A NOVEL APPROACH TO GRADING AND GENERATING FEEDBACK FOR PRACTICAL ASSESSMENTS

Paul Bourne and Sean Lancastle
Southampton Solent University
United Kingdom

Abstract

This paper discusses the design and implementation of a practical assessment using Microsoft Excel to automatically grade and produce feedback based upon heavily quantised scores. The aims are to design an assessment that encourages experiential learning, efficiently and consistently produces effective feedback, and produces an appropriate distribution of grades that link to the intended learning outcomes. The pedagogic merits are discussed and some operational considerations. A pilot project is evaluated based on the tutor perspective with the intention to discuss the student perspective in a future paper. Findings from the pilot suggest that the project was largely successful, with the core aims met.

Background

When developing assessment practices, it is important to consider the context and profile of the students involved. Not all practices are likely to work in every scenario.

About the Institution

Southampton Solent University is a post-1992 UK institution with a remit for widening participation; Solent's strategy (Southampton Solent University, 2015) includes aims to recruit undergraduate students from non-traditional educational backgrounds and socio-economic groups. These are generally first generation applicants with vocational qualifications whom may have limited experience of the teaching and assessment strategies typically employed within Higher Education (HE). This poses challenges when developing inclusive assessments that do not rely upon prior experience or understanding of HE (Duke, 2015). Solent validates its units, levels and courses against learning outcomes and assesses students using grade marks that use criteria linked to these outcomes.

The Media Technology Programme comprises a range of Bachelor of Science degree titles in which the students learn about the development, systems integration and operation of the technical equipment used within the broadcast and audio engineering fields. They are essentially applied electronics degrees with a focus on developing the contemporary technical and personal skills required by related industries. The broadcast industry is suffering from an aging demographic and skills gap (Poray, 2012), and the course team is working closely with a variety of professional partners to help address these issues. The intention is to use this as an impetus to develop new work-based leaning models as discussed by Marshall (2016, p.153).

The course team has run Employability Self Evaluation (ESE) surveys with student cohorts to better understand their confidence in a variety of areas of employability. This aligns with the work of Jones and Sant (2013) and their capital compass model. The

results have shown that Media Technology students generally lack confidence in personal and professional networks and self-efficacy.

Review of Typical Assessment Strategies

Several texts discuss the merits of common assessment strategies in detail such as Race (2015) and Harris and Bell (1996). The discussion below is not intended to provide a detailed literature review but briefly summarises the elements that a new approach should aim to preserve and avoid. Baartman, Bastiaens, Kirschner, & Vleuten (2007) also provides a useful list of quality criteria to consider for assessments.

Presentations and viva voces. It has been said that the best way to test your understanding of a topic is to try to explain it to someone else (Rusczyk, 2016; Paul, 2011). Presentations require the students to understand a topic and impart some knowledge onto an audience. Therefore, students hone communication skills that are vital to their employability as well as demonstrating knowledge. However, many students dislike speaking in front of groups, and the pressure can mean they don't demonstrate their true academic ability. Viva voces can go some way to alleviating these issues as the interaction provides a mechanism for reassurance from tutors and allows them to use their professional judgement to extract knowledge from the students. Furthermore, the tutor is able to provide some immediate verbal and nonverbal feedback resulting in a more heuristic learning experience that aligns with the ideals of constructivism and "assessment for learning" (Biggs & Tang, 2007, p.21, p.201). The mode of assessment used in this project is a viva voce structured around a portfolio of lab work.

Comparison with other methods. Several other methods of assessment are routinely used within engineering courses including formal exams, in-class tests, time-constrained assignments, reports and portfolios. Formal exams and in-class tests are widely used to assess knowledge and are reasonably quick to prepare and mark. Grades should be consistent with the use of model solutions and *key word* marking, which enables multiple tutors to mark submissions. However examinations are known to cause anxiety and can encourage superficial learning by students focussing on rote retention of facts albeit depending on the students' approach to their learning. (Boud & Feletti, 1997; Biggs & Tang, 2007) Furthermore, students from vocational backgrounds may not be rehearsed in exam strategy; students who have previously been assessed by other means – such as BTEC students who tend to complete portfolios – may score lower than peers from A-levels encompassing exams. The disposable nature of exam questions also necessitates a constant stream of new questions, increasing the risk of errors. In-class tests retain many of the characteristics of exams but aim to reduce anxiety by using a familiar setting. Sequences of shorter tests can be formative and encourage reflection. The Media Technology programme uses in-class tests with a *limited open-book* format in an attempt to balance knowledge with understanding and encourage reflection.

Time constrained assignments (TCAs) include assessed practical activities and aim to establish what students can *do* rather than what they *know*. These encourage students to learn through application and experience, and so the assessment should be representative of what the student can do rather than recall. This requires the students to commit to the activities and undergo a deeper learning experience that features concrete experience and reflection (Kolb, 1984). These have traditionally been used in medical

and engineering disciplines, but they are time-consuming, and it can be difficult to provide feedback beyond how the task should have been completed.

Reports and essays are often based upon a project or specific learning experience. Like TCAs, they aim to assess what the student can do. They encourage the students to engage with the learning experience and to use the resources available to them to achieve the best outcome. However they are very reliant on the students' written communication and their ability to manage a project, often focussing on quite specific aspects of the unit. Tutors may become anxious when grading large numbers of reports as the feedback may not always align with the grades, and it becomes difficult to maintain consistency across the cohort (Biggs, 2003; Merry, Price, Carless & Taras, 2013).

Portfolio assessments require the students to wrap a number of tasks into a single body of work. This provides an opportunity for formative feedback on each of the tasks and allows the assessments to contain a wider range of topics than a single report. This can encourage student engagement throughout the unit and provide an assessment of the students' full range of abilities. Previous studies have suggested that students are motivated to complete work that they perceived to be graded (Gibbs, 1992) so portfolios can be an effective way to encourage completion of formative work and continuous reflection. However, portfolios can be very time consuming to create and mark, and there can be confusion over what students should include and how their work will be assessed (Race, 2015).

Drawbacks of presentations and viva voces. Viva voces are no golden bullet. They may penalise students with poor oral communication, and others may not show their abilities when faced with a person of authority (Race, 2015). Heavily structured interviews can also restrict the latitude for the respondent. Tutors should therefore use prepared questions as a starting point for unstructured discussion (Harris & Bell, 1996).

The Media Technology programme uses a range of assessment techniques but relies largely on reports to assess what the students can do. This reliance risks assessing their written communication more heavily than their knowledge, understanding or practical abilities. This is true even though the tasks themselves are largely practical. It is also time-consuming to read reports and to produce detailed feedback for large cohorts. Responses received by the programme from the *National Student Survey* (NSS) indicate that students expect quick and detailed feedback. In order to address this with financially viable staff-student ratios, a new approach is needed.

A New Approach

Based on the review of typical assessment strategies it was decided that the assessment strategy should:

- Encourage students to engage with learning experiences throughout the year
- Assess what the students can do as well as what they know
- Provide the students with some practical experience
- Provide detailed and prompt feedback
- Be efficient for tutors to mark consistently
- Allow multiple tutors to mark segments of the cohort

The ESE survey suggested Media Technology students struggle with networking and may not have established networks. Therefore, it is important to develop their communication skills. To overcome their low self-efficacy, it was felt this was best achieved using a viva voce as it provides a bidirectional communication channel. The subject of the viva voce is to be a series of laboratory activities that are run throughout the year. The sessions then become formative in nature, and the students are motivated to complete the preparation and reflection. The students are also required to keep a portfolio of their work as a prompt during the viva voce; the assessment serves as motivation to complete this good practice. The tutors will assess the conversation around a number of elements by providing the student with heavily quantised scores; the possible scores are restricted to a value out of around three rather than out of 18 possible grade-marks or even a percentage as is common.

Implementation

The design of a new assessment should start with the learning outcomes for the unit such that a grade reflects the extent to which a student has met these objectives. The process was therefore broken down into a series of stages as depicted by Figure 1.

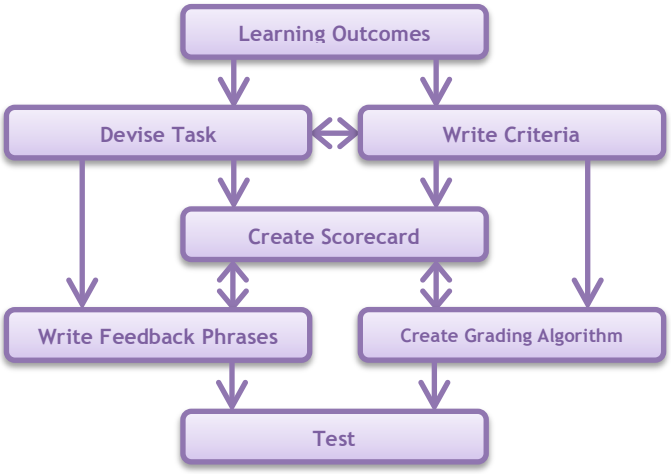

Figure 1. Interdependencies of stages when devising practical assessments.

The Design Process

As indicated by Figure 1, the interdependencies between these stages were not nessarily simple, and the author had to maintain an holistic overview during the process. As with traditional assessments, the process started with devising a task and writing suitable criteria on which to base the assessment. These stages were not distinct as they tended to influence each other; the task chosen affected the areas being assessed, and the desire to assess certain areas inevitably had to inform the task. Once these were created, the author developed a scorecard. This comprised a number of elements for the tutor to score the student against. In order to reduce ambiguity – and thereby increase consistency – the elements are very specific and the scores heavily quantised for example, “Has the student cited external research?” with a binary response, or “Number of labs completed,” with a score of zero to three. The range and depth of the areas on the scorecard needed to be appropriate for the task and criteria and also generate enough data to form a range of sensible feedback and grades. It was important that the

assessment maintained the holistic characteristic of a practical assessment and avoided becoming analytical like an exam (Biggs & Tang, 2007). The feedback phrases needed to be written in such a way that they related to the student's performance and offered constructive feedback. An algorithm was then developed to combine the quantised scores from the scorecard and produce grades for each of the criteria. Tecniques to help with these aspects are discussed below. Deveoping a scorecard that was simple to complete but provided insight was crucial, and this part of the design process was iterative. Once all of these stages were completed, the assessment was tested by trying a variety of input data and reviewing the grades and feedback. Finally a pilot was conducted where the students were directly assessed against the marking criteria as well as using the algorithm.

Generating Feedback

When assessing a student, the tutor notes the scores for each of the elements on the scorecard and enters them into a spreadsheet. Feedback can then be generated using lookups within Excel. A separate phrase was prepared for each possible score such a shown by Figure 2. The phrases were written to be fairly specific and as constructive as possible, which relied on careful deisgn of the scorecard. The cell in the feedback sheet then references the score given and the lookup table. Phrases and words are concatenated to form larger, more natural paragraphs.

	A	B	C	D
1	**Element**	**Score**	**Notes**	**Feedback**
2	Number of Labs Completed	0	None	There is no physical evidence to suggest you have completed the activities.
3		1	Few	It was a shame you had only documented a few of the lab activities. Make sure you attempt all tasks.
4		2	Most	You had documented most of the lab activities; ensure you catch up with all activities in the future.
5		3	All	Well done for including all of the activites.

Figure 2. Example feedback lookup table.

Grading Submissions

Microsoft Excel features a range of formulae that can be helpful in developing the grading algorithm. The weightings for each element were aligned to the assessment criteria with specual attention given to boundary cases and which combinations of scores resulted in a pass.

At its simplest, each element that feeds into a criteria was graded using a sum of VLOOKUP tables. The sum of the highest scores was 100 with appropriate weightings between the elements. A grade could therefore be calculated. For exmaple, students were required to summarise a laboratory activty and draw conclusions from their experience. The tutor then asked a series of easy, medium and hard questions from a prepared bank.

In some instances, grades were also be multiplied by an element within a VLOOKUP table rather than a simple summation. This was useful where an element reflected the quality of several other elements within the criteria such as whether the student's portfolio was presented as a single organised entity.

Conditional IF statements were also useful to cap grades to avoid awarding grades above the appropriate criteria. For example, the marking criteria states that students must correctly format references for a B grade, so a conditional statement ensures that they cannot exceed a C even if they have extensive and professionally written research.

Impact Assessment

The impact of the assessment strategy was assessed in two ways.

Methodology

A sample of 10 viva voces was graded by the same marker using the automated scorecard and using a traditional rubric. The grades and feedback were reviewed by an independent academic to evaluate their merit and whether the students would be likely to generate a similar action plan for self-development based on the two sets of feedback. The distribution of the grades from each of three markers was compared for the population of 20 students per marker. Analysis included standard deviation, mean average, interquartile range; the focus was the consistency between markers and adherence to the university's expectations of a 'normal' unit. See Table 1 and Figure 3.

Results

Table 1

Grade Distributions for the Three Markers

	Average Grade	Standard Deviation
Marker A:	68 %	14.03
Marker B:	60 %	14.04
Marker C:	59 %	15.5

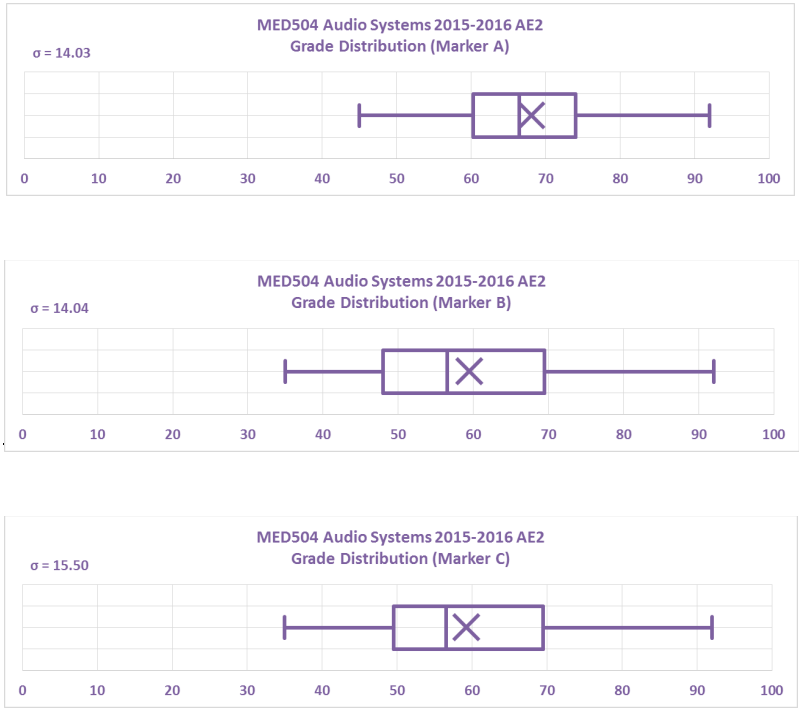

Figure 3. Grade distributions by marker.

Discussion

The feedback and grades were analysed after the entire population had completed the assessment.

Comparison of feedback. There were clear differences between the manual and automatic feedback comments. Firstly, the automatic system was able to provide more feedback than the manual method, simply because of the limited time available to the marker. This is likely to be of significant value to the student, as it is able to provide feedback on a greater variety of elements of the assessment than would otherwise be possible. Secondly, whilst supportive, the manual feedback tended to be more negative, leaning toward justification of the grade awarded with limited suggestions on how it could be improved. The automatic feedback was more balanced, in that it could identify areas that required improvement and suggest strategies to do so. Again, this is partly because of the time limitations of the manual marking process, but also partly because suggestions for improvement were written into each feedback phrase, guaranteeing their presence in the feedback.

There was also evidence of a third, subtler factor. The quality of the automated feedback was influenced by the time spent ensuring that the comments and phrases used were very clear to the reader. Manual, real-time feedback was more likely, in places, to be a little more ambiguous, probably as it was less rehearsed.

However, the manual feedback was, in almost every case, more personal. It tended to identify an aspect of feedback that was unique to the individual student's work that, in some way, separated it from that of the rest of the cohort. It is possible that this unique information is actually of limited pedagogic value to the student – there is not enough evidence of this in this study to be conclusive. However, the emotional reactions of students to obviously individualised feedback may be different to those for feedback which is automatically criterion-generated.

Grade distribution. The grade distributions are best compared using Table 1 and the box plots of Figure 3. As can be seen, the distributions are extremely close. The average grades from the three markers are within 9% with healthy and similar standard deviations. The averages show slight variance, but it is not unusual to find differences up to a grade (~12%) during moderation so the variance falls within what would normally be expected. The standard deviations and interquartile ranges suggest that all three markers awarded a reasonably wide range of grades and with a similar spread. The symmetry of the upper and lower quartiles and close proximity of the mean and median within each plot suggest the grades from each marker are evenly spread around the mean with little skew. It is worth noting that no outliers were identified during this analysis. The university considers a 'normal' assessment at level 5 to have an average mark between 48 and 68 and a standard deviation of above 5. This assessment falls well within those bounds.

Conclusions

Results from the pilot study indicate that the assessment strategy provides consistent grades with extremely similar distributions from the three markers. The feedback is comparable to that manually produced and is generally more thorough and constructive. The automatic feedback would, however, benefit from a couple of personalised

comments. The tutors involved found the assessment easy to administer with anecdotal comments such as "all assessment should be marked like this." One concern is that the format could disadvantage those with poor oral communication skills so a mixed diet of assessment within a programme is still imperative.

At present the tutors have no easy way to return the feedback to the students. Macros can be used within Excel to print the feedback to PDF titled by student ID so it should be possible to automate publishing these to the students via a backend script on the virtual learning environment.

The author intends to investigate the effectiveness of the assessment from the student perspective in due course.

References

Baartman, L., Bastiaens, T., Kirschner, P., & Vleuten, C. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review, 2*(2), 114–129 Retrieved from http://www.sciencedirect.com/science/article/pii/S1747938X0700019X

Biggs, J. (2003). *Aligning teaching and assessment to curriculum objectives.* Retrieved from https://www.heacademy.ac.uk/sites/default/files/biggs-aligning-teaching-and-assessment.pdf

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Berkshire, United Kingdom: Open University Press.

Boud, D., & Feletti, G. (1997). *The challenge of problem-based learning* (2nd ed.). London, United Kingdom: Kogan Page.

Duke, S. (2015). Assessing strategies for success for first generation college students at Merrimack College. *Education Student Publications, 7(1), 1-60*. Retrieved from http://scholarworks.merrimack.edu/cgi/viewcontent.cgi?article=1006&context=soe_studentpub

Gibbs, G. (1992). *Improving the quality of student learning.* Bristol, United Kingdom.: Technical and Educational Services

Harris, D., & Bell, C. (1996). *Evaluating and assessing for learning* (Rev. Ed.). London, United Kingdom: Kogan Page.

Jones, R., & Sant, R. (2013). *The Solent capital compass model*. Retrieved from http://ssudl.solent.ac.uk/2567/1/Compass%20and%20Journey%20with%20references%20and%20Definitions.pdf

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall.

Marshal, S. (2016). *A handbook for leaders in higher education: Transforming teaching and learning*. New York, NY: Routledge.

Merry, S., Price, M., Carless, D., & Taras, M. (2013). *Reconceptualising feedback in higher education: Developing dialogue with students.* Oxon, United Kingdom: Routledge.

Poray, C. (2012). *The shortage of broadcast engineers – is there, or do we just need to know where to look?* [NAB Conference April 2016]. Retrieved from http://www.sbe.org/sections/documents/ShortageOfBroadcastEngineers.pdf

Paul, A. (2011, November 30). The protégé effect: Why teaching someone else is the best way to learn. *Time.* Retrieved from http://ideas.time.com/2011/11/30/the-protege-effect/

Race, P. (2015). *The lecturer's toolkit: A practical guide to assessment, learning and teaching* (4th ed.) New York, NY: Routledge.
Rusczyk, R. (2016). Learning through teaching. *Art of problem solving*. Retrieved from https://www.artofproblemsolving.com/articles/learning-through-teaching
Southampton Solent University (2015). *Building an excellent university*. Retrieved from http://portal.solent.ac.uk/documents/vco/southampton-solent-university-strategy-2015-2020.pdf

Author Details

Paul Bourne
paul.bourne@solent.ac.uk
Sean Lancastle
sean.lancastle@solent.ac.uk