006) Using data, student experiences and collaboration in developing probabilistic reasoning at the introductory tertiary level, *Proc 7th International Conference on Teaching Statistics*, International Statistical Institute.

Metz, K.E. (1997). Dimensions in the assessment of students' understanding and application of chance. In *The Assessment Challenge in Statistics Education*, eds. I. Gal and J. Garfield, Amsterdam: IOS Press, 223-238.

Moore, D. (1997). New pedagogy and new content: The case of statistics (with discussion). *International Statistical Review, 65*(2), 123-137.

Sadler, D.R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13, 191-209.