Talking about the elephant in the room: Improving fundamental assessment practices

Betty Gill
University of Western Sydney, Australia

Abstract*

This paper reports on an institution-wide strategy to improve first year assessment practices. Assessment is central to the student experience and to informing their developing conceptions of themselves as students. Despite this central importance, much national and international literature raises questions about the fitness-for-purpose of assessment practices in higher education. The reported strategy was developed in response to analysis of student feedback, which suggested, like the literature, substantial opportunity for improvement. Student feedback on their assessment experience was validated by an audit of first session assessment and used to inform the strategy. A significant improvement in quantitative and qualitative measures of student satisfaction across routine data sources is provided to demonstrate impact. This supports a conclusion that the first year student experience can be impacted by the systemic application of a small number of fundamental good practice assessment strategies which are outlined.

*This article was first presented at the 2015 STARS Conference in Melbourne, Australia in July 2015 and was selected via the peer review process as one of the top-rated papers. The author has kindly given her permission to have this article published in the conference issue of the Journal and it has undergone a further review by the editors to confirm it aligns with the Journal format.

Please cite this article as:


This article has been peer reviewed and accepted for publication in Student Success. Please see the Editorial Policies under the 'About' section of the Journal website for further information.

Student Success: A journal exploring the experiences of students in tertiary education

This work is licensed under a Creative Commons Attribution 4.0 Licence. As an open access journal, articles are free to use, with proper attribution, in educational and other non-commercial settings. ISSN: 2205-0795
Introduction

Commencing students, transitioning to university studies have particular learning needs, not the least of which is to be inducted into an academic culture, of which assessment practices and expectations are a major component. First year assessment practices are an important dimension of the first year student experience, essential for the foundational development of the necessary skills to be successful students. Assessment is central to the student experience and to informing their developing conceptions of themselves as students and as potential future graduates. Those experiences have enormous potential to impact student confidence; to elicit questions about whether university is really for them; whether they can succeed at university? Tinto (1993) argues a student’s initial experience at university, particularly in relation to assessment, can impact significantly on their long term engagement and retention. Within a widening participation agenda, such issues take on increased significance and yet, despite pockets of excellence and innovation, there is little evidence of systemic improvement in assessment practices in the higher education sector if student satisfaction is the measure.

Despite the central importance of assessment to the student experience, much research literature raises questions about the fitness-for-purpose of assessment practices. Following analysis of comments made by 94,835 graduates from 14 Australian universities, Scott (2005) reported assessment (standards, marking, expectations, management and feedback) as a recurrent “hot spot” (p. viii) consistently receiving a higher ratio of “needs improvement” (NI) to “best aspect” (BA) comments. Clear assessment requirements and an understanding of what is to be produced are key (Scott, 2008) and not unreasonable expectation of students. Similarly, Price, Carroll, O’Donovan and Rust (2011) report that the UK National Student Survey consistently shows assessment processes as the aspect of their study students report least satisfaction with. They conclude that this could not come as a surprise to anyone in higher education, since for many years’ researchers in the field have been telling the “same story” (p. 479).

The FYE literature consistently emphasises the importance of quality assessment to assist student transition and the development of the necessary skills to be successful students. The dominant message is the need for a fundamental reconceptualization of the purpose of assessment. As Krause (2006) argues, we need “a shared understanding of how to integrate assessment and feedback structures into [and across] the curriculum so as to enhance student learning” (p. 6). In other words, shifting focus from “assessment of learning to assessment for learning, particularly in the first year (p. 6). Further articulation of key aspects of assessment practice addressed in the project and literature supporting the measures are provided later in the paper as the project is outlined.

Evidence supporting the need for improvement

Routine student evaluative feedback - assessment related

Student feedback across a range of routine institutional data sources was reviewed in 2012. Over two consecutive years—2011 and 2012—respondents to the University of Western Sydney (UWS) Commencing Student Survey (CSS) rated “clear assessment requirements” as being of highest importance in supporting their learning. In both years this item showed the largest gap between student \((n=2,371)\) importance rating (4.632 on a 1-5 scale) and their performance rating (3.941) (gap of 0.695 in 2012).

A review of the 2012 CSS qualitative comments reinforced this primacy. The ratio of NI to BA comments for the category “Assessment Task” at 3:1 was the highest of all categories. It also
attracted the largest number of comments (921) in terms of what students nominated would be of most help to them (“Most Help Now” [MHN]). When categorised into subcategories, the highest proportion of these comments related to “assessment expectations” (NI=272 and MHN=302), with “timing” (NI=112 and MHN=302) being the 2nd highest.

The assessment-related items (3 of 13) on the Student Feedback on Units (SFU) evaluations had over many years consistently received lower satisfaction ratings than other items. Review of qualitative comments for 1st Session, 1st year units relating to “Assessment Tasks” showed a NI:BA ratio of 2:1. This contrasted to all other categories where BA comments exceeded NI comments. By way of an example of this contrast, student comments concerning “Staff” (quality and teaching skills, etc.) indicated a reverse ratio of just one NI to every two BA comments (1:2). The highest proportion of NI comments related to “expectations” (n=3,123), followed by “timing” (n=1,414), “feedback” (n=1,862), “assistance” (n=1,001) and “load” (n=1,089).

Other corroborating evidence:

Further validation of the student experience was sought through verbal feedback from staff working in support programs across UWS. Anecdotal advice from staff working in the Library Roving program, where students can, at their initiation, receive immediate advice and assistance from academic literacy staff, confirmed that many students find it challenging to understand assessment instructions - what is being asked of them, what is expected, and how to go about approaching assessment tasks? The support staff, all with higher degrees, reported difficulties themselves at times in clearly and unambiguously comprehending some assessment tasks, thus increasing the challenge of providing clear explanations to direct and guide student understanding and approach to assessment.

Validating student feedback - the student voice

In light of this evidence, along with literature suggesting improvement of assessment as a priority for higher education, the need to examine assessment practices to understand and validate the student perspective, was a first step to identify improvement opportunities.

Institutional audit of 1st session assessment practices

In 2013, a review of 1st year 1st session assessment was undertaken. Following endorsement by the UWS Senate Education Committee\(^1\) (SEC) course teaching teams were requested to collaboratively review unit assessment plans and provide a course level report documenting:

1. A whole-of-course assessment schedule showing assessment tasks, size and timing across all core units, as well as an estimate of the time an average student would need to complete each assessment task. Course teams were asked to reflect on and confirm that the amount and spread of workload throughout the semester was appropriate for a full time 1st year transitioning student (Higher Education Academy, 2012, Kift, 2009, Krause, 2006);

2. The identification within the schedule of an appropriate early, low risk assessment as advocated by Kift (2009); and

\(^1\) The Senate Education Committee has responsibility for promoting quality in Learning and Teaching.
3. Confirmation of consistency in descriptors used for each type (mode) of assessment common across units, and the criteria and standards applied (Kift, 2009).

Reports were endorsed through each School Academic Committee (SAC), the peak school academic quality body, prior to submission to the SEC, encouraging further peer review and discussion to ensure the broadest awareness and engagement with these quality issues.

An additional component of the project involved professional staff (Learning Advisors - Academic Literacy) collaborating with teaching teams to review unit assessment information across a sub-sample of large courses to more comprehensively assess the extent to which:

1. Assessment information was clear and provided in a language accessible to and appropriate for students transitioning to university and a new academic discipline;

2. Assessment types common across units used consistent terminology; criteria and standards reflected common understanding of requirements and expectations, with common (consistent) marking rubrics used for the same assessment type; and

3. Exemplars were provided to demonstrate standards and expectations to students.

As well as reviewing materials, where opportunities were identified, they worked with academic staff to revise those materials and enhance the clarity of assessment information and to develop annotated exemplars. Along with the peer review processes embedded in the reporting aspect of the project, this was designed to facilitate reflection and discussion of existing assessment practices and encourage quality improvement. Encouraging staff to critically reflect on student workload at a whole-of-course level was also an integral part of the project.

**Assessment improvement foci**

**The importance of making student information clear**

Devlin, Kift, Nelson, Smith and McKay (2012) emphasise “the importance of making expectations clear for low SES students in language they understand” (p. 26). They report both staff and students citing the “benefits of thorough explanations of assessment requirements and criteria, and the use of accessible language and examples to ensure student understanding” (p. 26) [emphasis added]. This need, whilst articulated specifically for low SES students, is also particularly relevant for first in family students and unquestionably represents fundamental good practice, of benefit for all students transitioning to university study, and beyond. As Meyers and Nulty (2009) remind us, adequacy of assessment information is necessary to establish a sense of purpose and direction for students, and thus to engage them in assessment processes and learning from assessment.

**The importance of considering total student workload**

Student workload is important to understanding the student experience with assessment load and timing being prominent in NI comments. “No surprise” many academics would respond, “they would say that!” But how do we know if a student’s total workload is appropriate and manageable? The modularisation of courses, where units are often planned and taught in relative isolation (Higher Education Academy, 2012) resulting in secret unit business, militates against such
knowing and requires deliberate strategic action to overcome. How commonly this occurred, was something we sought to understand.

It is generally expected that a UWS student spend 10 hours per week on their studies for each 10 credit point unit, including all learning and teaching activities such as class attendance, prescribed readings, preparation for tutorials, research and completion of assessment, examination preparation and sitting time. It also should include independent learning, of which Karjalainen, Alha and Jutila (2006) stress “thinking time” as a component, and a basic precondition for learning. For a fulltime student, this equates to 40 hours per week, but will peak at different points across the semester when assessment items are due. If however, a number of substantive assessment items are due in the same week, or over consecutive weeks, then significant pressure will be placed on students, with resultant anxiety. This is particularly true for transitioning students who need to develop appropriate time management skills, at the same time as coming to terms with university norms and expectations.

Study time expectations have changed little over recent decades despite a significant change in the student body with many more working high average hours (Universities Australia, 2013) or with caring responsibilities. Acknowledgement and consideration of these aspects in planning curricula and workload is important. This is not to argue for reduced standards, but for intentional and coordinated course level planning of curricula, including assessment, so as to ensure reasonable and manageable workload, to maximise learning impact and not overwhelm students. Interestingly, Price et al. (2010) noted a UK Higher Education Policy Institute study “estimated students needed around 30 hours study per week to achieve the learning outcomes set for full-time study” (p. 486) - a 25% decrease on the UWS norm.

To what extent we plan unit learning and teaching activities and specifically assessment activities with a clear understanding of the amount of time an average student would need to devote in order to successfully complete those activities, particularly beginning students, is unknown. What should a total student workload look like? Actual workload is however only one component which can lead to student feelings of overload. How that workload is scheduled, the clarity and consistency of communication, and unfamiliarity with expectations will also impact, particularly on transitioning students. Conflicting and confusing instructions and advice; feelings of inadequacy and of not belonging will also compound feelings of overload and student anxiety. A student experiencing overload and anxiety will not be capable of efficient learning, nor the development and demonstration of academic skills needed for success. Indeed they may not be capable of any action. Feelings of paralysis—not knowing where to start, or what is expected—will inevitably result in inaction manifesting as non-submission, non-attendance and ultimately withdrawal, often informally. Given a significant number of students do not formally withdraw, but rather simply “disappear”, reinforces this scenario.

The project thus sought to gain some insight into the issue of student workload, additional to scheduling and clarity of information, by asking teaching staff to estimate the amount of time it would take a student to complete assessment tasks. Determining workload, or the right amount of time needed to complete a task, including reading, researching and thinking, is an extremely complex task with limited guidance available to assist in making such judgements (Karjalainen et al., 2006) so considerable variation in estimates was expected.

Assessment workload and the quality issues identified above are important not just for the student experience. Academics teaching, often
very large first year units consisting of student new to university from diverse backgrounds, face significant challenges. The comments of students in the CSS and SFU surveys indicate the desire of students for greater clarity in assessment tasks, in particular understanding of expectations. Inevitably, first year teaching teams are the first port of call for student enquiries. Increasing clarity, coherent planning and consistency in communication of assessment requirements, offer significant potential to reduce staff workload. It is clearly important, as Kift (2009) stresses, that assessment workload is manageable for both staff and students. Taking a whole-of-course approach to the planning of assessment, enabling student learning and assessment to be integrated and scaffolded horizontally and vertically, along with clarity of assessment information, can help to achieve this outcome for both staff and students.

What was learnt from the audit?

So what did we learn about 1st session assessment practices at our university? First and foremost, the reports and review of existing practices validated many of the NI comments made by students, reinforcing that improvements in some fundamental aspects of assessment offered significant opportunity to better support student transition, learning and retention. Similarly, they offer improvement for staff workloads, albeit requiring time for consultation, collaboration and integrated development in the planning stage. A summary of some findings and discussion of the significance of each issue follows.

Scheduling of an early low-risk assessment

Few courses reported scheduling of an early low-risk assessment of the type advocated by Kift (2009). In the vast majority of cases where there was an early assessment item(s) in the schedule, it was not an appropriate task to fulfil the goals articulated by Kift (2009) as they were most often a quiz. Only a small number of courses had low-risk written tasks. It appeared from responses that the purpose and function of having an early low-risk assessment task was poorly understood.

Scheduling of assessment tasks

The number of assessment items, their spread and degree of integration across units are important for structuring learning for all students, no more so than for transitioning students. This includes the extent to which task sequencing enables the opportunity for students to build skills, develop their understanding of different types of assessment tasks (genres) and associated expectations, and to learn from feedback. Across UWS, there was large variation in the workload expected of students and the number of individual tasks per unit. The completion of an assessment schedule across all core units within a program emerged as common practice for many, but a first for many other programs.

The requirement to complete an assessment schedule initiated discussions among teaching teams and within SACs, leading to modification of assessment schedules in some courses, and raised awareness of the importance of such planning and scheduling more generally. That being said, there were some practices the audit highlighted that raised concern and offered the opportunity for improvement of both the student and staff experience of assessment. One course assessment schedule demonstrated these concerns. The assessments across four core units included only quizzes or tutorial participation up until week 7 when three substantive (20, 30 and 50% weighting) written items were due across three units. Each of these pieces of written work represented three different genre (essay; annotated bibliography and case study), each a new experience for students, as even university level essays differ significantly in their expectations to those secondary school level students complete.
If we accept that assessment, particularly for commencing students, should be focused for learning – that it should help students to develop and build skills and understanding of academic and disciplinary conventions and expectations, with opportunity to learn from feedback, then the implications of this scheduling are significant, including:

- The lack of any early low risk written task and subsequent feedback prior to three substantive tasks being due in the same week, means that students had no opportunity to evaluate their skills or to be referred to services for assistance prior to having to produce all three significant pieces of work;

- No opportunity for students to learn from and utilise feedback from one written assessment on their writing, expression, referencing, etc., to inform subsequent work;

- Adding to the complexity confronting students, the three different writing genres, required differing structures and other requisites, all at the same time;

- Three groups of academics confronted with the same errors or inadequacies in student writing, expression, referencing, etc., likely to each spent time providing feedback on these same issues.

Whilst this example is at the extreme in terms of number and types of substantive assessments due in the one week, it does not represent an aberrant picture across courses. Other course schedules also included a significant number of different genres, with at least one requiring eight different assessment types. This places a large demand on students studying in their first session who are required to understand and adapt to the requirements of such a broad range of genres in limited time. This is particularly concerning, as academics generally appear to underestimate the complexity of such tasks for commencing students. For example, Wilson (2011) asserts that essays in the first year are a source of great angst for both students and staff, and may be more complex than we (staff) think.

Many units tended to have smaller, more frequent items, with lower percentages such as online quizzes. This strategy is apparently adopted to engage students with the learning materials – to “keep them actively engaged”. However, as Boud et al. (2010) note:

> Many separate low-value pieces of assessment can fragment learning without providing evidence [to either students or staff] of how students’ knowledge and skills from a unit of study are interrelated. This is often compounded across subjects, leading students to experience knowledge [and skill development] as disconnected elements.

(p. 3)

**Student workload and estimates of student time required to complete assessment tasks**

Significant variability was evident across staff estimates provided for similar tasks (i.e. same task type and word limit). There was large variation in average weekly workload between units in the same and different courses, with ranges of 20 hours in one unit to 12 hours in another unit within the one course. Though recognising the lack of reliability estimates provide, such differences suggest little coherence and coordination in planning and raise questions about what relative importance or relevance students may perceive of the differing units. The importance of total workload in understanding the student experience and of overall workload expectations being reasonable and manageable for students should not be underestimated. Once workload expectations are determined and assessed as reasonable and manageable
for students, it is desirable that students explicitly understand these expectations.

**Assessment descriptions and marking criteria consistency and provision of exemplars**

In her *First Year Curriculum Principles*, Kift (2009) asserts the importance of consistency in communication of assessment expectations and advocates the need for consistent criteria and standards, naming of assessment tasks and use of assessment verbs. Whilst course self-reports generally indicated that different assessment types used consistent terminology, this was contradicted through the review of assessment information within the sub-sample. Considerable variation was noted in use of assessment terminology in these courses. The university has sought to address this issue by developing a list of assessment modes, with clear definitions. Across the university, the relationship between marking criteria and assessment descriptions was at times inconsistent. The lack of clarity was compounded by the absence of exemplars demonstrating standards and expectations to students. Exemplars were uncommon across the courses reviewed despite student feedback consistently communicating a desire for them and the UWS Assessment Policy strongly advocating their use. This is of concern given that, as the Higher Education Academy (2012) asserts, the best way to demonstrate standards and expectations is through provision of, and discussion around, anonymous exemplars of different responses to assessed work.

**Clarity of purpose and appropriateness of language for transitioning students**

Practices in outlining and explaining assessment tasks varied greatly, even among units in the one course. Marking rubrics were not always included in unit documentation and when they were, marking criteria on occasions did not accurately reflect the stated assessment aims. Occasionally, different marking rubrics were used in documentation than were provided on the unit’s LMS site. Other problematic areas include the quality of rubrics in terms of their ability to enable students to improve performance and their facility for objective and reliable marking of large cohorts. Language used was generally judged to be appropriate. However, at times, confusing and sometimes contradictory information, along with inconsistent use of terms across units, combined to limit clarity and understanding.

**The institutional improvement strategy**

Following the audit report, SEC endorsed four recommendations for implementation in 2013:

1. Collaborative planning and assessment mapping for all courses, with mapping endorsed by SACs to ensure peer review as to the appropriateness of the assessment schedule (including range of types/genres, number and spread of items) for the students’ stage of study;

2. Development from this mapping of a course level student assessment schedule to be made available to all commencing students in each course;

3. Implementation of an appropriate *early low-risk assessment* item within a core 1st year unit to identify students who are not engaged or who need additional support with core academic skills, and/or academic advising; and

4. Implementation of peer quality review processes within all courses to ensure the quality review of student learning guides, including assessment information.
Evidence of positive impact of Assessment Improvement Strategies:

Student feedback on a range of routine institutional surveys undertaken in 2013 and 2014 were compared. These years were chosen because in 2013, awareness raising and improvements were instigated as a result of the audit process activities, and in 2014, the above recommendations were implemented institutionally.

<table>
<thead>
<tr>
<th>Table 1: CSS reported importance and performance gap (2013 and 2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SFU Item</strong></td>
</tr>
<tr>
<td>Clear Assessment Requirements</td>
</tr>
<tr>
<td>Clear Learning Guides</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Mean satisfaction ratings on SFU items 2013 – 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SFU Item</strong></td>
</tr>
<tr>
<td>Assessment Activities</td>
</tr>
<tr>
<td>Assessment Guidelines</td>
</tr>
<tr>
<td>Workload</td>
</tr>
</tbody>
</table>

Analysis

Mean ratings for quantitative items in 1st year 1st session SFU surveys were compared and a paired t-test applied to assess statistical significance, as well as the more conservative Cohen’s d test. Significance levels were set at .05 for the t-test (i.e. p<.05). For the Cohen’s d differences are said to be “meaningful” when d=/>3. Comparative ratios of BA and NI comments provided by students in SFUs are provided – a function of dividing BA by NI comments. A ratio of 1 indicates the same number of BA to NI comments; <1 indicates more NI relative to BA comments and >1 more BA than NI comments. Qualitative comments (BA and NI) were treated as categorical data enabling a Chi2 test for statistical significance.

Table 2 shows a positive improvement in mean student ratings for relevant SFU items. Each are statistically significant (p < .05), however using the more conservative Cohen’s d test do not meet the =/>3 threshold.

Table 3 shows the change in BA:NI ratios between 2013 and 2014 for comments categorised as relating to assessment tasks, as well as for the major and relevant sub-categories reported, each of which shows a positive improvement. The change in ratio for the assessment task category is significant (p=.011), as are two of the three sub-categories reported, with the biggest most significant change occurring in the sub-category relating to assessment expectations (p=.000). The

Changes in the gap between student importance and performance ratings on key items in the CSS are provided, with no appropriate significance test available.

Results

Table 1 demonstrates a declining gap between student mean importance and performance ratings on two relevant items between 2013 and 2014.
number of comments in the feedback sub-category—not an explicit focus of the strategy—though improved, were not significantly different ($p=.178$).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment Task</td>
<td>5,510</td>
<td>10,799</td>
<td>0.51</td>
<td>2,524</td>
<td>4,588</td>
<td>0.55</td>
<td>.011</td>
</tr>
<tr>
<td>Sub-categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expectations</td>
<td>1,429</td>
<td>3,123</td>
<td>0.45</td>
<td>780</td>
<td>1,419</td>
<td>0.54</td>
<td>.000</td>
</tr>
<tr>
<td>Feedback</td>
<td>622</td>
<td>1,862</td>
<td>0.33</td>
<td>288</td>
<td>771</td>
<td>0.37</td>
<td>.178</td>
</tr>
<tr>
<td>Timing</td>
<td>435</td>
<td>1,414</td>
<td>0.31</td>
<td>304</td>
<td>764</td>
<td>0.40</td>
<td>.003</td>
</tr>
</tbody>
</table>

**Table 3: Ratio of BA to NI comments for SFU assessment categories**

Discussion and conclusion

The consistency in quantitative and qualitative improvement in student feedback across relevant items of the CSS and SFU support the conclusion that systemic implementation of these fundamental good practice principles institutionally has had a significant impact on student satisfaction with their assessment experiences.

The audit project undertaken in 2013 was important in raising awareness of fundamental aspects of assessment impacting the student experience and for validating the student voice. It has resulted in positive discussions within course teaching teams and schools as they engaged with the process, and has been the impetus for ongoing change initiatives within schools. The implementation of key recommendations arising from the initial project seeks to consolidate an institutional cultural change process, designed to ensure a holistic, integrated and developmental approach to assessment across units becomes standard practice at UWS.

Effective collaboration among teaching teams, oriented towards the constructive alignment and assessment of course learning outcomes, is essential to achieve quality outcomes (Higher Education Academy, 2012). The fundamental assessment practices addressed in this project are a core and necessary step in this process, essential to supporting student transition, skill development and the circumstances for optimising student success. If the elephants in the room, viz.

- assessments reflecting a lack of clarity and inconsistent application of standards and expectations;
- secret unit business resulting in uncoordinated assessment planning; and
- potentially unmanageable, indeed unknown, student workload;

are not addressed, then student learning and the student experience will be negatively impacted. Additionally, the efficacy of co-curricular support programs, which are designed to support student learning and the development of academic skills and on which universities expend considerable resources, will be limited by these fundamental inadequacies. The strategies implemented in this project, with the exception of conceptualising and quantifying student workload, are a reasonably “light touch” and thus achievable without additional investment.

We are in the early stages of exploring the complex task of validly quantifying the time needed for students to complete assessment
tasks and other learning activities in order to gauge overall workload expectations. It is clear however that to support student transition the first year curriculum, including assessment, needs to be intentionally and coherently designed, planned and scheduled. This however requires significant cultural change to a whole-of-course approach, making cross unit integration and collaborative planning normal business.

References


