

Benchmark



The chiseled horizontal marks that surveyors made in stone structures, into which an angle-iron could be placed to form a "bench" for a leveling rod, thus ensuring that a leveling rod could be accurately repositioned in the same place in the future.

[Wikipedia, "Benchmark (surveying)"]

Picture By JeremyA, CC BY-SA 2.5,

<https://commons.wikimedia.org/w/index.php?curid=493066>

Benchmarking Academic Standards in ICT

Chris Johnson

**Executive Officer, ACDICT
Honorary Associate Professor in
Computer Science, ANU**

10 minute version



Why benchmark learning outcomes? Who cares?

0. *The Regulator: TEQSA - HESF*

1. Students

2. Employers (who represent the public interest, in having graduates as trained workforce)

3. Other Stakeholders (governments, providing a large part of the money; parents/voters)

4. Academics (doing modern education management, within collegiate peer-based traditions)



UK experience – grade inflation

- English universities report grades centrally: numbers of graduates at class H1, H2i, H2ii.
 - *Australian universities do not report grade numbers of pass graduates: HD, D, Credit, with merit etc. terms vary*
- The English accept that there is a problem with **grade inflation:**
 - **Measure: the proportion of H1 and H2i graduate grades has increased.**
- It is too late to revert to previous standard, maybe not too late to halt inflation.
- The system of benchmarking by external examiners appears part of the problem.

Is there a problem of grade inflation in our assessment?

If pass rates improve, or if the proportion of High Distinction grades increases – is this because of

- better teaching 😊
- better learning
- better innate abilities of students

How would we know if there has been an improvement across the discipline?

Is there a difference which university a BIT comes from?

Is every pass level degree at equivalent standard? Are HD grades the same?

How would we know? What can we improve? Whose standard is “right”?

- University of [Go8]
- [Technology Network] University
- [Independent Research] University
- [Regional Network] University
- Non-aligned ex-advanced education college university
- University of somewhere / something



Benchmarking academic standards

- There is rising **global interest** in academic quality and the development of **valid indicators of graduate outcomes**
- **Existing measures of quality:** student feedback, experience, and satisfaction surveys, self-reported learning outcomes
 - These all have a questionable relationship to **educational outcomes**

Benchmarking academic standards (2)

- **Other Proxy measures:** graduation rates, employability, success in further higher education study, entry cutoffs (QILT)
 - These are crude measures that cannot be related to education, rather than input qualities or employment conditions; unreliably vary with location in place and time
- **AHELO:** OECD Assessment of Higher Education Learning Outcomes project. An international trial of standardised exit testing of graduates in a discipline
 - *trialled but not accepted 2015.*

Academic standards

- The aim has been to define standards for graduate learning outcomes – not for “content”
- Recent projects in Australia: Threshold Learning Outcomes for broad disciplines [OLT 2010]
 - Engineering and ICT combined
 - 5 items
 - no distinction made between graduates at AQF 7 (ICT) and AQF 8 (Eng)
- Compare with established UK QAA processes: Subject Benchmark Statements (SBS)
7-8 pages



Why benchmark? – 1. students

- “For students, assessment standards provide guidance for their learning and allow them to monitor their progress, and ultimately, the standards will be used to judge their performance.”
 - (Price et al, 2008, in Rust)
- *Better informed students = better engaged students = better learning by students*

Why benchmark? 2. employers

- interest in warranted skills of graduates as trained workforce recruits
- Employers perceive poor work readiness of graduates (anecdotally)

and at the same time

- **Employers do not report any big lacks in technical knowledge and skills (systematic surveys)**

Why benchmark? – employers.2

Do the TLOs provide what the employers want?

– *ask them* ! [2009 ALTC project, Koppi and Nazhdy]

- Employers say: graduates generally **meet needs** in relevant ICT knowledge, theoretical principles, literacy, numeracy, computer languages; the balance of fundamentals vs technical is mostly OK
- but graduates **lack**: commercial awareness, some lack Project Management, written communication, knowledge of business processes
- Recent graduates say: we lacked interpersonal skills, business and industrial knowledge, want WIL



Why benchmark? –

3. government as stakeholder

- Higher education is a large chunk of the Commonwealth budget: want to managing the cost while increasing the quality and quantity
- Interest in numbers of graduates as trained workforce, in the national interest
- Response to concerns of employers and students' parents as vested interests/votes

Why benchmark? –

4. academics

- Part of Academic Quality Improvement processes (measure it in order to manage it)
- Do quality management in academic hands rather than government's
 - build on tradition of academic peer review – collegial approach
 - do not expose an external numerical or formulaic performance indicator to government interference
 - note the government proposal/threat to use pass rates/retention as KPI for \$ (NZ 2016, Australia 2018)

Desirable requirements for benchmarking methods

- **Economical** in academic effort
- **Effective** in determining quality in ways that can measure improvement
- **Responsive** to stakeholder needs and **convincing** to stakeholders
- **Reliable**: provides an indicator that correlates with actual (perceived) quality, does not drive behaviour in unwanted directions
- Acceptable to university academic **principles**: non-intrusive, independence, collegiality



Possible Methods to benchmark

1. Define common standards collectively across the discipline, then have each university do internal comparisons and internal reviews
 - Unreliable, difficult, *unconvincing*
2. Appoint external examiners throughout subject delivery and assessment
 - Expensive, unreliable, *Intrusive*
3. **Peer external academic review** of indicator subject results, after results
 - feed responses into normal review/improve cycle
 - reliable, cheaper, **less** intrusive



3. Peer external review (2)

Appoint external peer academic(s) who review assessments in selected subjects, after results decided: for

1.degree of fit of designed assessments to subject's designed goals (learning outcomes) and rubrics;

2.internal compliance of grading with rubrics;

3.match of design and achieved assessment standard to own experience

using selected **indicator** subjects, such as:

1. Final individual project (as done in Go8 Engineering)

2. Capstone project (ICT)

3. Common "Advanced" subjects

4. First year introductory programming subject



Possible Methods to benchmark

1. Define common standards collectively across the discipline, then have each university do internal comparisons and internal reviews
 - Unreliable, difficult, ***unconvincing***

Possible Methods to benchmark

1. Define common standards collectively across the discipline, each uni do internal comparisons
 - Unreliable, difficult, *unconvincing*
2. Appoint external examiners to participate throughout subject delivery and assessment process (UK model)
 - All subjects. Examiner acts to moderate results on the spot, has indirect effect on processes. Expensive? inconsistent? unreliable *Intrusive*.

Possible Methods to benchmark

1. Define common standards collectively across the discipline, each uni do internal comparisons
 - Unreliable, difficult, *unconvincing*
2. External examiners participate throughout subject delivery and assessment process
 - All subjects. External acts to moderate results on the spot, but no change in processes. Expensive? inconsistent? unreliable *Intrusive*.
3. **Peer external academic review of subject results, after assessments completed**
 - Use selected indicator subjects, feed responses into normal cycle of review/measure-improve. Cheaper?



3. Peer external review (1)

Appoint external peer academic(s) to review assessments in selected subjects, after results have been decided, to evaluate:

1. How well the subject as delivered and assessed fits to the stated subject and graduate learning outcomes, and to **common academic standards**: questions, rubric standards, submitted work, marking quality
2. Internal compliance of assessment to marking rubrics
3. Standard of grading compared to other institutions – the community

Formative feedback report.



3. Peer external review (2)

External peer academic(s) who review assessments in selected subjects, after results have been decided
fit, compliance, assessment standard

using selected **indicator** subjects, such as:

1. Final individual project (as done in Go8 Engineering)
2. Capstone project (can apply to ICT, given the ACS accreditation requirements)
3. Common “Advanced” subjects
4. First year introductory programming subject



Issues to consider

Is the benchmarking method

- **Efficient** to execute once and repeatedly (economical)
- **Effective** and **responsive** to make comparisons of the quality of graduate outcomes achieved, effective as a driver of **improvement**
- Give **convincing** results for stakeholders
- provide **reliable** indicators after universities respond with action (cannot be gamed)
- Acceptable to **academic principles**



What to compare? 1

Final individual project

- Common, similar in all 4 year Engineering programs
- Students' final report demonstrates achievement of the Threshold learning outcomes: abstraction, modelling, problem solving, communication in writing (and possibly verbal)
- Student grade in the project closely aligns with degree honours grade = holistic quality indication
- *But* the individual project subject is not common in 3 year CS or IS programs



What to compare? -- 2a

ICT Capstone project

- Common to all CS and IS (ACS accreditation)
- The capstone project develops and assesses TLOs in needs, problem solving, modelling, communication, self management, work relevance
- Integrates technical learning over program (*but* this is indirect, usually not directly assessed)
- Requires teamwork, relation to industry
- Contexts vary between universities (open-endedness, relate to industry client, ethics)



What to compare? -- 2b

ICT Capstone project

- Ways of assessing projects vary widely
 - process reports, minutes of meetings, design documentation, individual reflective journals, client feedback, mentor review, presentations, software artefacts, quality of user experience, design rationale
 - **Larger volume of assessment** than most subjects
- Assessment of the team project subject often makes a weak distinction between team and individual (can an individual fail in a good team? How can I star?) – flat grading profile, weak as an overall indicator



What to compare? -- 2c

ICT Capstone project

- Con:
 - The volume and variety of assessment material makes comparison assessments expensive or partial
 - The difficulty of teamwork vs individual work makes some universities hard to compare
 - Assessment is holistic and indirect on some LO s
 - Assessor may be unfamiliar with project topic
- Pro:
 - The capstone project includes demonstration of many of the graduate learning outcomes
 - Project assessments indicate quality well enough to confirm the assessed grade in whatever topic



What to compare? -- 3a

Selected Advanced subjects

An “advanced” subject [ACS definition]

- Has a prerequisite chain (*but* some universities favour *assumed knowledge*) (implies is in second half of program)
- Has a majority of LOs at Higher Bloom levels: (apply), create, evaluate, analyse (*but* some universities not yet using LOs)
- all programs have some Advanced subjects [ACS accreditation]

What to compare? -- 3b

Selected Advanced subjects

Rationale for benchmarking advanced subjects

- Reliable indicator: internal moderation in the university ensures that all core subjects will reflect the overall standard graduate standard (is this true?)
- selected advanced subjects can be used as a proxy indicator (*if unperturbed*)
- Select common subjects, assessor can be familiar with subject matter at same level

What to compare? -- 3c

Selected Advanced subjects

Pro:

- a small number of comparable subjects can be assessed on written materials, cheaply by lecturers with existing expertise
- subjects with comparable content are common over unis (*?test this assumption*)

Con:

- non-technical graduate outcomes vary in these subjects or are absent in the formal exam, assessed mainly in assignments

What to compare – 4.

Introductory programming

- Common, similar learning goals everywhere
- Good volume of research literature
- Easy to compare
- But – goals and teaching qualities are distant from later years and from degree program outcomes
- **At too low a level to be useful**

Effectiveness: Quality improvement

- A classical quality improvement cycle: measure-analyse-improve *is implied*
 - 1.Measure:** Review, compare and report
 - Internal or external evaluation and reporting
 - 2.Analyse:** curriculum & teaching methods
 - Personal or institutional reflection and review
 - 3.Improve:** curriculum & teaching methods
 - Redesign, modify, approve; develop teaching skills, apply to practice

Selected Advanced subjects

- Can we find comparable subjects
- Pick similar content using ACM curricula for CS (strong) and IS (weak)
 - Computer Networks
 - Algorithms and Complexity
 - Advanced Database
 - Enterprise Systems
 - Data and Information management/ Business Intelligence/ Analytics/ Datamining

