Paper based on this talk at the end of this document: CETL_MSOR_07_paper

Queensland University of Technology

Roles of assessment in learning in statistics and mathematics

Helen MacGillivray
School of Mathematical Sciences, QUT
Director, QUT Maths Access Centre
Australian Carrick Senior Fellow, 2007
President-elect, IASE (International Association for Statistics Education)
Visiting Fellow, UK CETL in university-wide maths & stats support

Juggling hats?

Assessment also requires balancing

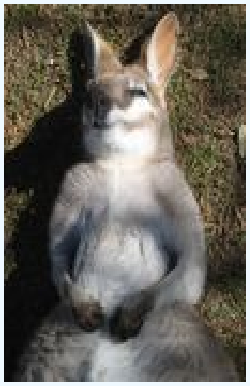
"Students learn only for assessment!"

Naturally

Students think what is assessed must be what is of value

If we value learning, assessing must be for learning

Engagement of students?



This presentation

Like mathematical proofs, end product doesn't reflect evolution to it

- **Overall comments on assessment**
  - **General HE & stats educ lit**
  - **Criteria, standards, objectives**

- **4 "stories" illustrating aspects of assessment in different contexts**
  - **Intro data analysis**: service & core, teaching data investigation
    - **Components of assessment balanced over different objectives**
  - **Intro prob & dist'nal modelling**: core & semi-service; unpack, analyse, extend, link with data
    - **Assessment for learning problem-solving**
  - **2nd year linear algebra**: core & semi-service; looks towards industry problems, applied research, computational maths
    - **Balance of continuous & tests, theory, applications & computing**
  - **2nd year engineering unit**: $\frac{1}{2}$ data analysis; $\frac{1}{4}$ dist'ns; $\frac{1}{4}$ comp maths
    - **Balancing components, workload, objectives**

- **Comments throughout on: alignment with objectives, balance, integrated assessment & learning packages, group work, collaborative work, plagiarism**

From general HE literature …..

- ***If learning really matters most, then our assessment practices should help students develop .. skills, dispositions, and knowledge…..***
- **Angelo, T., 1999, Doing assessment as if learning matters most. Bulletin of the American Association for Higher Education.**
- ***Students study more effectively when they know what they are working towards….. Students value assessment tasks they perceive to be 'real'***
- **James, R., McInnis, C., Devlin, M., 2002, Assessing learning in Australian universities. Melbourne: The University of Melbourne Centre for the Study of Higher Education**

Objectives of learning & assessment must be clear

.....reflected in statistics education literature

- **Care & indepth consideration of objectives, goals, contexts, content. Hogg (1991), Vere-Jones (1995), Moore (1997)**
- **Emphasis on data, statistical literacy & reasoning Cobb (1999), delMas (2002), Garfield et al (2002)….**
- **In a survey of US statistics educators, *of all areas of statistics education, assessment practices have undergone the least reform* Garfield et al (2002)**
- **Calls for *statistics educators to assess what they value* (Chance, 2002)**

- ***Explicit aligning of assessment with objectives* features in both the general higher education (James et al, 2002) and statistics education literature (Gal and Garfield, 1998).**

Aligning of assessment with objectives

Like mathematical proofs (& this presentation!), an iterative process

Components of assessment ⟷ objectives
to produce an assessment, teaching & learning package that is

integrated, balanced, developmental, purposeful, with structured facilitation of student learning across the student diversity

- **This needs identification of**
  - **Purpose of the learning**
  - **What the cohort are bringing to their learning**
  - **How the students manage their learning**
  - **The students' perception of its roles for them**

For wide range of backgrounds, programs, motivations, study skills……

Recent pressures for staff in tertiary assessment

- **Seeking balances & paths amongst:**
  - Formative, summative, flexible, continuous, rich, authentic
  - Generic graduate capabilities
  - Work-integrated learning
  - Criteria & standards referenced assessment
  - HE fads, generalisations & arbitrary rules

- **Plus challenges of:**
  - Avoiding over-assessment
  - Politics of pass rates & attrition & standards
  - Increasing diversity of student cohorts
  - Instant gratification generation
  - Workloads – students & staff

Criteria & standards referenced assessment

The term 'criteria-referenced assessment (CRA)' is often interpreted as meaning verbal descriptors of standards

Not so

- ***in criteria & standards-referenced assessment it is the <u>configuration</u>* (Kaplan, 1964) *<u>or pattern of performance</u>*' Sadler (1987) *which is used for ranking or reporting a level of achievement* ….**

- **Good packages have inbuilt configuration or pattern of performance**
- **Configuration comes from**
  - **construct of formative & summative assessment aligned with objectives & learning across cohort**
  - **construct of timing, types & weights of tasks**
- **Exemplars help to identify characteristics of each component of assessment, with verbal descriptors for salient criteria**

Statistical Data Analysis 1

science, maths, surveying, educ….: approx 500 pa

- **Theme is basic statistical data concepts and tools & using them in real data investigations.**
  - **Separate phases – tools & building blocks of procedures, concepts and procedural skills**
  - **Synthesis – choosing, using, interpreting, combining in whole data investigations**
- ***Structure, examples & learning experience built around real data investigations from first ideas through to report***
  - **Planning, collecting, handling, graphing, summarising, commenting on …. data**
  - **Categorical data – chisq tests; principles of testing hypotheses; p-values**
  - **Revision of normal; standard errors; confidence intervals and tests for 1 & 2 means, proportions, variances. Tolerance intervals**
  - **ANOVA & exp'tal design (via software): interaction (2-way), multiple comparisons, checking assumptions. Unbalanced data**
  - **Multiple & polynomial regression (via software): interpretation, diagnostics, re-fitting**

Learning & assessment package

- **Computer-based practicals on datasets from past student projects**
- **Worksheets with full solutions**
- **Fortnightly quizzes of fill-in-gaps & short response type; out Sunday, in by Friday: best 5 out of 6 contribute 10%**
- **Workfolder containing their ongoing work on the worksheets and their marked (collected) quizzes: 3%**
- **Whole semester group project in planning, collecting, analysing & reporting data investigation in context of group choice: 20%**
- **In-semester test (similar to quizzes 1-4): 10%**
- **End of semester exam (similar to quizzes 1-6, more on 5, 6): 57% ***

Quizzes, test, exam: exemplars + exemplar processes
Quizzes & test formative & summative; exam summative
Assistance given for quizzes – most important aspect is DOING them

****For a few years also an optional essay on how statistics revolutionised science in the 20th century: 10% if improved overall result. Dropped because (i) almost never improved result (ii) attracted students who could least afford the time. Objective not worth student & staff effort***

Research on numeracy/maths & statistical reasoning of cohort

- **Numeracy/maths on entry: highly diverse – see Wilson & MacGillivray *Counting on the basics: mathematical skills amongst tertiary entrants*, (2007) IJMest 38(1), 19-41**
- **General statistical reasoning on entry: Wilson and MacGillivray, *Numeracy and statistical reasoning on entering university*, 7th International Conference on Teaching Statistics (2006) http://www.stat.auckland.ac.nz/~iase/publications/17/C136.pdf**

- **Numeracy & level of maths stood out as most important predictors of general statistical reasoning**
- **Fish question greatest discriminator between core & advanced maths preparation**
  - ***A farmer wants to know how many fish are in his dam. He took out 200 fish and tagged each of them. He put the tagged fish back in the dam and let them get mixed with the others. On the second day, he took out 250 fish in a random manner, and found that 25 of them were tagged. Estimate how many fish are in the dam.***

Own choice group project

- **Teaches & assesses data investigation & synthesis of procedure choice & interpretation**
  - Other assessment can focus on operational knowledge & skills - tools & building blocks of procedures, concepts and procedural skills
- **Group because task needs a group**
- **Guidelines & descriptors of 3 criteria with standards given** (MacGillivray, *Criteria, standards and assessment in statistical education*, Proceedings International Statistical Institute, 55th Session, 2005)
- **Feedback on proposal + ongoing help; they propose – we advise**
- **Use of past datasets in class demonstrations and practicals**
- **Access to past projects, including assessments, and model reports**
- **Each group receives a written assessment report with comments & marks for the 3 criteria**

Criteria, standards & exemplars. Formative & summative

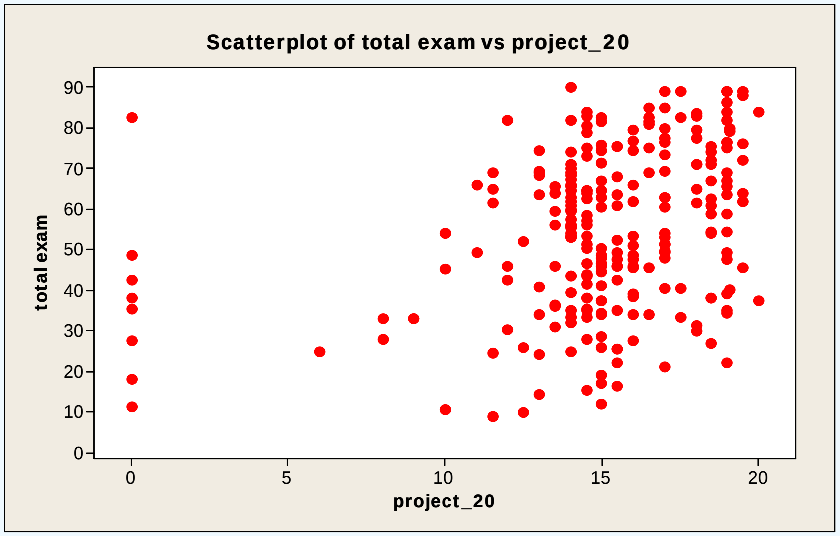
Own choice group project

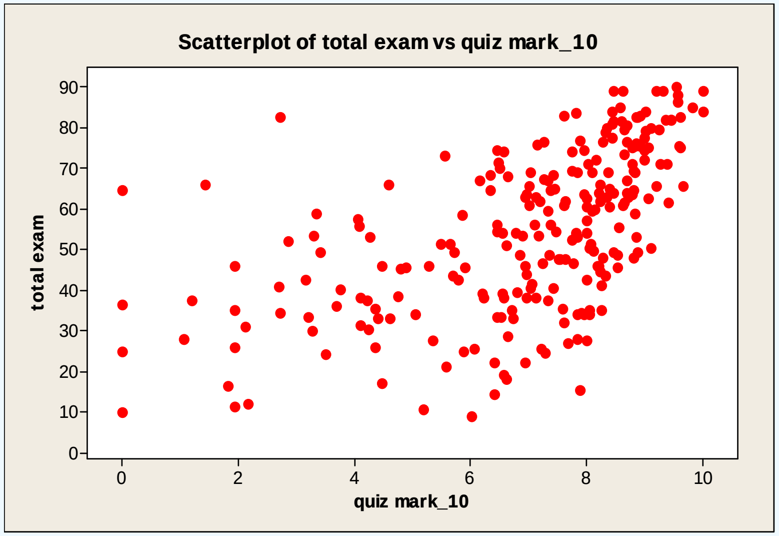
- Criteria
  - (i) Identifying context and issues; planning and collecting of data; quality of data and discussion of context/problems
  - (ii) handling, processing, preparing & understanding data & issues; exploring and commenting on features of the data
  - (iii) using statistical tools for statistical analysis and interpretation of the data in the context/issues
- Group problems?
  - They form groups, we help as necessary using pracs
  - Dropouts after week 8 can cause some problems but solvable
- Plagiarism?
  - Projects retained & designated "published". To copy = 0
- Contributions balance? Not a problem with right emphasis on project as learning experience
  - Almost never in (i)
  - Seldom in (ii); allocation of tasks helps in (ii) & (iii);
  - Leaders in (iii) tend to learn more, need less revision for exam & do better…..and learn by helping others

Just a few recent titles

- Still time for play?
- How long can you suck?
- Talking your ear off
- Gym junkies
- Gifted hands
- Ah McCain you've done it again
- An analysis of alcohol induced loquaciousness
- Investigation into student internet usage
- Maritime museum usage
- pH of river
- Optical illusions
- Voluntary student unionism: to join or not to join
- We love muffins
- Human curiosity
- Holding breath
- Usage of the 15 min workstations in the GP library
- Strength of our athletes
- Where are all the single people?
- Seed germination
- The big news about breakfast
- Music and the people

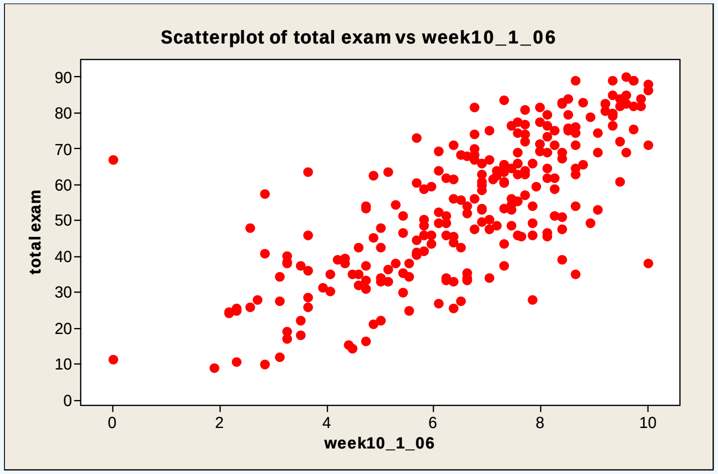
Graphs 2006 sem 1





Low +ve relationship ✓

Moderate +ve relationship ✓



High +ve relationship ✓

Statistical modelling 1

All maths programs, maths electives, maths educ: approx 120

- **Builds skills and foundations in concepts & thinking in**
  - **Intro probability, conditional arguments, distributional and stochastic modelling for**
  - **applications in a wide range of areas, from communication systems and networks to traffic to law to biology to financial analysis**
- **Analyzes prior understanding/ misunderstandings**
- **Links with data, observation and simulation**
- **Links with and consolidates 1st yr calculus & algebra skills**

Whole approach is problem-solving & modelling

Statistics education "reform": "more data & concepts, less theory, fewer recipes" (Cobb, 1992).

It's time to apply this in teaching probability & distributions

Formative components of assessment

- Initial general probability reasoning questionnaire (PRQ) to seed thought & discussion (introduced 2004)
- Class activities, simulations, selected computer modules, worksheets with unlimited help
- Each topic has preliminary experiences or exercises or discussion points (development completed 2005)
  - prior knowledge, foundations & seeds
  - perceive, unpack, analyse, extend

"Using what we already knew to learn other stuff was really good and helped us learn other stuff"
A student definition of constructivism perhaps?

Formative/summative & summative components: all oriented to problem-solving

- **Four assignments based on class activities, examples and worksheets, with problems in authentic contexts 20% before 2006; 16% in 2006**
  (Assistance available. Collaboration – yes; straight copying rare)
- **Group project. 2 everyday processes that could be Poisson (free choice); data collected; Poisson-ness investigated by combination of tests and graphs**
  10%.
- **End of semester exam. Problem-solving based on activities, worksheets, assignments; ranging from simple to slightly complex in life-related authentic contexts. Students design & bring in own summaries (4 A4 pages)**
  70% before 2006; 66% in 2006

Some examples from group projects

- ***Australian Rules (football) grand final***
- ***Time spent on phone***
- ***Pedestrian traffic in mall***
- ***Time to be served icecream***
- ***Occurrences of "Harry" per page in a Harry Potter book***
- ***Traffic on a pedestrian bridge***
- ***Distribution of leaves on tiles***
- ***Behaviour of ants***
- ***Arrivals & service at library***
- ***Distances between coffee shops***
- ***Service in "fast" supermarket checkout***
- ***Time between customers wearing high heels.***
- ***Time between changes of a baby's nappy***

New assessment component in a problem-solving environment

Problem-solving environment Gal et al (1997)

"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."

- **Formative assessment & assignments designed for managed optimal learning but students needed greater persuasion to learn through trying ('ave a go ….)**
- **Some topics identified as most needful of immediate involvement of students in active problem-tackling in an environment that maximises engagement & learning**

Tutorial group exercises, 2006

- 4 practicals structured for immediate "hands-on" learning.
- Groups allocated; different groups for each practical.
- No compulsion to complete exercise; credit for participation.
- Assistance available as required.
- Full collaborative work required, with groups ensuring that explanations were shared within the group.
- Participation in each of these four special tutorials contributed 2% to the overall assessment.

Evaluation of new component

- **Qualitative**
  - **Tutors and students voted experiment success.**
  - **The tutorials were buzzing, and early departures were practically non-existent.**
  - **Student opinion was that four was the ideal number.**
  - **Other tutorials benefited significantly.**
- **Quantitative**
  - **Assignments provide exemplars for exams**
  - **Data support that assignments most important in predicting exam (as desired!)**
  - **In 2005, assignments score depended on group project & PRQ score**
  - **In 2006, assignments score depended *only* on tut group exercises score for participation – strategy worked!**

2nd year linear algebra unit

maths+others e.g. maths educ, physics, eng – approx 80-90

- **Mixed student cohorts with often bimodal results**
  - **Balance of theory and practice?**
- **Some changes in continuous assessment – did they help or impede student learning?**
  - **Challenge of student engagement**
- **Interface of first and second level courses**
  - **first level courses respond to school/tertiary interface**
  - **first year units – which are best predictors?**

- ***The examples and learning experiences in unit are motivated by higher level needs in mathematics generally & particularly computational mathematics, & by applications based on experience with industry problems.***

Assessment package, 2003 & 2005

- **2003**
- **21% continuous assessment**
  - 3 Maple group assignments totalling 21%
  - mid-semester exam 15%
  - final examination 64%.

- **Lecturer's observations plus feedback:**
  - Maple group assignments too heavy for 7%
  - Students needed more structured help with their learning

- **2005**
- **40% continuous assessment**
  - 2 Maple group assignments totalling 24%
  - 3 "homework" quizzes totalling 16%
  - final examination 60%. Similar in style, format and level to 2003

Analysis of data: assessment components

- **For both continuous assessment programs, a test-type component and a Maple group assignment component combined as best predictors of exam**

- **Exam has applications but no actual Maple use, providing support of the claims in the literature, that both theory and practice contribute to overall learning and understanding in linear algebra**

- **Reassurance that the change in the continuous assessment program is not detrimental to performance, and appears to assist in learning**
- **Lecturer's concerns about high marks in the 2005 continuous assessment program are reflected by only 25% of the variation in exam marks being explained**
- ***but the challenge of how to grade the continuous assessment can be tackled with confidence in the program's facilitation of student learning across the theory and practice components of the unit***

Analysis of data: 1st year predictors

- **Formal prerequisites**
  - **1st level calculus unit and**
  - **1st level introductory linear systems and analysis unit, with the brief synopsis**
  - ***linear systems and matrices; vector algebra; coordinate systems; introduction to abstract algebraic systems; complex numbers; first and second order differential equations.***
- **Entry to 1st yr units via advanced mathematics in senior school or equivalent 1st yr unit.**
- **Alternative prerequisites 1st yr engineering maths**
- **Other compulsory 1st year units for maths degree are an introductory unit in computational mathematics, Statistical Data Analysis 1 & Statistical Modelling 1.**

Analysis of data: 1st year predictors

Data are complex because of different pathways. But best single predictor amongst 1st yr units, of performance in 2nd yr linear algebra in 2003 & 2005* was Statistical Modelling 1.

Synthesis of techniques and problem-tackling with new contexts, theory and applications appears to be the common thread linking these unlikely partners

* Note: changes in the 1st year units since then have probably changed this

2nd year engineering maths unit

all engineering programs - approx 450-520

- **Unit “new” in 2007 but composed of sections common across previous engineering units**

Content in 2007:

- **Statistical data investigations & analysis (1/2 unit)**
- ***As in Statistical data analysis 1; as given in all eng programs since 1994***
- **Introductory numerical analysis (1/4 unit)**
- **Introduction to random variables & distributional modelling, including linear combinations of normals, goodness-of-fit & introduction to reliability (1/4 unit)**

Level of unit

First year work in Science and Maths

- **Statistical data analysis 1**
- **Numerical component extract from 1st yr unit**
- **Intro rv's & distributions extract from Statistical Modelling 1**

But

- **It's different**
- **It's not straight calculus/algebra & any of these that are needed must be at fingertips in new contexts because of amount of material**
- **The statistics (both parts) full of new concepts & new ways of thinking**

‘tis always thus in Australian eng courses

Because

- **of the philosophy of Australian eng courses (whether new, old or middling)**
- **engineering needs the most technical maths faster than any other discipline**

AND

- **engineering needs the most maths generic skills faster than any other discipline**

Advantages of stats being in 2nd year eng are….

(i) ***they’re 2nd years in some ways & they have better maths thinking than most other disciplines***

(ii) ***they start reflecting on their studies (I’ve been listening to & observing 2nd (or 3rd) year eng students for over 30 years)***

Disadvantages of stats being 2nd or 3rd year eng are….

(i) ***they think they’re 2nd years in every way***

(ii) ***it’s stats & they’re eng students***

(iii) ***many tend to think it’s less important than other units***

Learning & assessment package

focus is on learning by doing

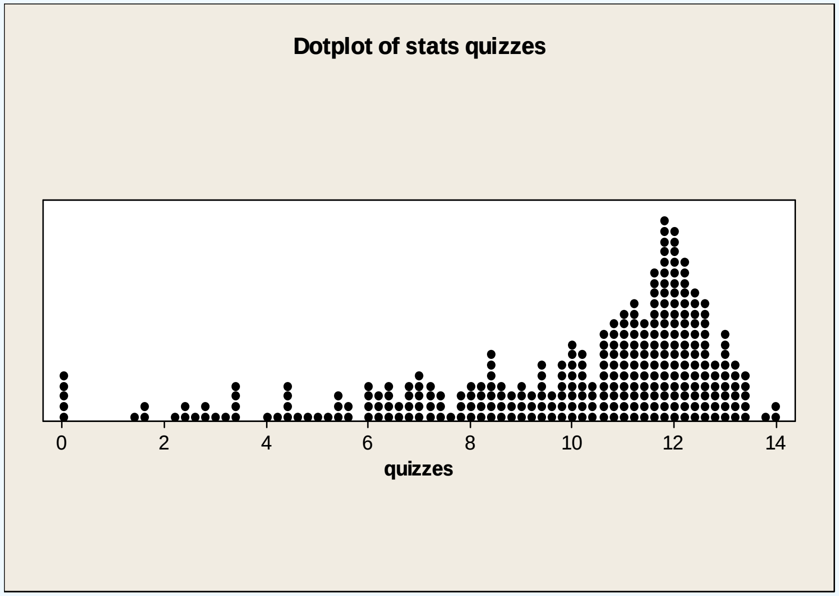
- Computer-based practicals on datasets from past student projects
  - Weeks 1-6 on statistical data analysis
- Worksheets with full solutions
  - For all sections: 15 worksheets in total

- Stats (¾): five quizzes of fill-in-gaps & short response type 14%
- Whole semester group project in planning, collecting, analysing & reporting data investigation in context of group choice 20%
  - As for all eng since 1995 & as in Stat data Analysis 1
- Numerical analysis (¼): assignment 6%
- End of semester exam (based on quizzes & w'sheets): 60%
  - Ensures overall coverage correctly proportioned

Quizzes, assignment, exam: exemplars + exemplar processes
Project: criteria, standards & exemplars

Quizzes, project & assignment formative & summative
Assistance given for quizzes – most important aspect is DOING them
Exam summative

Assessment data: stats quizzes



Stats quizzes designed for efficient & effective learning.

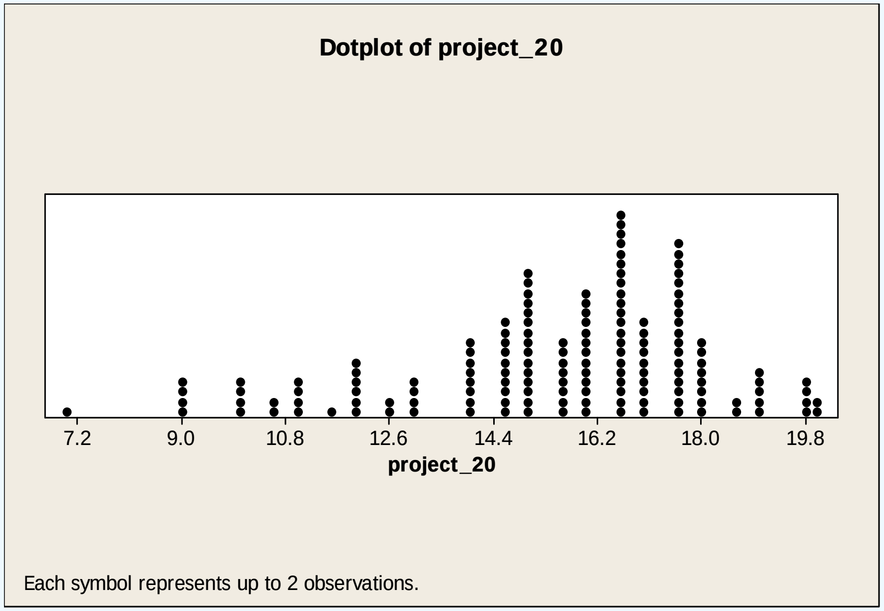
Evidence of value over years & units.

Plot good – why?

Strategy introduced late '90's in an MBA ½ unit with highly diverse cohort with FT jobs. Then developed further in eng unit when data analysis became a ½ –unit module; strategy used to decrease time demands so as to keep the full project.

Unexpected & amazing side effect in eng unit was drastic reduction in copying. Students still worked together but argued/explained instead of copying. Similar effects observed in Statistical Data Analysis 1.

Assessment data: stats project



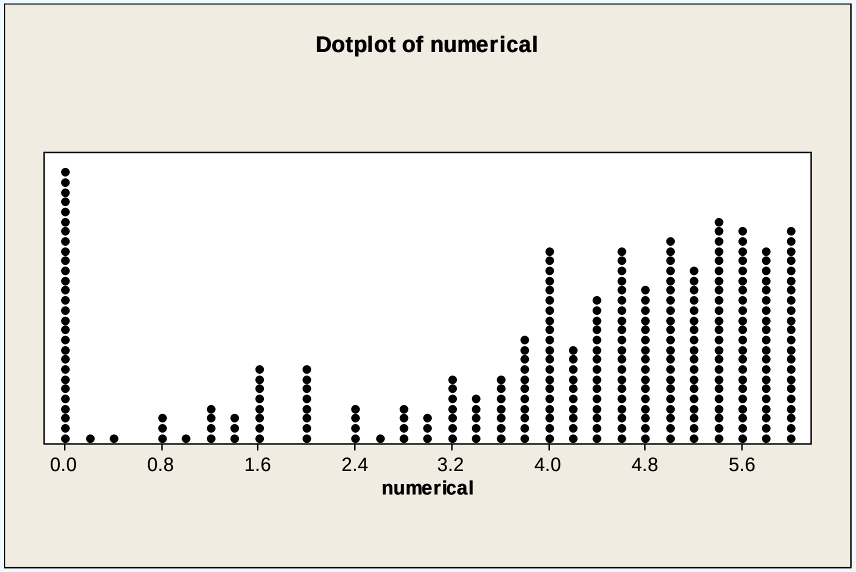
Project teaches & assesses synthesis of planning, thinking, understanding, choice of procedures and interpreting output.

Practicals designed to provide learning for project as well as for unit content.

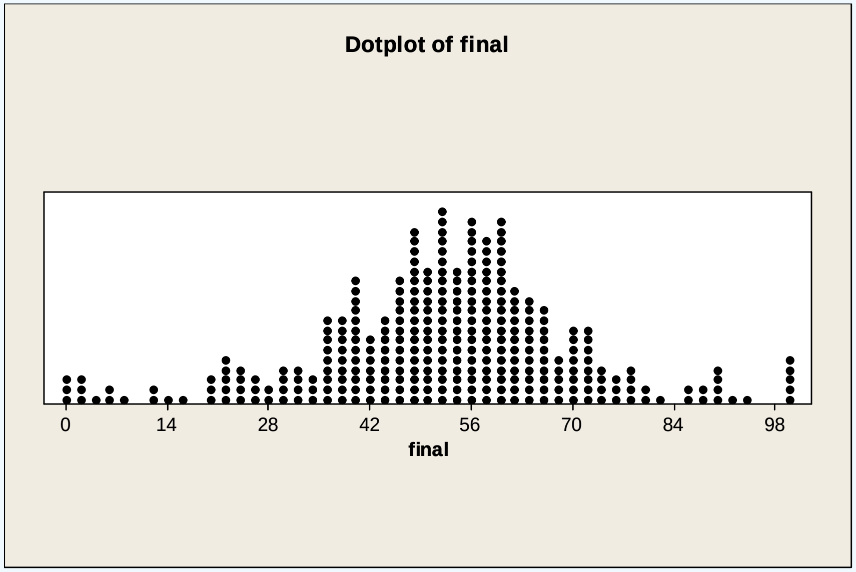
Engineering projects about same standard over past decade.

Areas of choices 2007:
Most popular was transport! 21% on some type or aspect of transport.
17% observational (usually on people); 16% experimental; 12% food/drink;
8% work or study related; 5% each on computing, media, sport, surveys;
3% each on house prices/rentals & on other prices/retail

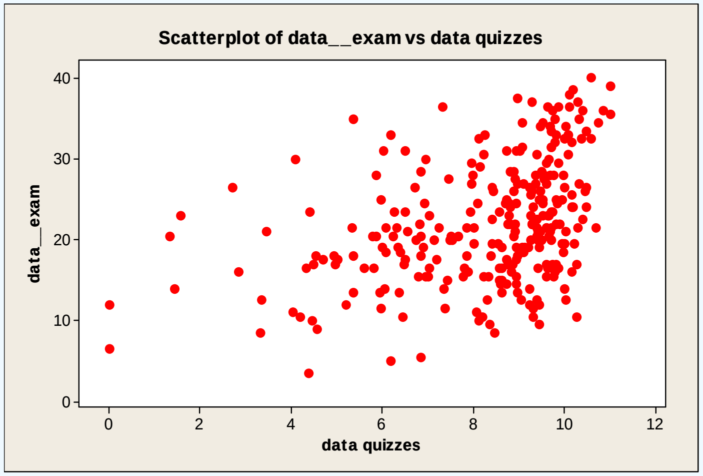
Assessment data: numerical assignment, overall



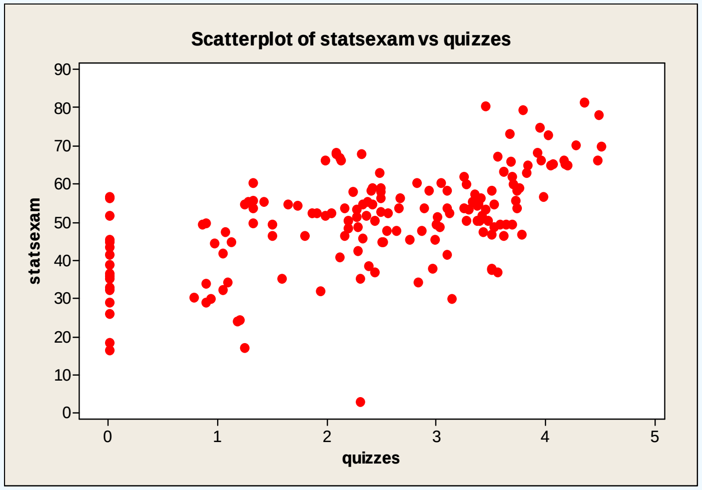
Plot indicates problems – why?



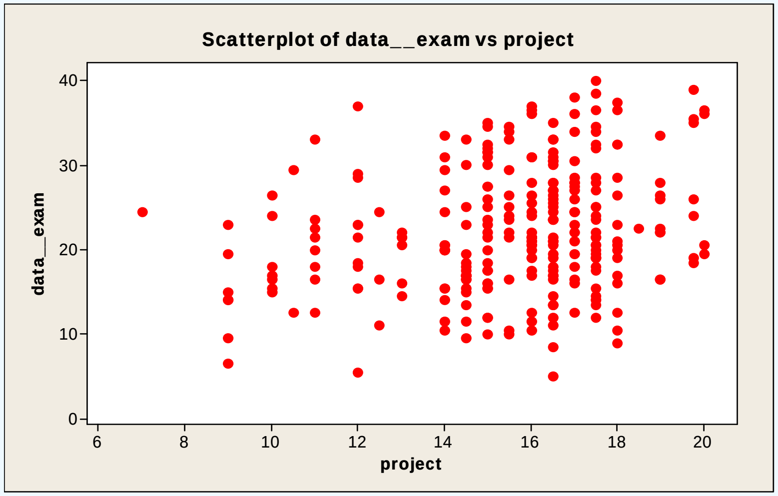
Assessment data: data quests on exam vs data quizzes, project



2007: Relationship good – a bit too much variation



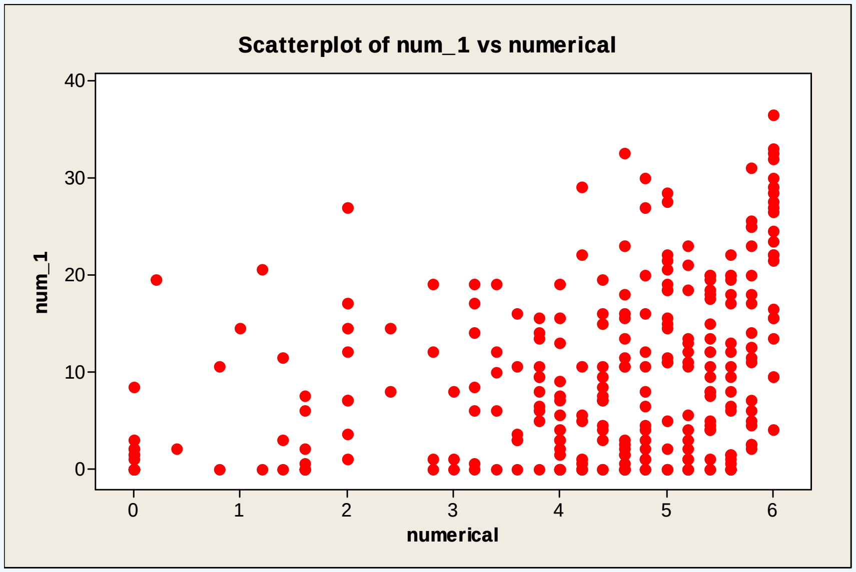
2002, elect engs: excellent relationship. Less variation, partly because half size of 2007 class

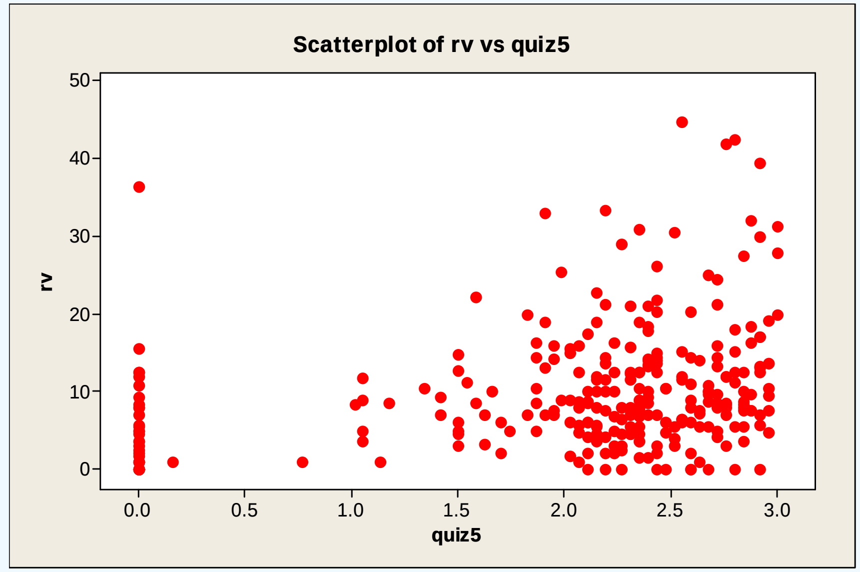


Consistent over years & units; relationship & variation as it should be. Some relationship but project assesses different objectives

Assessment data: num. quests exam vs num. assign; dist'n quests exam vs dist'n quiz

These two need consideration – why?

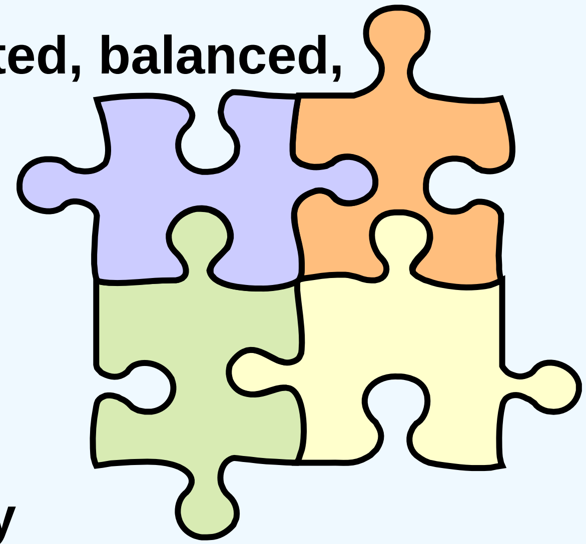




Too much collaboration: why?
Change in engineering course
⇒ inequitable backgrounds
Assignment not difficult but long & detailed
Too much other assessment because of new eng faculty rule

The students seemed to be drowning in assessment in other units weeks 8-11. In weeks 10-12 they tried, with many valiantly doing last quiz. Many were glad to be able to do project weeks 12, 13, but had difficulty engaging with new work.

Conclusions: Assessment for learning



- **Each item/task/component has role in integrated, balanced, developmental, purposeful learning package**
- **Learning objectives ↔ assessment**
  - **What is of value in this item/task/component?**
  - **How do we learn & assess this objective?**
- **Structured for facilitation & management of student learning across the cohort diversity**
  - **What balance of formative/summative does this task have?**
  - **Is this task manageable & correctly weighted for purpose?**
  - **Are the purpose & criteria of task clear within package?**
  - **Do we know enough of students' pasts, presents & futures?**
  - **Have we clearly communicated on collaborative & individual work?**
- **Above assist in preventing plagiarism**

And explore, analyse, interrogate & interpret DATA

Thank you for your attention

Questions, comments, debate,……?

Roles of assessment in learning in statistics and mathematics

Helen MacGillivray
School of Mathematical Sciences
Queensland University of Technology, Australia

Abstract

In the nexus between principles and practice in tertiary assessment in statistics and mathematics, the variety and extent of demands and pressures on assessment packages can sometimes appear overwhelming and even contradictory. Amidst the balancing of formative, summative, flexible, continuous, rich and authentic assessment with demands for criteria and standards-referenced assessment, and 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 statistics and mathematics 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. This paper discusses issues, challenges and strategies in designing and implementing integrated assessment and learning packages in statistics and mathematics particularly in early undergraduate years for both service and core courses. Comments on plagiarism, cooperative and group work are included. Examples are given in both statistics and mathematics, and similarities and contrasts in different assessment packages are highlighted.

Introduction

Calls for tertiary educators to assess what they value, to identify learning objectives, and to align assessment with objectives, appear in both general and discipline-specific higher education literature emphasizing the role of assessment in learning. Although this sometimes causes tensions, it is important for disciplines to be pro-active in analysing, developing and proclaiming the pedagogical aspects of their disciplines, including points of agreement or otherwise with general higher education literature and viewpoints. Assessment is often discussed as if it is an entity in itself, whereas assessment is, and should be, an integral part of a holistic approach to learning and teaching. 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. Naturally students learn for assessment. From their perception, what is assessed must be of value. 'Good' learning and assessment packages in introductory tertiary mathematics/statistics units are integrated, balanced, developmental, purposeful packages with well-structured facilitation of student learning across the cohort diversity.

After some brief comments on literature on assessment in general higher education, and in tertiary teaching of mathematics and statistics, including use of criteria and standards, this paper focuses on four examples illustrating aspects of assessment in different contexts. These examples are from one university but the contexts and characteristics of the student cohorts are analogous to many. The term unit is used here to refer to any module, subject or course for which a student can enrol and receive a grade on completion. Each case includes a summary of the information about the unit and student cohort necessary for any discussion about assessment. The first example is an introductory data analysis unit that is both a service and a core unit, whose central theme is the planning, implementation, analysis and reporting of data investigations. Its learning and assessment package is balanced over components of its objectives. The second example is an introductory probability and distributional modelling unit that is core for mathematics majors but also has enrolments from other degree programs. It progresses through unpacking, analysing and extending concepts, knowledge and skills, with emphasis on linking with real contexts and data. Its learning and assessment package is built around learning problem-solving and modelling skills.

Both these units are first year units. A second year linear algebra unit that is also core for mathematics majors but with enrolments from other degree programs, looks towards industry problems, applied research and computational mathematics. The assessment challenges in this unit are the balance of assignments and tests, and of theory, applications and computing. The fourth example is a second year engineering unit that consists of three components. Half of the unit is introductory data analysis as in the first year unit above, one quarter is introductory theory related to distributions, and the remaining quarter is an introduction to computational mathematics. Hence the challenges are very much of balance – balancing components, workload, objectives and development of skills and knowledge.

Each of the four units includes both group and individual assessment tasks. The collaborative tasks are oriented to particular objectives and are of a nature that calls for group work. A characteristic of good learning and assessment packages in mathematics and statistics is that assessment tasks that carry weight towards the overall result but that are due during semester, are formative *and* summative – these are referred to here as formative/summative. The feedback on such components of assessment is vital for student learning. End of semester components of assessment are summative only. Components of assessment that carry no weight are formative only – in mathematics and statistics such components are primarily for learning and only partially for self-assessment. The overall balance of assessment components and their weights is a key consideration in the gradual development of student skills and operational knowledge.

The combination of qualitative and quantitative information on student performance is important in evaluating the effects of components of assessment in the overall learning package for a unit. In all four examples discussed here, examination of data on student performance in components of assessment assists considerably in understanding strengths and weaknesses of the assessment schema and in the ongoing process of developing aligned assessment strategies and objectives.

Objectives, criteria and standards

The following two quotes represent two important aspects of the roles of assessment in learning in general in higher education

If learning really matters most, then our assessment practices should help students develop .. skills, dispositions, and knowledge…..[1]

Students study more effectively when they know what they are working towards….. Students value assessment tasks they perceive to be 'real' [2]

Such thoughts are reflected within a more discipline-specific approach in the statistics education literature with authors urging care and in-depth consideration of objectives, goals, contexts and content [3], [4]; emphasis on data, statistical literacy and reasoning [5]; and calls for statistics educators to assess what they value [6].

The above quotes and references all entail clear identification of objectives and explicit aligning of assessment with objectives [2], [7]. Aligning of assessment with objectives is an iterative and reflective, rather than sequential, process. Identification of valued assessment tasks helps in the identification of objectives and vice versa. The process involves 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. In mathematics and statistics, particularly at the introductory level, this must also take account of a wide range of backgrounds, programs, motivations and study skills. In principle, this process is not overly difficult, but recent pressures for staff in tertiary assessment can tend to obscure or complicate issues. These pressures include seeking balances and paths amongst considerations such as:

- formative, summative, flexible, continuous, rich and authentic concepts of assessment
- generic graduate capabilities
- work-integrated learning
- criteria and standards-referenced assessment
- higher education fads, generalisations across disciplines and arbitrary rules;

plus challenges such as:

- avoiding over-assessment
- the politics of pass rates, attrition and standards
- the increasing diversity of student cohorts
- the desire of some for “instant gratification”
- managing and controlling workloads of students and staff.

Perhaps it is not surprising that a report of a survey of US statistics educators, [5], comments that of all areas of statistics education, assessment practices have undergone the least reform.

Contrary to the fears of tertiary staff who have been exposed to only the verbal descriptors of criteria and standards referenced assessment, the messages in leading research in this area are consistent with what is regarded as desirable by staff and students in assessment in statistics and mathematics. Marks are certainly valid as measures of achievement against criteria. Whether letters, words or marks are used, the critical aspect is identifying what they represent. In criteria and standards-referenced assessment, it is the *configuration* or *pattern of performance* [8] that enables standards referencing. 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. In [8] the importance of *exemplars* (such as marked past student work, representative assessment tasks and model solutions) is emphasized in identifying the characteristics (or criteria) of each component of assessment, with verbal descriptors to draw attention to salient criteria at different points.

Example 1: Statistical Data Analysis 1

This unit is core in all science or maths degree programs, including education with maths as a teaching area. Approximately 600 students a year enrol over two semesters. The theme of the unit is basic statistical data concepts and tools and using them in real data investigations. The objectives cover two overlapping aspects: tools and building blocks of procedures, concepts and operational skills; and synthesis of choosing, using and interpreting statistical procedures in whole data investigations. The structure, examples and learning experiences are built around real data investigations from first ideas through to report. The content can be seen in chapters 1-8 of [9] and progresses from planning, collecting, handling and exploring data through to ANOVA and multiple (including polynomial) regression, with the procedures of the last two topics done only through a statistical package. The learning and assessment package (with assessment weights) consists of:

- computer-based practicals on datasets from past student projects (formative only)

- worksheets with full solutions (formative only)
- fortnightly quizzes of the fill-in-gaps and short response type, with the best 5 out of 6 contributing 10%
- a workfolder containing the student's ongoing work on the worksheets and their marked (collected) quizzes, contributing up to 3% in proportion to workfolder content
- a whole semester group project in planning, collecting, analysing & reporting in context of group choice (20%)
- an in-semester test, similar to quizzes 1-4 (10%)
- an end of semester exam, similar to quizzes, with more weight on the content of chapters 5 and 6 (57%).

Various aspects of the own-choice project, its impact on learning and teaching, and examples from over 1500 students projects, are included in [10]-[13]. Research on numeracy and its connections with statistical reasoning for the cohorts in this unit can be found in [14], [15]. For a few years there was also an optional essay on how statistics revolutionised science in the 20th century, contributing 10% if the mark improved the student's overall result. This item was dropped because it almost never improved a student's result, and because it tended to attract, and distract, students who could least afford the time. It was an appealing assessment item for staff and some students, but, on the basis of data, we concluded that it was not a sufficiently valuable component of the overall package to warrant student or staff effort.

The own-choice group project teaches and assesses investigation and synthesis of procedure choice and interpretation. It is group because the task needs a group. The students receive feedback on their ideas, and assistance throughout the semester. They have access to past projects and model reports, and past datasets are used in class demonstrations and practicals. Three criteria are given with guidelines, descriptors and standards, and each group receives written comments and marks for the three criteria. These are:

(i) Identifying context and issues, planning and collecting of data, quality of data and discussion of context/problems

(ii) handling, processing, preparing & understanding data & issues; exploring and commenting on features of the data

(iii) using statistical tools for statistical analysis and interpretation of the data in the context/issues.

The students form their own groups, with help from us as necessary. Projects are retained and are designated as 'published' material. Students may refer to past projects if they wish. They know that a copied project is given zero marks. In over 1500 projects, this has occurred fewer than 10 times. All members of a group sign a cover sheet and receive the same mark. Within groups, there are almost never uneven contributions in criteria (i), seldom in (ii), and within-group allocation of tasks assists in avoiding them in (ii) and (iii). Students who emerge as leaders in criteria (iii) tend to learn more, need less revision and do well overall.

For the first semester class of 2006 of which 260 students sat the end of semester exam, Figure 1 gives the plots of the total marks on the end of semester exam paper versus their project mark, and versus their total for the quizzes. The quizzes provide guided learning of tools, procedures and operational skills, and also provide exemplars for the exam. Both these contribute to the moderately strong positive relationship in exam mark versus quiz totals. The greater variation of exam marks amongst the lower quiz marks reflects the mix of abilities amongst those who tend not to do regular assignment work. Although the project helps develop procedural knowledge, it is designed to teach and assess more of the synthesis aspect of the objectives, so the low but slightly positive relationship with the exam marks is appropriate. [Fig 1 here]

Example 2: Statistical Modelling 1

This unit is core in all maths degree programs, including education with maths as a teaching area. It is also taken as an elective by some science and other students. Approximately 120 students a year enrol. The unit builds skills and foundations in concepts and thinking in introductory probability, conditional arguments, distributional and initial stochastic modelling for applications to a wide range of areas, from communication systems and networks to traffic, biology and financial analysis. The unit analyses and extends prior understanding (and misunderstanding); links with data, observation and simulation; and uses, and hence consolidates, foundation calculus and algebra skills in new and different contexts. The whole approach is development of problem-solving and modelling.

The formative components of the learning and assessment package include:

- an initial general probability reasoning questionnaire (PRQ) to seed thought and discussion (introduced in 2004)
- class activities, simulations, selected computer modules, and worksheets with unlimited help
- preliminary discussion points or exercises at the start of each topic to identify and unpack prior knowledge in order to analyse and extend. Development of these was completed in 2005.

More information on the unit and roles of the preliminaries can be found in [13].

The formative/summative and summative components described below, with their % contribution to the overall result given in brackets, are all oriented to problem-solving:

- four assignments based on examples and worksheets, with problems in authentic contexts (20% before 2006; 16% in 2006)
- a group project in which two everyday processes that could be Poisson are chosen in free choice, data collected, and Poisson-ness investigated through tests and graphs (10%)
- an end of semester exam involving problem-solving based on activities, worksheets and assignments, ranging from simple to more challenging in life-related contexts (70% before 2006; 66% in 2006)
- four special tutorials structured for immediate hands-on learning through group exercises, introduced in 2006 (8%).

Students design and bring their own summaries to the exam. For the four special tutorials, students are allocated to different groups for each and full collaboration is required, with groups ensuring that explanations are shared. Assistance is available and credit is for participation, with the group’s attempt signed by all group members. Solutions are provided later and the groups’ attempts are collected but not marked because the credit for this component is for participation. For more information on the group project see [12], [13], and on the new tutorial group exercises, see [16].

Introduction of the tutorial group exercises component was based on observation of the most and least successful learning styles of the students as part of the research and reflect component of the “field” research cycle. It was decided that more active learning was needed to develop skills in 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 in [17], 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*.

Tutors and students voted the experiment an outstanding success. The tutorials were buzzing and early departures were practically non-existent. Other tutorials also benefited significantly through increased attendance and participation. Student opinion was that four was the ideal number. As the assignments provide exemplars for the exam, it is not surprising that analysis, whether in terms of best subsets, forward or backward fitting, shows that they are most important in predicting exam scores. In both 2005 and 2006, in the full models for exam scores in terms of all other variables, only the assignments (p=0.000) and the PRQ score (p=0.009 in 2005, p=0.032 in 2006) were significant. In both years, the relationships of the assignment marks on the other non-exam components of assessment were then investigated. In 2005, significant predictors of the assignments score were the group project mark (p = 0.06) and the PRQ score (p = 0.016). In 2006 the tutorial group exercise contribution was the *only* significant predictor (p=0.004) of the assignments score. It must again be emphasized that the score from the tutorial group exercises is solely for participation, and that they are intended to develop learning and problem-tackling. Hence the above analysis provides further support for the qualitative evidence that the tutorial group exercises are fulfilling their intended roles as significant enablers for all students.

Example 3: second year linear algebra unit

This unit is core in all maths degree programs, and some physics programs. It is also taken as an elective by some engineering students. Between 70 and 90 students a year enrol. The examples and learning experiences in the unit are motivated by higher level needs in mathematics generally, particularly computational maths, and by applications based on experience with industry problems. Although the students are all maths-oriented, there are differences within the cohort, for example between those doing a double degree in maths and business and those doing double degrees in maths and information technology or engineering. The challenges of balancing theory, applications and computing in linear algebra are thus augmented by the cohort diversity. To encourage student engagement and hence student learning, some changes were made in the continuous (that is, during semester) assessment between 2003 and 2005. A number of questions were raised, including: did these changes help or impede student learning, and does the computing help in the theoretical learning?

The assessment package in 2003 consisted of:

- three Maple group assignments totalling 21%
- a mid-semester test worth 15%
- a final exam worth 64%.

The lecturer's observations plus student feedback were that the Maple group assignments were too heavy for 7% each, and that, similarly to example 2 above, students needed more structured help with their learning. In 2005, the assessment package was adjusted to:

- two Maple group assignments totalling 24%
- three quiz-style assignments totalling 16%
- a final exam worth 60%.

The final exams were similar in style, format and level in both years. They have both theory and applications but no Maple use. More details of the content and the assessment can be found in [18].

Analysis of data from the assessment components showed that, for both continuous assessment programs, a Maple group assignment and a 2005 quiz or the 2003 test combined as best predictors of the exam [18]. This provides support of claims in the literature that both theory and computing contribute to overall learning and understanding in linear algebra, and reassurance that changes in the continuous assessment program are not detrimental and do appear to assist in learning. The lecturer had concerns about the grading of the continuous assessment, but could now tackle this with confidence in the program's facilitation of student learning across the theory and practice components of the unit.

Example 4: a second year engineering maths unit

This unit is core in all engineering programs, with 350-450 enrolled in second year programs. The unit was new in 2007 but composed of sections common across previous post-first year engineering maths units. The first half of the unit is as in Statistical Data Analysis 1, and as given in all engineering programs since 1994. The second half is half introductory numerical analysis and half distributions, linear combinations of normals, and introduction to reliability. The level of the unit is first year work in Maths and Science programs, but more compressed. Apart from the compression, its challenge for engineering students is that it is not straight calculus and algebra, and any calculus and algebra that are needed must be at students' fingertips in new contexts. Also, all the statistics is full of new concepts and new ways of thinking.

The focus in all mathematics and statistics units is learning by doing, but this is emphasized even more in this unit. The assessment package consists of:

- for the data analysis half, computer-based practicals on datasets from past student projects (formative only)
- worksheets with full solutions (formative only)
- five quizzes as assignments on the statistics parts, with quizzes being of the fill-in-the-gap or short response type (14%)
- whole semester group project in planning, collecting, analysing and reporting a data investigation in context of group choice (20%)
- numerical analysis assignment (6%)
- end of semester exam, ensuring overall coverage correctly proportioned (60%).

The statistics quizzes are designed for efficient and effective learning, and there is evidence of their value over years and units. The strategy was introduced in the late 1990's in a short Master of Business Administration (MBA) unit with a highly diverse cohort of students holding full-time jobs. It was then developed further in an engineering unit when the data analysis became a half-unit module. The aim was to decrease demand on student time while retaining the full data investigation project. An unexpected and interesting side effect in the engineering unit was a great reduction in copying. Students still worked together on the quizzes but argued with, and explained to, each other instead of copying. Similar effects have also been observed in Statistical Data Analysis 1. The most important aspect of the statistics quizzes for students is the doing of them, but mathematics and statistics staff often worry about the contribution to the overall result of components of assessment in which assistance is provided, and in which students can help each other. As described above, the format of the quizzes has encouraged constructive student interaction. Staff assistance focuses on helping students to find or work out the answers for themselves. Figure 2, illustrating the distribution of the totals for the five statistics quizzes in 2007, shows that although many students obtain, as expected, good scores, very few obtain top marks. The skewness to the left is due mostly to non-submissions. As in example 1 (Statistical Data Analysis 1), there is also a strong positive relationship between the exam performance and the statistics quizzes, demonstrating their value in the student learning. [Figure 2 here]

The group project data investigation is the same as in Statistical Data Analysis 1. This strategy was first developed for the engineering statistics units/modules and has been part of their program since 1995. The engineering data investigation projects have been of a similar standard over the past decade. The projects are both significant learning and assessment tools. Because one-third of the project marks are for the ideas, plan, collection and quality of the data, and because assistance is available throughout the semester, failing project marks are associated with little effort. Sufficient effort and reasonable to good learning tend to produce project marks in the range of 12 to 17, depending on ability. Projects that achieve the highest marks demonstrate statistical thinking, skills and interpretation beyond what could be expected in an introductory unit. Figure 3 shows the project marks for 2007, demonstrating a shape similar to that of the quizzes in Figure 2, and that few projects achieve the highest marks. [Figure 3 here]

In contrast to Figures 2 and 3, Figure 4, showing the distribution of marks on the numerical assignment, indicates problems. Far too many students chose not to do the assignment, and there are far too many high marks, indicating unhealthy collaboration. It became clear during assistance to students and during the marking that the assignment was too long, and its purpose within their learning was unclear to the students. Two additional external problems were that the assessment load in other units was heavy at the time, and, due to changes in the engineering course structure, some of the students had Matlab but most did not. The assignment was designed for Excel but Matlab was permitted if students wished, so that the new engineering course structure made the assignment inequitable. The combination of all these factors tended to lead to an inappropriate amount of collaboration with less learning. [Figure 4 here]

The exam data and its relationships with components of the continuous assessment were also considered. The relationships between the marks on the data analysis section of the exam and the project and the relevant statistics quizzes were very similar to their equivalents in Statistical Data Analysis 1 as discussed above. The relationships between the numerical part of the exam and the numerical assignment, and the distribution questions on the exam and the relevant statistics quiz are shown in Figure 5. It is clear that the numerical assignment did not provide much help with student learning in this area, for reasons as discussed above. The second plot is indicative of “burnout”. Although there is some positive relationship, the main feature of the plot is the increased variation of performance. Indeed, the students had heavy assessment tasks due in other units by week 10, 11 (due to a new “fad” rule in their faculty that 40% of assessment must be completed by week 11), so that even if they valiantly worked through the last quiz, many had difficulty in truly engaging with new work in the last few weeks of semester. [Figure 5 here].

Conclusion

In assessment for learning, each component of assessment has a role in an overall integrated, balanced, developmental and purposeful learning package. Assessment is aligned with learning objectives in an iterative and ongoing process that asks of each assessment component, what is of value that is being assessed, and of each objective, how is this objectives learned and assessed. The learning and assessment package is structured for facilitation and management of student learning across the cohort diversity. Although the examples in this paper are for units in only one university, they have diversity over types of topics, objectives and cohorts. They illustrate how different components can be used and structured so that the combination aligns with an overall set of objectives, balancing formative and summative, group and individual tasks, assignments and tests, and workload. Selection of types and combinations of components of assessment depends on the nature of the unit. Identification of meaningful objectives specific to a unit can help staff in designing learning and assessment activities, and students in understanding their purpose. The importance of exploring, interrogating, analysing and interpreting data is emphasized and used to illustrate aspects of examples that were successful, and some assessment tasks with problems. All examples included group and individual work, with the group tasks being of the type that *need* a group, and being owned by the group. One example demonstrated the use of organised collaborative work to facilitate learning in a problem-solving environment Some interesting indicators emerged of characteristics of "homework" assignments in mathematics and statistics that appear to discourage copying on one hand, and not discourage it on the other. Over-long "homework" assignments whose learning purposes are not clear may increase tendencies to copy, in contrast with more pithy structures which are clearly designed for students to learn core operational knowledge and skills.

Acknowledgement
The author thanks the referee for suggestions that improved the paper.

References

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

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

[3] Hogg, R. V. (1991). Statistical Education: Improvements are Badly Needed. *The American Statistician* 45: 342-343.

[4] Vere-Jones, D. (1995). The Coming of Age of Statistical Education. *International Statistical Review* 63: 3-23.

[5] 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

[6] 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

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

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

[9] MacGillivray, H.L. (2005) *Data analysis: Introductory Methods in Context*, 2nd edition, Pearson Education Australia

[10] MacGillivray, H.L. (1998). Developing and synthesizing statistical skills for real situations through student projects. *Proc. 5th Int Conf on Teaching Statistics*, ISI, 1149-1155.

[11] MacGillivray, H.L. (2002) One thousand projects, *MSOR Connections* 2(1), 9-13.

[12] MacGillivray, H.L. (2005). Helping Students Find Their Statistical Voices. *Proc of the ISI/IASE Satellite on Statistics Education and the Communication of Statistics*, Sydney, Australia: ISI, Voorburg, The Netherlands. http://www.stat.auckland.ac.nz/~iase/publications.php?show=14

[13] MacGillivray, H.L. (2006) Using data, student experiences and collaboration in developing probabilistic reasoning at the introductory tertiary level*, Proc 7th Int Conf on Teaching Statistics*, ISI. http://www.stat.auckland.ac.nz/~iase/publications/17/6B4_MACG.pdf

[14] Wilson, T.M. and MacGillivray, H.L. (2007) Counting on the basics: mathematical skills amongst tertiary entrants, *IJMest* 38(1), 19-41

[15] Wilson, T.M. and MacGillivray, H.L. (2006) Numeracy and statistical reasoning on entering university, *Proc 7th Int Conf on Teaching Statistics*, ISI. http://www.stat.auckland.ac.nz/~iase/publications/17/C136.pdf

[16] MacGillivray, H.L. (2007). Weaving assessment for student learning in probabilistic reasoning at the introductory tertiary level**.** *Proc of the ISI/IASE Satellite on Assessing Student Learning in statistics, Guimaraes, Portugal*: ISI, Voorburg, The Netherlands. http://www.swinburne.edu.au/lss/statistics/IASE/CD_Assessment/index.htm#papers

[17] 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.

[18] MacGillivray, H.L. and Turner, I. (2005) Components of learning and assessment in linear algebra, *Proc of the 5th Southern Hemisphere Conference on Undergraduate Mathematics and Statistics Teaching and Learning*, 74-81

Fig 1 Plots of the total exam mark versus the project mark and the quiz mark in Example 1, Statistical Data Analysis 1, for semester 1, 2006

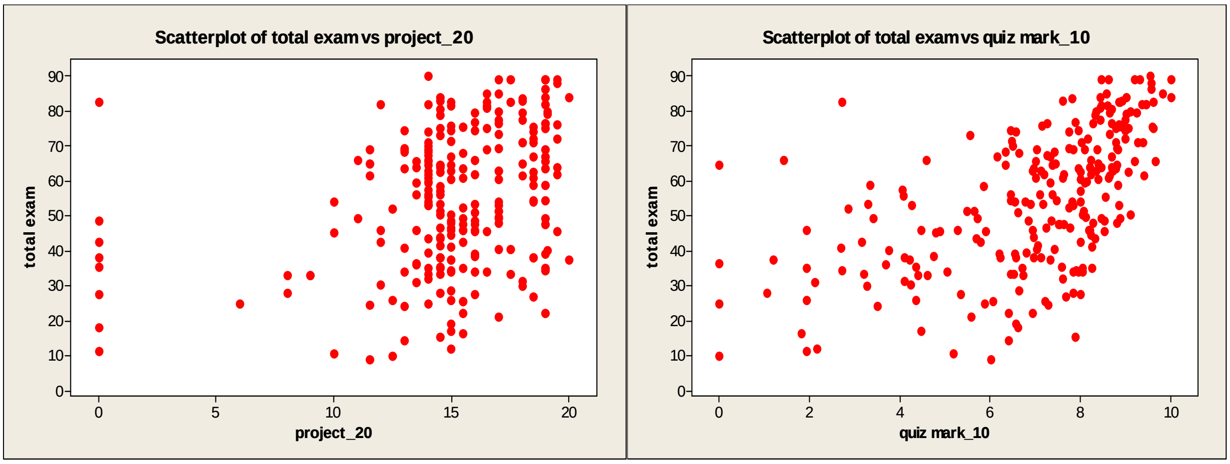


Fig 2 Plot of the total for the statistics quizzes in Example 4, a second year engineering unit, in 2007

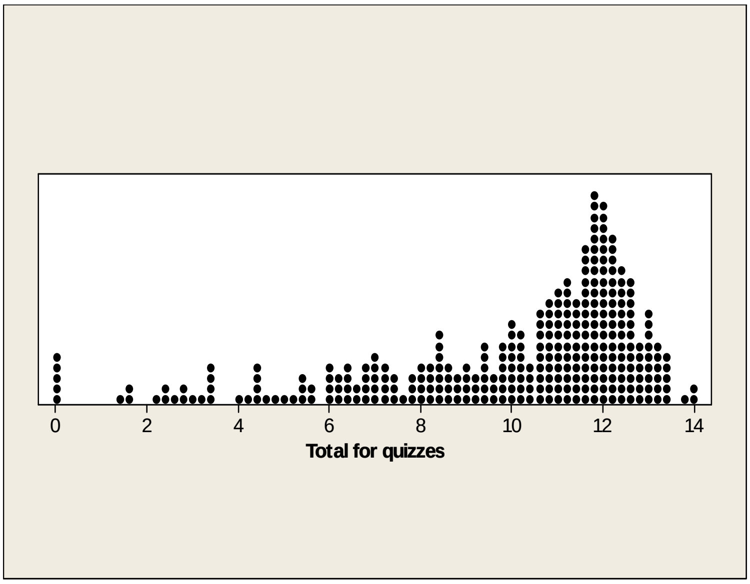


Fig 3 Plot of the project marks in Example 4, a second year engineering unit, in 2007

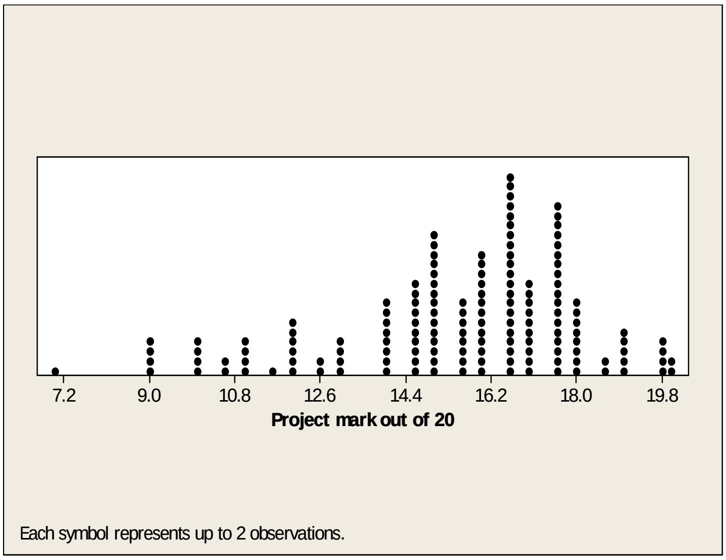


Fig 4 Plot of the numerical assignment marks in Example 4, a second year engineering unit, in 2007

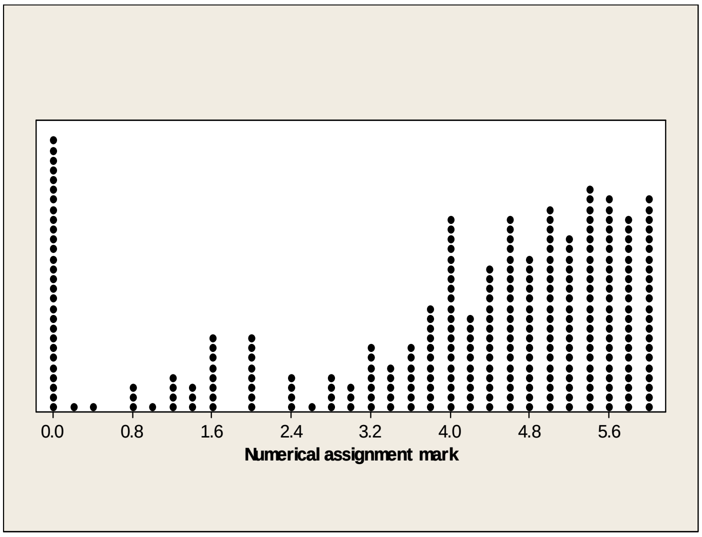


Fig 5 Plots of exam marks versus the corresponding assignment mark and quiz mark in the numerical section and the distribution section in Example 4, a second year engineering unit, in 2007