

ASSESSMENT OUTCOMES: FACTORS INFLUENCING THE DISTRIBUTION OF MARKS

Dr Paul Alexander, Tertia Worst

Curtin Business School, Curtin University of Technology
Perth, Western / Australia
p.alexander@cbs.curtin.edu.au; t.worst@cbs.curtin.edu.au

Abstract

Assessment is an integral and crucial part of university education. Much research has been reported on its different forms, its outcomes, and the means of maximising its reliability and accuracy.

Assessment outcomes must meet the needs of several diverse groups of internal and external stakeholders. Students require suitable grades and expect equity and reliability of marks. Employers require evidence of competency. Assessment statistics are a health indicator for courses; they impact community and professional perception of courses and graduates, and are key to teachers' performance management. Universities use assessment statistics as an input into rankings and they even impact the institution's "brand". Such a diverse range of objectives of assessment outcomes may therefore generate conflicting forces which can impact the statistical distributions of marks.

This paper is an exploratory study to determine the factors influencing the marks distribution patterns of a series of 37 units delivered as part of a logistics course over a period of 18 months.

It finds clear factors such as quality of student intake, methods of delivery and differences in teaching and content all influence the patterns, but there are statistical indicators of more subtle influences, generated by the needs of the stakeholders. It also leads to clear directions of further research and a possible model, which focuses on organisational culture, policies and motivations of assessors and the assessment process and their interaction.

Keywords - Assessment, marks distribution, assessment influencing factors, assessment policy, university.

1. INTRODUCTION

Assessments are often expressed in terms of marks comparisons to expected models or historical norms. General assessment purposes and their role in the pedagogical system are well understood by Schools, but when specific assessments are set for individual units, the exact role they may have is not as clear.

While there are myriad assessment forms, each with well documented characteristics, benefits and shortcomings, in terms of their fundamental aims they can be seen as achieving one or more of three main objectives.

1.1 Assessment for measuring performance

The first objective is to allow measurement of students against standards. Such standards are seen as crucial and operate at several levels [1]. They can inform, and indeed are often the basis for planning and delivering instruction, and for reporting that to the public [2]. In this they intertwine with policy making and as such may become an emotive issue. For example, making literacy scores identifiable by school on web sites in Australia has resulted in a major debate about the public's right to rank schools versus complex causal issues. The result has been a clash between Government, teachers and other stakeholders [3]. Clearly assessment as a performance measure has powerful and widespread acceptance across society as a whole, but not necessarily the same meaning.

1.2 Assessment to rate competence

At an operational level, assessment may also be used to ensure student competence [4] and as a measure of success of teaching efforts [5]. This results in many assessment strategies that ultimately must make sense to students, teachers and parents [4]. These strategies must serve a motivational as well as a cognitive, informational purpose and also form the basis of a fundamental shift in practices and learning expectations.

1.3 Assessment as part of the learning process

Much recent focus has been on assessment not just as a tool for measuring, but as a means of learning itself.

This objective can actually be complex in its outputs. The crucial distinction is between assessment to determine the status of learning and assessment to promote greater learning [5].

Learning assessment requires a continuous, iterative flow of information about student achievement that feeds back into the learning process, and not merely to check on it. Underpinning this are several linked strategies [4] shown in fig. 1.

Learning assessment is more wide ranging than assessment for measurement, since it is embedded deeply into courses and all stages of the learning process.

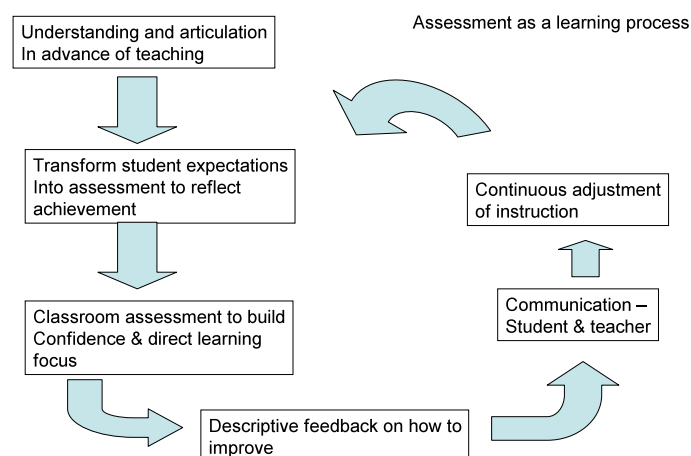


Figure 1: Assessment as a learning process

1.4 Assessment and stakeholders' expectations

Evident and integrated into assessment objectives are the collective aims of many stakeholders, each with diverse and potentially inconsistent requirements. In designing assessments, the needs and priorities of these stakeholders must be balanced.

In commercial terms stakeholders are typically defined as "an individual, group, or organization impacted by the decisions and actions of an organization" [6]. In the context of tertiary assessments, these stakeholders include students, teachers, schools, universities or colleges (marketing, branding and financing arms), employers, the community at large and certifying bodies to name but a few. Such a diverse group has different demands on an assessment system. What's more, contemporary course structures must take all this input on board in both design and operation of the assessment system. A modern course is rarely created in an academic vacuum and must therefore fulfil many requirements; from serving as a competency guarantee for employers to a viable revenue stream for universities.

Biggs [7] recognises this in a general sense and sees assessment as a balance between "what is likely", "what we want", "what is possible" and "what is allowable" (fig. 2).

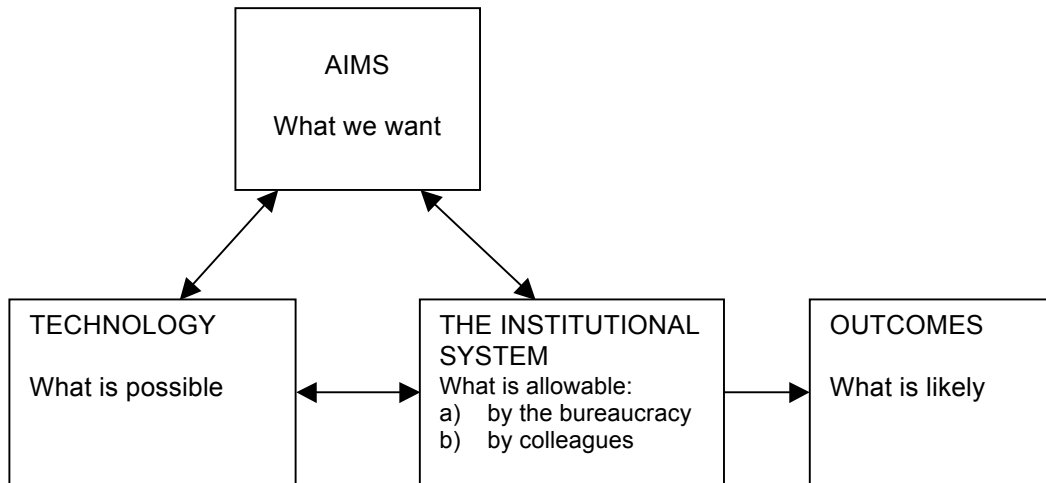


Figure 2: Balancing different requirements in tertiary education, (after Biggs [7])

Biggs also notes that other factors can be important in how these interactions occur. For instance quantitative marking systems, which are adequate for construing some kinds of learning and which are more readily seen as objective; and qualitative approaches, which should be the appropriate framework for enabling decisions flowing from most tertiary educational aims, but which may be open to interpretations that are different for different stakeholders.

Not only do external factors impact the nature and objectives of the assessment but the assessment itself may be important in the social order of tertiary institutions. There is an interaction between the assessment and the organisation itself. For instance, administration effort and policies are linked closely to the specific assessments carried out and may compromise or otherwise modify the formats being chosen. Accreditation and management functions are simpler for more standardised, quantitative forms of assessment.

Measurement of assessment is almost universally expressed as “marks”, “percentages”, and “grades”, with a tacit acceptance of how that impacts different stakeholders. They create expectations in students, employers, schools and other stakeholders. 49% (a “fail”) for instance is seen as poor, while 50% (a “Pass”) is seen as acceptable, though the validity, interpretation and importance of these will be perceived differently by different groups.

Such measures may also influence the learning cues provided to students. A test making up a high proportion of an overall unit assessment may be seen as higher priority by students than a low proportion test regardless of the contribution to learning that test is to make [8]. This may also result in digression of students' perceptions from teachers' intentions of what is required to be learned.

There is also an inertia created by these multiple stakeholders, since changing assessment systems at any level will be met with different enthusiasm by different stakeholders [9]. Thus, a change enthusiastically embraced by one group of stakeholders may be resisted by others. However, Biggs [10] notes that the rise of academic audits and adoption of institute wide performance indicators does make practices and approaches visible to many of the stakeholders and creates a dynamic tension that can at the least make measures understood better by the stakeholders, and may also lead to reform and realignment to agreed goals.

1.5 Perceptions of accuracy in assessment

Much work has been devoted to accuracy in assessments [1, 11-14]. Assessments inherently suggest there is a standard reference point which is “the truth”. However, different stakeholders have different interpretations of this truth, and this therefore leads to questions of the accuracy of an assessment in that context. An assessment policy that is used to support filtering (for example, failing the lowest 15% of a class) will result in perceptions of inaccuracy when seen in terms of demonstrating competency; in this example those in the failed group may nevertheless have demonstrated suitable competencies.

1.6 Assessment as a balance of interests

So, it can be argued that a specific mark achieved by a student, its perceived accuracy and its very meaning is the result of a complex interaction between teacher, students, curriculum, teaching and assessment methods and institutional systems. While it may be tempting to therefore assume a subjectivity of the whole process, it is nevertheless valuable and possible to determine the contribution of the underlying factors which can, if not providing a positivist standard, then at least recognise the differences and also the stakeholders.

Such an approach requires an exploration of assessment outcomes from the perspectives of the various stakeholders.

2. RESEARCH METHOD

This research is an exploratory study to determine the factors that may contribute to particular assessment patterns. Analysis of a sample of 37 results from different units conducted in the logistics discipline at an Australian University over a period of 3 semesters (18 months) across 3 campuses by six lecturers was performed. Units are semester-long subjects undertaken by students. Each must be passed in order to receive credit. In all cases, three forms of assessment are required. One of these must be an exam worth 40% to 50% of the overall score. Other assessments vary from continuous assessment to large assignments, and typically include a group and an individual assessment.

2.1 Distribution of grades

For each unit, students are assigned a mark and a grade according to the system shown in table 1. Some grades carry particular significance to the student and the University. For instance, a F-in or F often requires the student to repeat the unit with attached time and money penalties. Other significant mark are 7 or 8, which are considered distinctions and carry significant kudos for the student and also can serve as entrance requirements into some units.

Each of these events flowing from significant marks also requires specific processes to be conducted by both the teacher and administration resources. For instance an F with a near pass mark may result in appeals which require considerable time to complete.

Table 1: Grading system

Grade	Mark	Application
F-In	0-99	Fail – Incomplete / Insufficient assessment; discretionary use where a student has not completed or submitted all required assessable work for the unit.
F	0-99	Student has failed a unit classified as a Grade/Mark unit
5	50-59	Student has passed a unit classified as a Grade/Mark unit and achieved a percentage mark in the range specified without supplementary assessment or without a pass having been conceded
6	60-69	
7	70-79	
8	80-89	
9	90-99	
10	100	

Marks are assessed independently by the teacher but must be cross checked by a co-examiner and by representatives of the school collaboratively via a board of examiners. Such checks are to provide consistency and also to identify “outrider” distributions of results.

For each unit distributions of the grades of all students were examined for patterns to elucidate underlying factors. Sub-groups of undergraduate versus post graduate units, and a comparison of lecturers was also considered in order to expose other potential factors. The two lecturers with extracted assessments were asked to comment on the distributions of their unit marks.

No attempt was made to analyse the data statistically and a focus was maintained on patterns.

2.2 Factors influencing assessment

An exploratory investigation to find various influencing factors was conducted. This consisted of a search for tangible policies, statements, processes, practices, marketing statements, and other sources.

These factors were assessed as to their possible influences on the assessment distributions observed.

3. RESULTS

3.1 Overall marks distribution patterns

Two different patterns are evident (fig 3.). A bi-modal pattern is seen in 27% of the units while the remaining show a single mode of distribution with characteristics approaching a normal distribution.

Fail rates, and the significant impact and administrative overhead they carry are low in the mono-modal units, though higher in bi-modal units; in some as high as 30%. These high-fail units are subjected to extra scrutiny which involves a lecturer's report and further comment from the Head of School. Only 12% of units exhibited such a distribution.

Exam marking requirements also require extra administration in the form of head of school scrutiny and approval if more than 40% of students exceed a grade of 7. This occurs in 29% of all units.

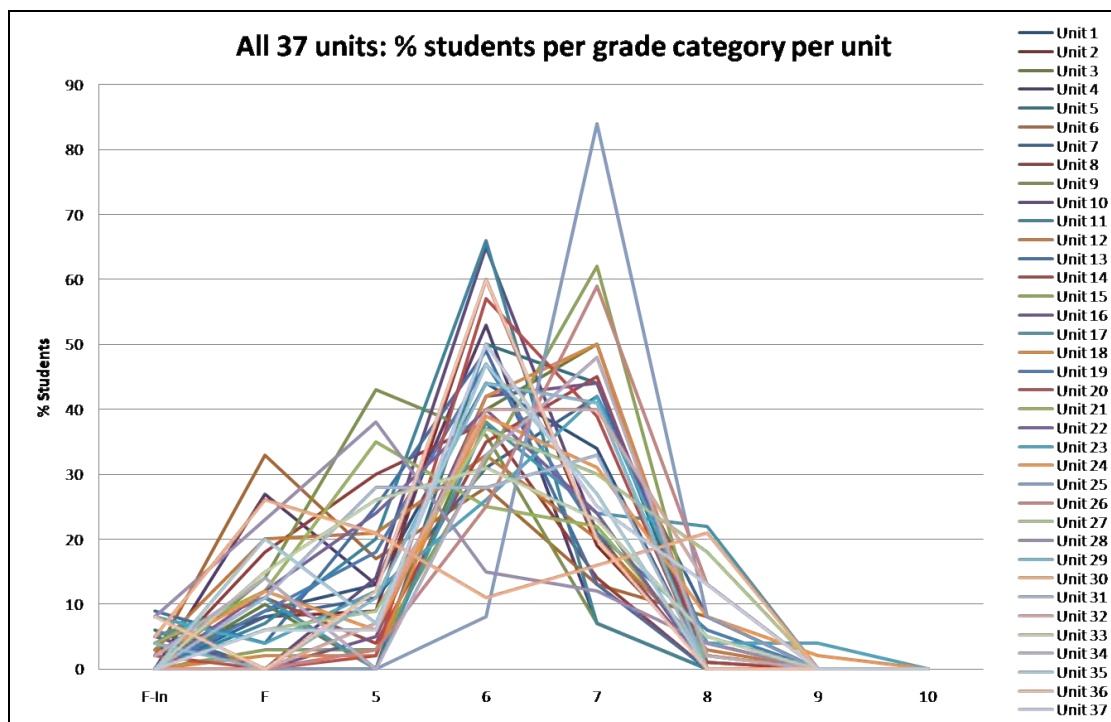


Figure 3: Grade distributions of logistics units

3.2 Differences between lecturers

Two lecturers teaching more than 10 units over the period of the study were selected and distributions of their units extracted (Fig 4a and b). As lecturers are stakeholders in the assessment process, and they themselves have a complex set of goals, such an analysis can yield extra information.

The two distinct patterns (mono- and bi-modal) noted in the overall units are again present for both

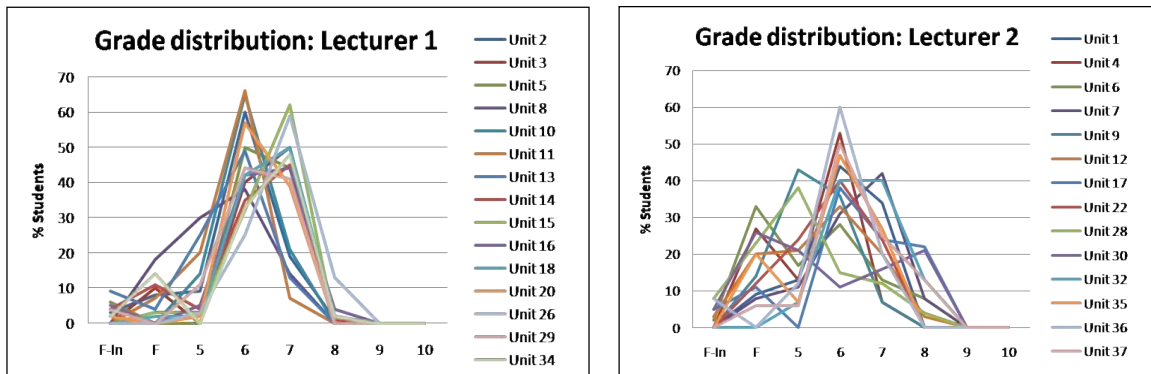


Figure 4a & 4b: Grade distributions for two lecturers

lecturers, but there are significant differences in their frequency depending on the lecturer.

Lecturer 1's graph has a much closer grouping than Lecturer 2's, where the bi-modal pattern is much more evident. While the spread of marks follows different patterns the modal values are relatively similar.

3.3 Postgraduate and undergraduate units compared

To assess the impact that quality of students and maturity, as well as many other less apparent factors may contribute, distributions were extracted for postgraduate and undergraduate units separately (fig 5a and b).

Compared to the undergraduate units, post graduate units display a narrower spread of results and are more likely to be mono-modal. An overall lower failure rate, a higher distinction rate and a higher average is also visible though not tested statistically.

Postgraduate students are likely to be mature age students already in the work force. Additionally, while the impacts of an F grade may under some circumstances be mitigated by the student undertaking a supplementary exam, which if they pass, allows them to pass the unit, this option is not open to postgraduates. The impact of a F is higher for those students.

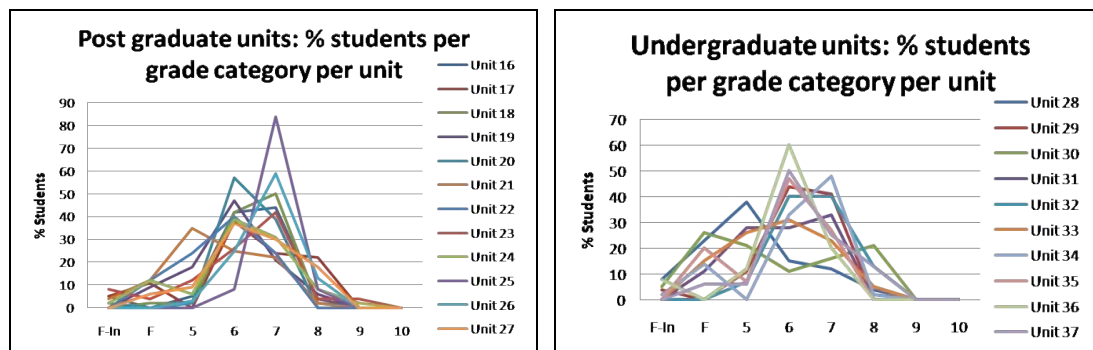


Figure 5a & 5b: Grade distributions, undergraduate & postgraduate

3.4 Factors influencing assessment

A survey of factors with the potential to influence marks distribution was collated. These are shown grouped into categories better representing possible stakeholders' interests, with examples found (Table 2).

Table 2: Factors contributing to marks distribution patterns

	Influencing factor	Source & evidence
1	"Please explain" Flags on particular marks distributions	<ul style="list-style-type: none"> • Directives on exams • Avoidance and querying of extremes • Head of School approvals required for certain distributions
2	"Equity" versus "Filtering" policies	<ul style="list-style-type: none"> • School Policy statements stating "pass based on merit"
3	Distribution of marks compared to other units	<ul style="list-style-type: none"> • Statistics reported to School and at University level • Procedures to identify and justify outlier statistics.
4	School position to generate student quality perceptions	<ul style="list-style-type: none"> • Units "too hard" or "too easy" are examined. Marks distribution is a major contributor to this assessment
5	Student assessment of lecturer	<ul style="list-style-type: none"> • Student assessment system (feeds into staff performance system). Perception that "hard" lecturers may not be scored highly by students.
6	Student appeals	<ul style="list-style-type: none"> • Associated with particular marks
7	Administration related to failures & appeals	<ul style="list-style-type: none"> • Process involving School, administration and data entry staff. Four approvals and assessments typically required.
8	Self perception of lecturer	<ul style="list-style-type: none"> • Discussions with two lecturers in this study
9	School recruitment requirements	<ul style="list-style-type: none"> • Student entry level requirements impact quality of cohort.
10	University positioning	<ul style="list-style-type: none"> • Based on analysis of market position against competitors. Assessed at course level.
11	Student expectations	<ul style="list-style-type: none"> • Based on university wide pass rates and perceptions of standards
12	Student quality	<ul style="list-style-type: none"> • See item 9
13	Impact of assessment on the student	<ul style="list-style-type: none"> • Postgrad and undergrad supplementary exam policies
14	Learning objectives	<ul style="list-style-type: none"> • Open book v's closed book exams • Exam coaching strategies
15	Quantitative vs qualitative unit content	<ul style="list-style-type: none"> • Unit outlines and exam formats

4. DISCUSSION

4.1 Patterns of distribution

The majority (71%) of all units display a mono model pattern with what appear to be normal distribution characteristics. When exam results are reported, mean and standard deviation statistics are calculated and reported along with them. Clearly, such a mono-modal normal distribution is an expectation for the examination. University level reporting also considers these statistics and each School is required to report on variations from university-wide norms.

However, a highly visible feature of assessment patterns in this study is the grouping of units into those displaying a bi-modal pattern and those that are mono-modal. This can be attributed to three main drivers which must be strong enough to counter the scrutiny that they will attract when reported at School and University levels.

Firstly, teachers tend to generate marks away from boundaries. In this instance the boundaries are indicated by the directive defining the acceptable norms for percentage 'failed' students and

percentage 'distinction' students. That is, there are "no go" areas of marks at these boundaries which have been adopted in the marks distribution. A student won't "only just fail" very often. If they are in such a position a review of the exams (sometimes in the marking process itself) will result in movement of the mark away from the boundary. The main boundaries identified in this study are "fails" and to a lesser extent distinctions (7 grade).

The second driver relates to the F-category ("fail"), which is both quantitative (based on assessment results) and qualitative (based on meeting certain pre-requisites) as opposed to grades 5-10 which are just quantitative (based on assessment results). Thus, the F-category lumps results due to different causal effects into one, which creates an anomaly in comparison to the other categories. For example, a student may see an F-IN (Fail – Incomplete) with a mark of 55% as an anomaly, though it might actually mean that an assignment was not completed. F-IN marks generate significant administrative overhead and also much stress for students who seek to reverse this.

Thirdly, some subjects are more quantitative in nature and a "tough" unit may generate more fails as there is less room for interpretation of results. In such units a bi-modal pattern could also form where some of the students in the unit understand the concepts well while others cannot.

4.2 Influence of the lecturer

Differences in both spread and degree of bi-modal patterning is evident when looking at units for individual lecturers. These differences can logically be attributed to several internal and external factors. Lecturers usually teach a portfolio of 5 or 6 units and they may simply be biased towards units that cause these patterns. As well, this bias may extend to the balance of postgraduate and undergraduate subjects with the differences in pattern that have already been observed.

Having said this, lecturers also appear to have a distinctive individual marking style, a marks distribution "fingerprint" based on clustering of marks which can be seen to some degree in fig 5a and 5b. This may be a combination of all sorts of personal decisions about unit assessment. For example, supporting the assertion of the previous section Lecturer 1 taught more qualitative units where interpretation of the results leads to a lower fail rate. Lecturer 2's units are more often quantitative and can't so readily be interpreted. Not so obviously based on a driver for bi-modality Lecturer 1 was less inclined to mark above an 8. This lecturer reserved such marks for "exceptional students delivering more than expected in that unit", a personal definition not formally defined at any organisational level.

Other individual factors may influence student distributions. This includes forms of coaching (lists of possible questions, chapters to concentrate on, and other such approaches) that can lead to higher (or lower) pass rates.

The nature and role of the exam in the unit also varies by lecturer. For instance Lecturer 2 saw the exam primarily as a test of students' accumulated knowledge and understanding. Lecturer 1 saw studying for the exam as an important learning tool as well.

The type of examination selected by the lecturer may range from open book exams, focusing on analysis and interpretation to closed book exams focusing more on memory. This is seen to impact the results, improving pass marks and narrowing the overall spread of marks.

4.3 Post graduate compared to undergraduate units

The differences between the patterns for post graduate units and those for undergraduate units could be attributable to two main factors:

Firstly, student quality and student maturity are likely to influence distribution. Post graduate students are typically mature-age students and full-time workers, displaying a higher level of commitment and organising capability. They may also have quite specific goals which they aim for in a more concerted manner.

Secondly, school policy allows supplementary exams for undergraduate units, but not for post graduate units, which may influence students' approach to studying, and also teachers; approaches to marking.

4.4 Stakeholder influences

The influencing factors identified in this study can be mapped onto the marks distribution (table 3). Each plays a role in shaping grade distribution patterns, ranging from formal policies/directives through informal marking guidelines to individual approaches and external factors.

Table 3: Stakeholder influences and factors

	<i>Influencing factor</i>	<i>Impact on marks distribution</i>
1	“Please explain” Flags on particular marks distributions	Bi-modal distributions
2	“Equity” versus “Filtering” policies	No direct evidence yet
3	Distribution of marks compared to other units	Normally distributed marks with constrained mean and standard deviation
4	School position to generate student quality perceptions	No direct evidence
5	Student assessment of lecturer	No direct evidence
6	Student appeals	Bi-modal distributions Higher pass rates in postgraduate
7	Administration related to failures & appeals	Bi-modal distributions Higher pass rates in postgraduate
8	Self perception of lecturer	Changes in pass and distinction rates
9	School recruitment requirements	No direct evidence
10	University positioning	No direct evidence
11	Student expectations	No direct evidence
12	Student quality	Higher pass and distinction rates
13	Impact of assessment on the student	Lower fail rates
14	Learning objectives	Changes in pass and distinction rates
15	Quantitative vs qualitative unit content	Bi-modal distributions

However, even though there are many influencing factors impacting the marks distribution of individual units, over time and across multiple units, student results display some strikingly similar characteristics. They all have relatively similar modal grades, and only a small rate of distinctions. However, within that constraint they do exhibit two distinct patterns, mono- and b- modal, and these are associated with different fail rates.

4.5 The Biggs model extended

This study demonstrates the wide range of factors, associated with different stakeholder groups, that impact the distribution of marks. In that, it is in line with the Biggs model presented earlier.

Fig. 3 shows an extended Biggs model. In this model student results are primarily driven by assessment choices from teachers and learning prioritisation from students. Other factors influencing grade results also play a significant role. The institutions’ policies, culture and market positioning impacts assessment results directly by setting policies, guidelines. Differences in the perception of competency levels also influence student results. These are then tested against standards and norms and this is used as a feedback mechanism to influence future assessment patterns.

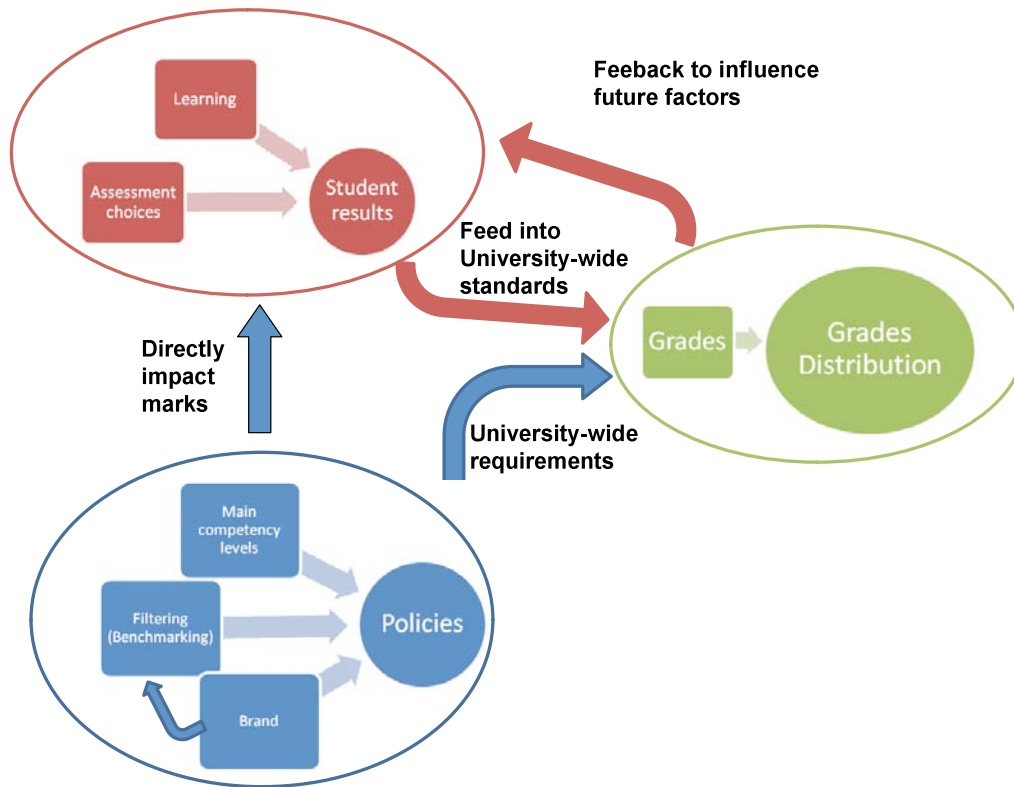


Figure 6: Influences and feedback in grading: the Biggs model extended

5. CONCLUSIONS

The role of statistical characteristics of assessments for groups of students is fraught but still valid. It is apparent that students' assessment results are influenced by a lot more than just the demonstration of the individual student's learning. Student assessment is a complicated area, in which institutional aims are reflected as branding, standards, recruitment practices and policies, all of which may directly or indirectly impact a teachers' approach to assessment. There are also practical considerations such as administration overheads aligned with particular assessment distributions.

In addition, teachers themselves have styles that alter marks distributions. Further, the stakeholders in the whole assessment process interact and directly or indirectly influence the assessments. Students for instance choose the university on brand, and brand drives "filtering", "normalising" and other assessment outcomes, which influences policy, which in turn impacts teachers' marking behaviour and consequently marks distributions.

So, is assessment a massively subjective area? The data may show differences in distribution, but it also shows considerable consistency. That this should be able to survive the many forces influencing it does help to demonstrate its validity, within of course, the context of different interpretations by the different stakeholders.

The study does however indicate some opportunities for Universities. Rather than imply assessment by means of these possible outcomes, maybe assessment needs to be an explicit part of university branding and be considered in any comparative ranking that institutions participate in.

There is also a need for institutions to clearly define the factors influencing assessment results and to ensure a balance between student-controlled influencers, teacher controlled influencers, and those outside of these areas of control.

To avoid indirect negative impact on student results, assessments should be clearly designed to meet their objectives, taking into account that these objectives are a product of policy, brand and culture.

The “dimensions” of learning are only partly aligned with assessments, but are directed by other objectives too.

A further requirement is that institutions should seek feedback from the outcome of these impacts on grade distribution and how this in turn influences policies and guidelines.

6. FUTURE RESEARCH

Should results be scaled? What is the importance of normal and other distributions? To what extent should student grades be influenced by the institution’s culture and policies, which in turn are related to its market position and stated goals? Are assessment objectives realised through the policies? Are grade distributions actually aligned with objectives of the policies?

In our current tertiary environment, with its assessment obsessive culture, where a percentage point can represent the difference between success and failure to an individual, these are avenues that are valuable to explore. This paper has indicated several lines of continuing research.

References

1. Dochy, F. and L. McDowell, *Assessment as a tool for learning*. Studies in Educational Evaluation, 1997. **23**(4): p. 279-298.
2. Pintrich, P.R. and E. V. De Groot, *Motivational and self-regulated learning components of classroom academic performance*. Journal of Educational Psychology, 1990, . **82**(1): p. 33-40.
3. Australian Broadcasting Commission. *Teachers to vote on League Table boycott*. 2010 [cited 2010 Jan 20, 2010]; Available from: <http://www.abc.net.au/news/stories/2010/01/19/2795549.htm>.
4. Stiggins, R., *Student-centred classroom assessment*. 1997, Old Tappan, NJ: Prentice Hall, Inc. 521.
5. Stiggins, R., *Assessment Crisis: The Absence of Assessment for Learning*. 2002. **83**(10): p. 758-765.
6. Institute, F.E.M. *Definitions*. [cited 2010 Jan 20, 2010]; Available from: www.training.fema.gov/emiweb/edu/docs/busind/Session%2020-%20Handout-I-Definitions.doc.
7. Biggs, J., *Assessing Learning Quality: reconciling institutional, staff and educational demands*. Assessment & Evaluation in Higher Education, 1996. **21**(1): p. 5-16.
8. Miller, C.M.L. and M. Palett, *Up to the Mark: A Study of the Examination Game*. 1974, London, England Society for Research into Higher Education. 131.
9. Crooks, T.J., *The Impact of Classroom Evaluation Practices on Students*. Review of Educational Research, 1988. **58**(4): p. 438-481.
10. Biggs, J., *Teaching for quality learning at University*, in *Teaching*. 1999, SRHE and Open University Press. p. 165-203.
11. Boud, D. and N. Falchikov, *Quantitative studies of student self-assessment in higher education: a critical analysis of findings*. Higher Education, 1989. **18**: p.: 529-549.
12. Calvert, B., *Demystifying Marking: reflections on developing and using grade descriptors*. Learning and Teaching in Higher Education, 2004(1): p. 93-97.
13. James, D. and S. Fleming, *Agreement in Student Performance in Assessment*. Learning and Teaching in Higher Education, 2004(1): p. 36-50.

14. Rowe, K.J. and P.W. Hill, *Assessing, recording and reporting students' educational progress: the case for 'subject profiles'*. . Assessment in Education, 1996. **3**(3): p. 309-352.