**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?**

3

# 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: ½ data analysis; ¼ dist'ns; ¼ 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).**

6

# 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

<span style="color:darkred">**integrated, balanced, developmental, purposeful, with structured facilitation of student learning across the student diversity**</span>

- ## 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**

  <span style="color:navy">**For wide range of backgrounds, programs, motivations, study skills……**</span>

# 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**
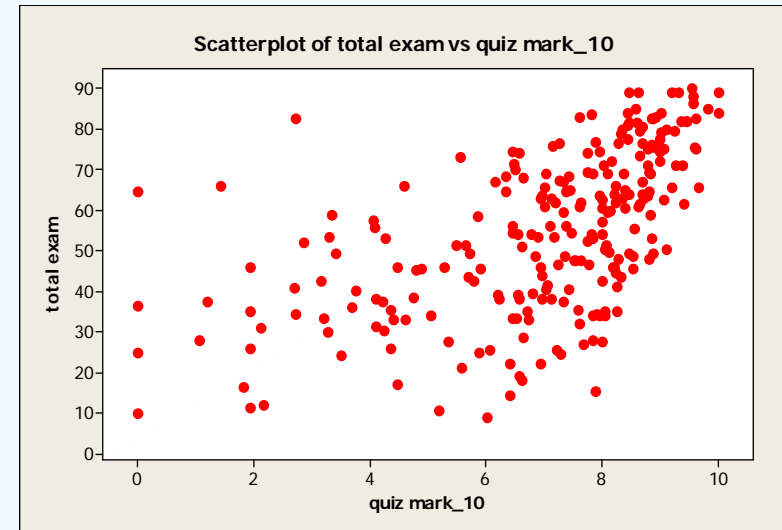
# Just a few recent titles

- 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
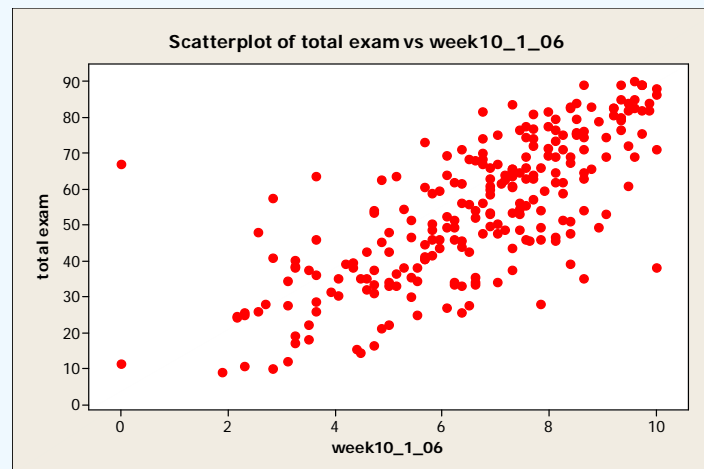- The big news about breakfast
- Music and the people

# Graphs 2006 sem 1


Scatterplot of total exam vs project_20


Scatterplot of total exam vs quiz mark_10


Scatterplot of total exam vs week10_1_06
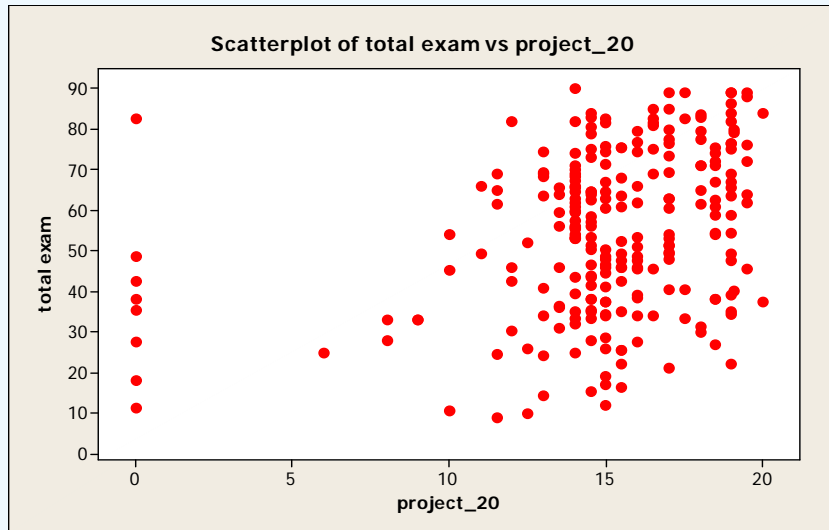
Low +ve relationship ✔

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 1$^{st}$ 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!

# 2<sup>nd</sup> 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: 1$^{st}$ year predictors

Data are complex because of different pathways. But best single predictor amongst 1$^{st}$ yr units, of performance in 2$^{nd}$ 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 1$^{st}$ year units since then have probably changed this

# 2<sup>nd</sup> 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 1$^{st}$ 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

**Dotplot of stats quizzes**
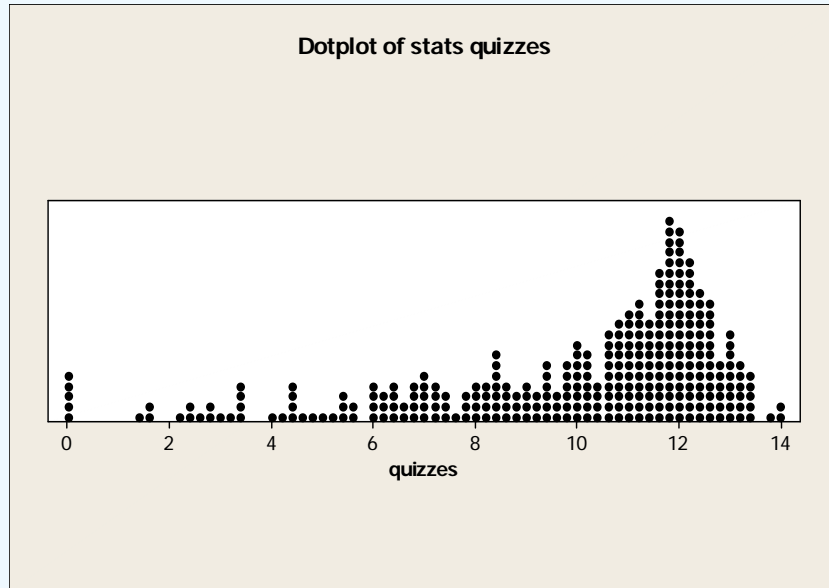


quizzes

**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

**Dotplot of project_20**



Each symbol represents up to 2 observations.
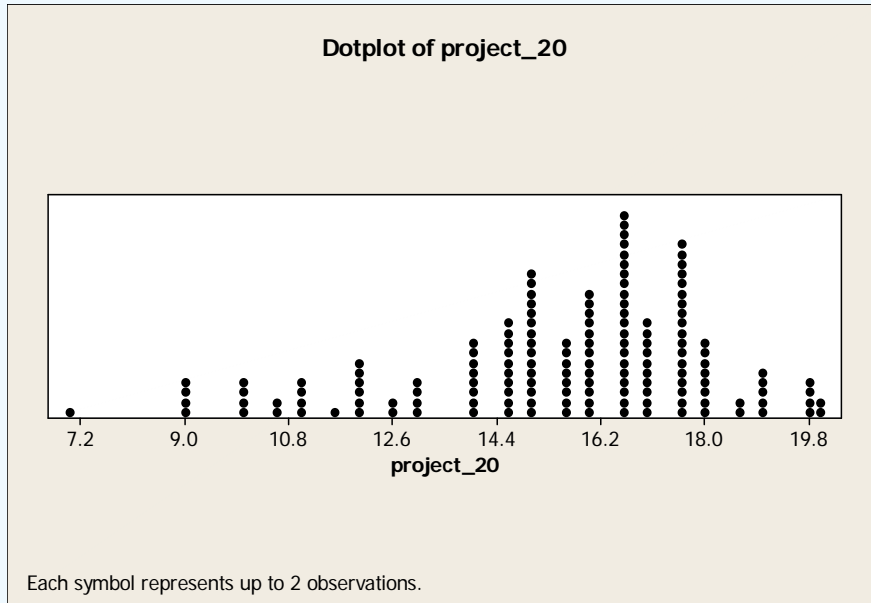
**Project teaches & assesses synthesis of planning, thinking, understanding, choice of procedures and interpreting output.**

**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**
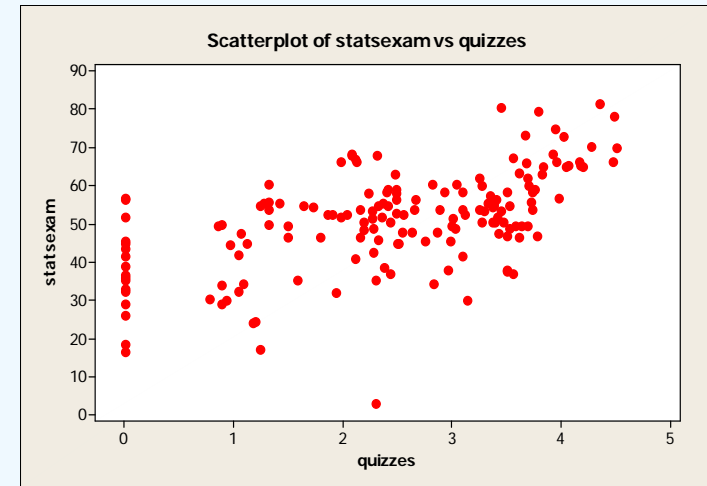
# Assessment data: numerical assignment, overall


Dotplot of numerical
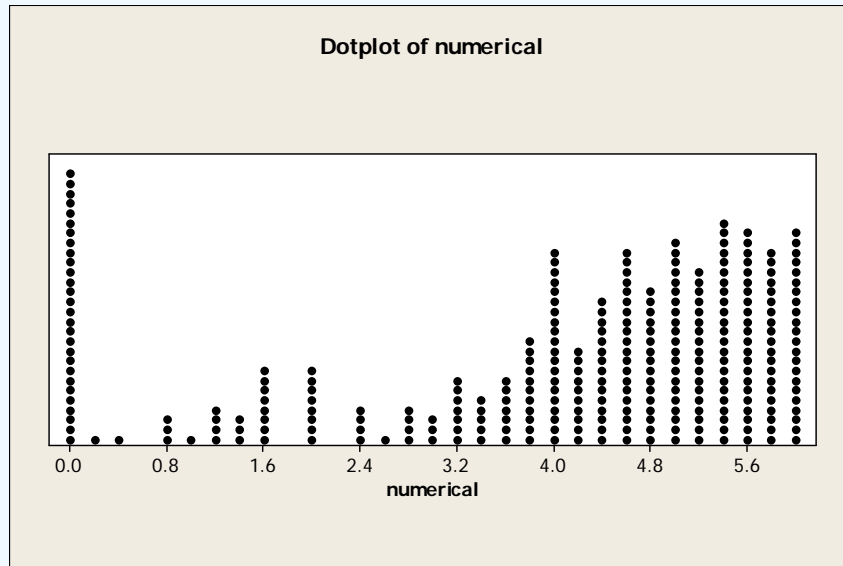
**Plot indicates problems – why?**


Dotplot of final
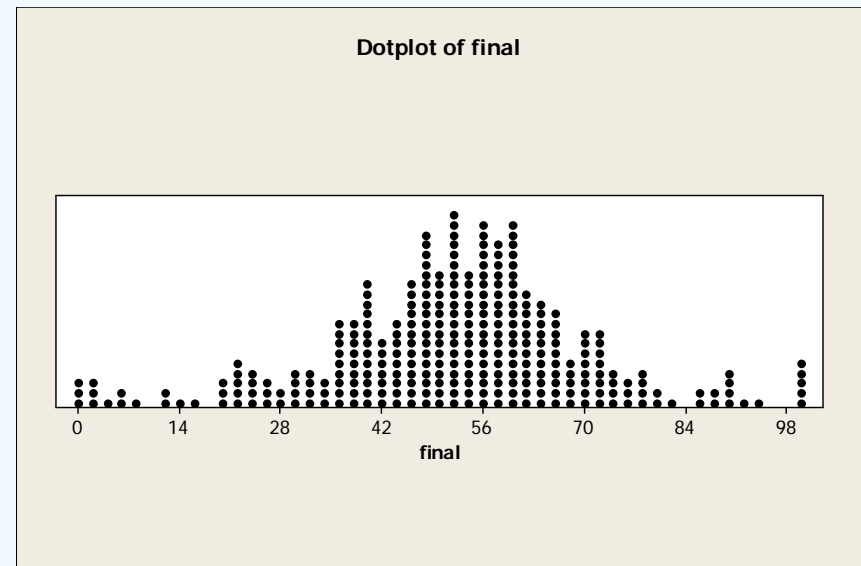
# Assessment data: data quests on exam vs data quizzes, project


Scatterplot of data__exam vs data quizzes
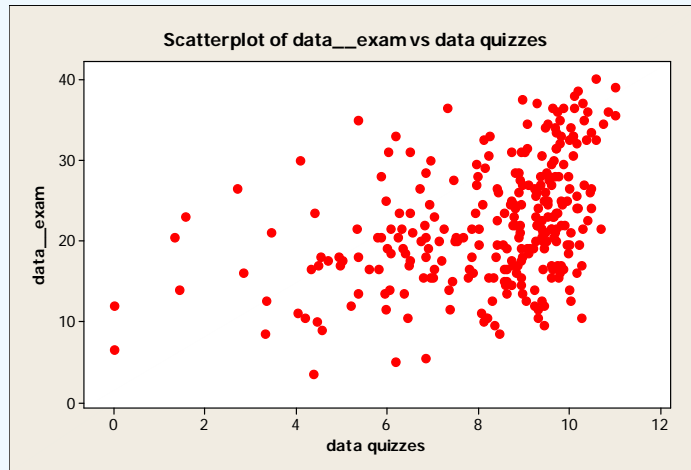
**2007: Relationship good – a bit too much variation**


Scatterplot of statsexam vs quizzes

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


Scatterplot of data__exam vs project
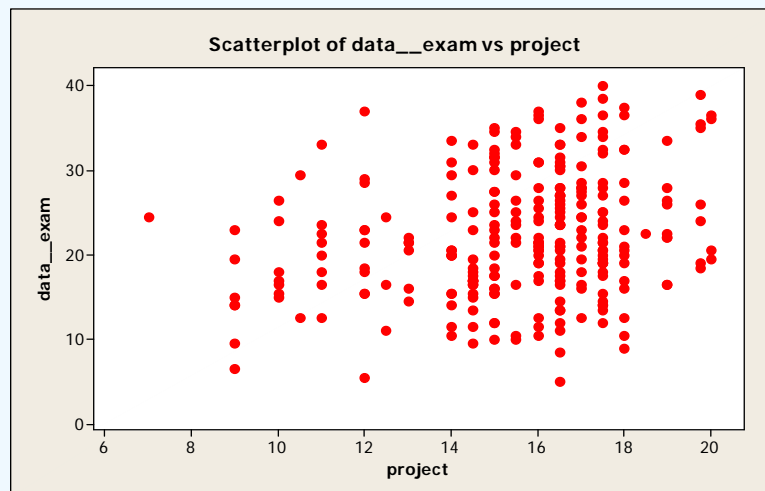
**Consistent over years & units; relationship & variation as it should be. Some relationship but project assesses different objectives**

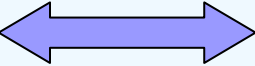**Assessment data: num. quests exam vs num. assign; dist'n quests exam vs dist'n quiz**

These two need consideration – why?
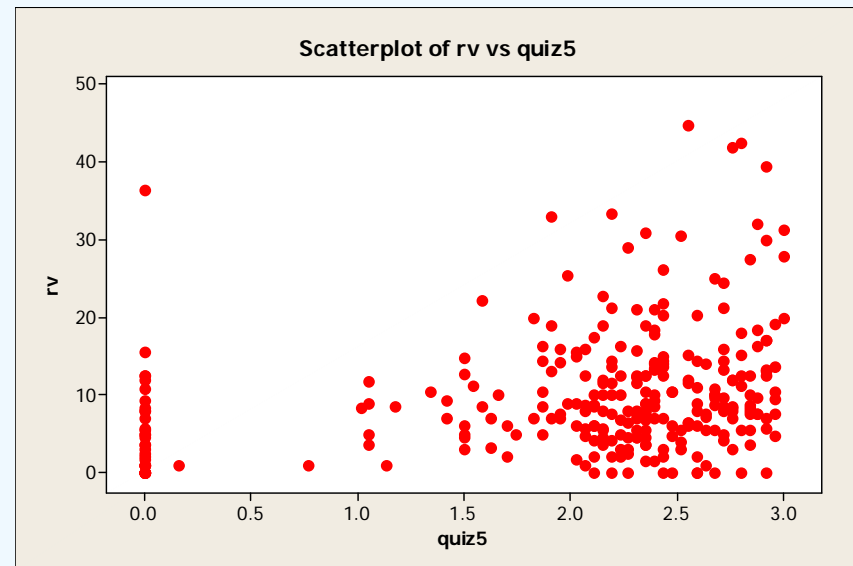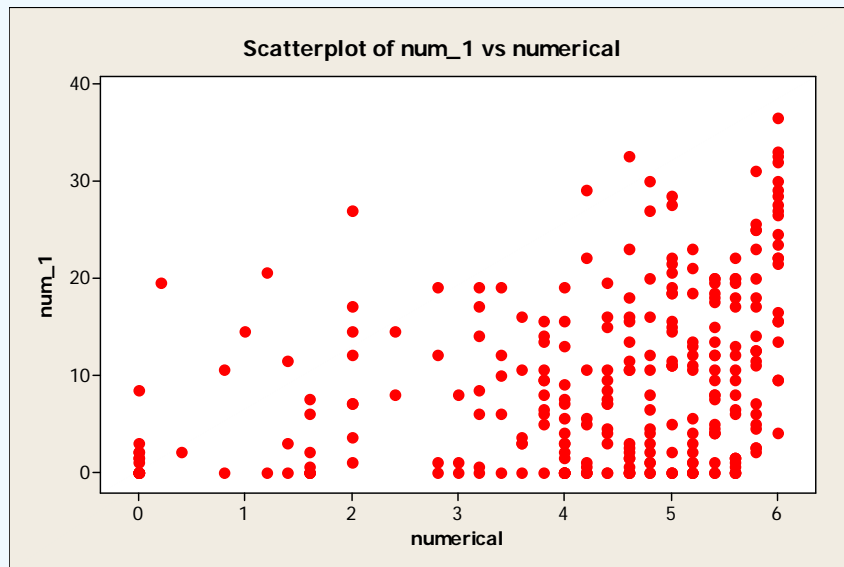
Too much collaboration: why?
Change in engineering course
→ inequitable backgrounds
Assignment not difficult but
long & detailed
Too much other assessment
because of new eng faculty rule

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.

37

# 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**
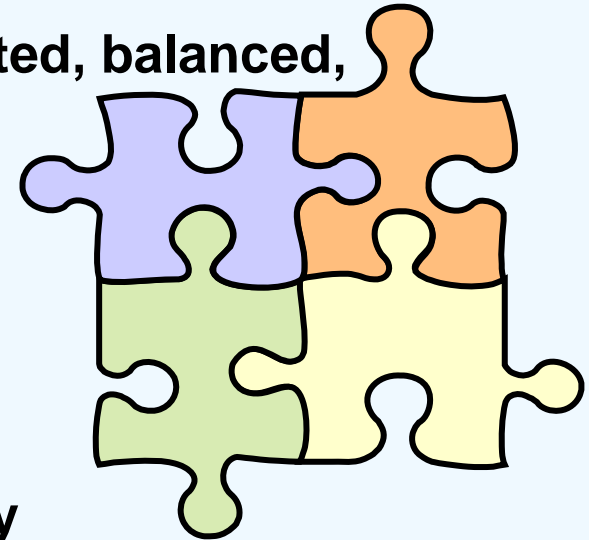
**And explore, analyse, interrogate & interpret DATA**

# Thank you for your attention

**Questions, comments, debate,……?**