

The use of threshold exams to change students' learning culture and provide assurance of learning

Keith Willey^a and Anne Gardner^b.

Faculty of Engineering & Information Technologies, University of Sydney^a, Faculty of Engineering & IT, University of Technology Sydney^b

Corresponding Author Email: keith.willey@sydney.edu.au

Introduction

Academics often express concerns about the assurance of learning from assessment tasks. In team based projects and individual assignments these concerns often focus on free riding team members and the possibility of assignment submissions being plagiarised and/or outsourced. Despite their shortcomings formal examinations are often favoured as a means of achieving learning assurance as students undertake them on their own in a controlled environment. However, while formal exams may mostly ensure the integrity of a student's submission (students own work) the claim that they provide assurance of learning is less clear. Even if the questions in the exam validly assess the associated subject learning outcomes students are frequently able to pass (achieve 50%) with unsatisfactory, and perhaps even no capacity to demonstrate learning in some subject topic areas, even though they may be regarded as covering requisite material. Furthermore, while undertaking these exams may highlight to students their learning deficiencies, they have only minor impact on their future learning as there is often no opportunity for feed-forward assessments (apply feedback to subsequent learning and assessment) to motivate them to address these identified learning gaps.

This paper reports on the impact of a threshold exam process on both student learning and assurance of that learning.

Background

Assessment is carried out for many reasons including for quality assurance, ie confirming achievement of learning outcomes, for provision of feedback to students, and for longer-term learning (Boud & Falchikov 2007). Much of the literature on assurance of learning focuses on the program, institutional or national level. While recognising the value in conversations at the institutional level about what we mean by learning standards, we argue, along with Sadler (2009, 2010) and Knight (2006) that the reliability of such learning standards depends on the quality of the assessment in individual units or subjects.

Sadler (2009, 2010) discusses the concept of assessment fidelity, defining this as "...the extent to which elements that contribute to a course grade are correctly identified as academic achievement" (p.728) Sadler (2010) discusses 'effort' and 'attendance' as examples of components of a subject grade that do not provide evidence of learning outcome achievement.

Sadler (2010) and Price et al (2011) also challenge us on the practice of progressive accumulation of marks throughout a semester from tasks set at a lower level than the threshold level for the subject (e.g. simple quizzes). There are two aspects to this issue, one is that marks from these early assessment tasks reflect learning at a lower level than is expected at the end of the subject, the other is that the understanding that students have in earlier stages of the semester may be significantly less than their understanding at the end of the semester. In either case this mark accrual process can "...misrepresent the level of achievement reached at the end of the course" (Sadler, 2010, p.735).

Assessment also provokes an emotional response in students. Boud & Falchikov (2007) report that:

“.. the emotional experience of being assessed is complex, and is a function of the relationship between the expectations and dispositions of a learner, relationships between learners and other people, the judgements made about learners and the ways in which judgements are made.” (p. 144 ch 11).

Emotional responses are part of everyone’s experience of being assessed. For some time, it has been claimed that assessment in general, and examinations in particular, can be very stressful. However, Gabriel and Griffiths (2002) and Palethorpe and Wilson (2011) report that while extreme anxiety disables learning, some degree of anxiety stimulates learning.

Willey and Gardner (2012) present a learning framework that suggests that learning is maximised when an assessment activity provides a well-designed learning opportunity and participants (students) approach the activity with a learning focus. Furthermore, for all assessment activities (both summative and formative) we recommend that academics should use scaffolding to reduce undesirable anxiety, promote engagement and ensure students understand and hence can benefit from the learning opportunity provided. We recommend academics use scaffolding to explain to students:

- why they designed the assessment activity the way they did,
- what learning opportunities the activity provides the students,
- how students can evaluate their learning from the activity, and
- how it is going to impact on their reality (enable them to see the world differently).

Considering these issues our aim in designing a summative examination process was to:

- make it learning oriented and include a feedforward component, allowing students an opportunity to respond to feedback and reassess their learning.
- only provide credit for demonstrated achievement against subject learning outcomes,
- increase assurance of learning in that students were able to demonstrate satisfactory learning in all subject topics, and
- move students to approach the exam with a learning focus.

This paper reports on the impact of a threshold exam process on both student learning and assurance of that learning in a first year engineering subject.

Approach

A large (>500 enrolments) multidisciplinary first year engineering subject was chosen for this study. Students were required to undertake a threshold exam in week 10 of a 13 week semester to demonstrate the subject learning outcomes at a satisfactory level. Students were required to achieve a score of 80% to ‘pass’ this exam, the aim being for students to have to demonstrate satisfactory understanding in each topic in order to achieve a pass and hence promote assurance of learning. The reasoning being it would be extremely difficult for a student to achieve 80% if they didn’t have satisfactory knowledge in all topic areas associated with the subject.

The four scaffolding recommendations previously discussed were clearly articulated to students. They were told that the intent of the exam was for them to have to demonstrate satisfactory knowledge in all requisite subject material; that the exam was learning oriented in that it provided immediate feedback allowing them to identify their strengths and weaknesses and recognise both misconceptions and topics that they may need to revise; and that the requirement to achieve 80% to pass this exam increased our assurance of their capacity to satisfactorily demonstrate the subject learning outcomes. In explaining the requirement to achieve 80% and demonstrate competency it was related to the New South Wales driving test. Like the threshold exam it has a high pass mark (90%) and to pass drivers must demonstrate all the requisite competencies to be given a driving license. This

analogy works well as most first year students have received their license within the last 2 years or know peers who have.

Students who failed to achieve 80% in the first exam were given the opportunity to attend a workshop and attempt a supplementary exam in week 14. In the workshops students could ask questions to clarify any misunderstandings and/or address any gaps in their learning. The fact that the exams were marked automatically allowed the instructor to pay particular attention to the most common mistakes in the workshops. These were identified as questions where a significant cohort of students took multiple attempts to achieve the correct answer. The instructor ensured that material addressing these misunderstandings was included in the workshop pre-work. In the week 14 workshop the students work in small groups to compare their pre-work answers before the instructor facilitated discussion on each question, clarifying any misconceptions and subsequently varying the question to check students' understanding.

The time between the first threshold exam and the supplementary exam provided students with the opportunity to revise material addressing gaps in their understanding identified through undertaking the first exam or during the subsequent workshop. This enabled students to feed forward the feedback they received from these activities.

Implementation

The exam was designed to have approximately 3 questions on each of the main subject topic areas. This means with 5 topics there is typically 15 questions. Depending on the number of questions students are typically given between 20 to 25 minutes to complete the exam. The exam is closed book and students are provided with paper on which to write any working out they may need to answer the questions.

The threshold exam consisted of multiple choice multiple attempt (MCMA) questions set at or just above the level of threshold learning outcomes for this first year subject. Students answered this MCMA exam using the Multiple-Choice/Multiple-Attempt mode in SPARK^{PLUS} which allows students to immediately identify if they have answered a question correctly. If they have selected the correct response a tick appears on the computer screen in the relevant square for that option, and the square becomes green. If they selected incorrectly, that response option becomes red and they consider the remaining options, and try again (see Figure 1). Instructors can set up the marking criteria for the exam. In the reported case students were given 100% when answering correctly on the first attempt, 50% when answering correctly on the 2nd attempt, 25% when answering correctly on the 3rd attempt and 0% for answering correctly on the 4th or 5th attempt. This answer and marking summary scheme is shown in Figure 2, and provides feedback for students on their strengths and weaknesses across different topic areas. For the student shown in Figure 2 the strongest areas are Topic 3 and 5 in which the student answered all questions correctly on the first attempt, while their weakest topic area is Topic 1 where they took two attempts to get two questions correct.

This marking regime is fairer than achieved using traditional multiple-choice questions. For example, if a student knows that 3 of the answers are incorrect but is unsure which of the remaining 2 answers is correct then they can achieve 50% of the marks available for the question for demonstrating their partial understanding of the associated topic. Most people can remember leaving an exam and then thinking of the correct answer when it is too late to demonstrate this knowledge nor for it to have an impact on their grade. The MCMA format allows a student who chooses an incorrect answer to be given immediate feedback allowing them to reconsider their thinking and examine the other answers to try and identify where they made a mistake and which answer is correct. Hence, they are learning and getting feedback while undertaking the exam, making MCMA a learning oriented form of assessment that promotes metacognition (students report thinking about their approach to answering questions).

The Multiple-Choice/Multiple-Attempt mode in SPARK^{PLUS} was inspired by the Immediate Feedback Assessment Technique (IF-AT) cards. In controlled trials, the IF-AT method was shown to promote both retention of learned material (Epstein et al 2002, Dihoff et al 2004, Brosvic et al 2005, Brosvic & Epstein 2007) and higher levels of independent learning (Brosvic et al 2005; Persky & Pollack 2008). The combination of immediate feedback and the capacity to think about and rework problems that they got wrong at the first attempt assists students in discovering gaps in their knowledge and areas of misconception. Each of these elements has the potential to increase deep learning (Persky & Pollack 2008). Hence if you do not have access to SPARK^{PLUS} you could replicate the design discussed in this paper using IF-AT cards (which was the way the authors initially facilitated threshold exams in 2011). SPARK^{PLUS} has the advantage of choosing to present the questions to students with the order of both the questions and the answers randomised. This further enhances assessment integrity by reducing the chances of a student being able to gain an advantage by looking at another student's screen during the exam.

View Answer Summary

Prev 1 2 3 4 5 6 7 8 9 10 11 12 13 14 Next

32 mins 34 secs elapsed

Submit

QUESTION 2

Activity	Predecessor	Duration (Days)	Budget cost (\$)	Actual progress (\$)	Actual costs (\$)
1		3	300	100%	250
2	1	5	500	100%	500
3	2	12	1200	50%	700
4	4	4	400	0%	0

In earned value analysis of a project, three dimensions need to be calculated:

1. the Budgeted Cost of the Work Scheduled (BCWS) (or Planned Value)
2. the Budgeted Cost of the Work Performed (BCWP) (or Earned Value)
3. the Actual Cost of the Work Performed (ACWP).

For the project shown in the image, what is the BCWP 20 days after the beginning of the project?

A) \$1450
 B) \$600
 C) \$2000
 D) \$1200
 E) None of the other answers are correct

A B C D E

Record of attempts required to answer each question

Correct on 2nd attempt

Figure 1: SPARK^{PLUS} Multiple-Choice Multiple Attempt (MCMA) assessment screen

While not the focus of this paper SPARK^{PLUS} provides analytics for instructors allowing them to identify which questions students had difficulty with, and the frequency with which incorrect answers were chosen and on what attempt. This allows instructors to identify common misconceptions and misunderstandings, as shown in Figure 3. In this case it is showing the cohort's first attempt responses to a question. As you would expect for a questions set at the threshold or satisfactory level most students selected the correct answer on the first attempt while answer C was the most significant distractor. This information is used to provide feedback to students and clarify misunderstanding in subsequent face-to-face sessions. Students are also provided with additional pre-work for the end of semester workshop within which these misunderstandings were further explored and discussed.

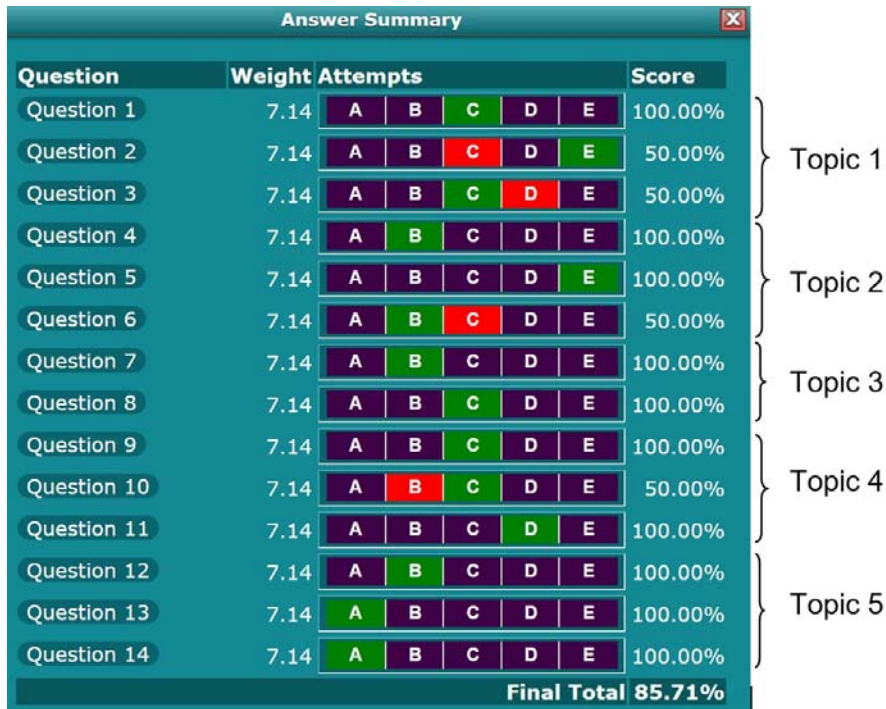


Figure 2: SPARK^{PLUS} Multiple-Choice Multiple Attempt (MCMA) student answer summary screen



Figure 3: SPARK^{PLUS} Multiple-Choice Multiple Attempt (MCMA) instructor analytics screen.

In this study discussion forum posts, a student survey, tutorial discussions and analysis of exam results were used as data to investigate how students' behaviour changed because of the threshold exam assessment, and the impact it had on their learning.

Findings / Discussion

A voluntary questionnaire was distributed to students after the threshold exam. Responses were submitted from 203 students of the total 520 enrolled, which represents a 39% response rate. Students were asked to indicate their response on a Likert scale from strongly disagree to strongly agree. The results for a selection of these questions are shown in Table 1 where Strongly Disagree and Disagree responses are aggregated as are the Strongly Agree and Agree responses. These results need to be interpreted keeping in mind

the relatively high failure rate (52%) of the first threshold exam and how this may have influenced student's responses.

The results indicate that the threshold exam was successful in motivating students (71%) to engage with the assessed learning outcomes with 64% agreeing that it motivated them to learn all the assessed topics. Students commenting that:

Without passing the threshold exam one could not pass the unit and one needed 80 percent to pass it. This motivated everybody to revise very well and work hard but also added a lot of pressure and stress.

The threshold exam made me study hard on all the material we have covered in the semester, due to the high pass-mark required. Therefore, I believe this engaged me more with the work and hence I understand everything to a greater degree.

Threshold exam pushes me to deeply learn every topic.

Failing the whole unit if not passing was the motivation for learning.

Since the exam had a pass mark of 80%, it made me study for it and look back at lecture slides and learn the content. Although I thought 80% was a little too high and got my nerves running quite a bit.

Table 1: Survey results for Threshold Exam

	Strongly Disagree / Disagree (%)	Strongly Agree / Agree (%)
Did the threshold exam motivate you to engage with and learn the material assessed?	29	71
The required 80% pass rate for the threshold exam motivate me to learn all the assessed topics	36	64
The required 80% pass rate caused me to feel stress and anxiety.	15	85
Even though the threshold exam used the multiple-choice multiple attempt format and assessed material at the threshold or satisfactory level required to pass the unit the required 80% pass rate is unfair - it should be less.	23	77

Most students who disagreed that the threshold exam motivated them to learn reported that they are always motivated to learn the material in a subject and hence having a threshold exam had no impact on the way they approached their study or whether they would have studied all of the topic areas:

I will learn the material received from lecturer or tutor whatever threshold exam will be held or not.

I wanted to get more than 80% anyway so I would have set that pass mark for myself.

Nope, 80% or higher would be my goal mark anyway.

The required pass rate didn't really motivate me to learn as my approach would have been the same if it was just a normal exam.

There was much stronger agreement (85%) when students were asked whether the required 80% pass mark caused them to feel stress and anxiety with the majority (77%) agreeing that the pass mark should be less:

Yes, because there was such a high pass mark. This caused my peers and I to PANIC AND STRESS an unwarranted amount about the exam.

Instead of motivating me to learn and thoroughly understand new material, the 80% pass rate placed more pressure and stress on me.

It also caused me to stress out all week and definitely made me feel the worst on the day.

When asked to suggest a fair pass rate remembering that the exam is designed to test learning outcomes at the threshold or satisfactory level the most common responses suggested between 65% and 70%. A small cohort of students expressed a belief that 50% is a pass mark in any assessment and it is never fair to require students to achieve more to pass.

A less common reason provided by students for the threshold exam causing stress, not associated with the 80% pass mark, included receiving immediate feedback when they selected an incorrect answer. Students mostly liked the multiple-choice multiple attempt format as it gave them a chance to recover from a silly mistake and/or demonstrate their partial understanding of the material assessed on the question even if they got wrong at the first attempt. However, students who were getting many questions wrong reported the immediate feedback increased their anxiety:

I mostly felt threshold exam to be stressful when answering the questions and immediately knowing the answer. It really made me nervous during the exam when I saw a lot of mistakes.

Another stress issue raised by a small cohort of students who believe assessment should always be accumulative felt stressed because passing the threshold exam was a requirement of passing the subject. They felt this requirement was unfair. These students felt that if they passed all of the assignments and group work activities then they should pass the course irrespective of how well they scored in the threshold exam. These students commented that:

It seems unfair that the 20% weighted exam can make you fail the entire subject if you fail it.

It isn't a good way to learn at all as telling a first year student that they are a failure isn't a great way to start a 4 year degree.

Several students who failed the initial threshold exam expressed an opinion that the questions were unfair, and it was the unfairness of the questions in the exam rather than their level of knowledge and understanding that prevented them from passing:

Most of the questions were vague and ambiguous, as there was not a clear 'right' or 'wrong' answer. The questions could be interpreted in multiple different ways, which was very frustrating as the 'correct' answer was often not how I would interpret a situation. For example, one of the questions was regarding the HOC, and gave a situation where a guard was put in place, eliminating any risk of injury. In the tutorials this was taught as an elimination, however in the threshold exam this answer was incorrect. Even if students like myself put countless hours into study and preparation for this exam, it is very frustrating to find questions that are not based off the content we studied.

Despite feedback being provided on the first exam and subsequent formative assessment activity it was not uncommon that it was only in the workshop for students who had failed the initial exam that these misunderstanding and knowledge gaps were addressed. The format of pre-work, peer sharing, problem variation and confirmation (Willey and Gardner 2012) was necessary to prevent some students from making the same mistakes again. These problems seem to result when students had a process focused or mechanical approach to learning that favoured learning how to answer a question in a particular context, and then trying to apply the same rules to a different context where the same rules or reasoning could not be applied. This attitude was similar to the attitude reported by Marton, Dall'Alba and Beaty (1993) with

“learning as memorizing and reproducing” (p.286). The workshop activities addressed these issues with many of these students volunteering that they now realised they failed the initial threshold exam because of their misunderstanding and knowledge gaps rather than the questions being unfair.

Conclusions

This study demonstrated the potential of a threshold examination process to improve confidence in assurance of learning. Post-exam workshops provided students with an opportunity to first identify and then address their misconceptions before attempting a threshold exam. Student comments show that many of them who initially failed the threshold exam often focused on mechanical level conceptualisations of learning which challenges us to change learning culture.

References

- Boud D & Falchikov, N (2007) Rethinking assessment in higher education. Routledge.
- Brosvic G, Epstein M., Cook M. & Dihoff R. (2005) Efficacy of error for the correction of initially incorrect assumptions and of feedback for the affirmation of correct responding: learning in the classroom. *The Psychological Record* Vol. 55 pp. 401-418.
- Brosvic G and Epstein M. (2007) Enhancing Learning in the Introductory Course. *The Psychological Record* Vol. 57 pp. 391-408.
- Dihoff R., Brosvic G., Epstein M. & Cook M. (2004) Provision of Feedback during Preparation for academic testing: learning is enhanced by immediate but not delayed feedback. *The Psychological Record* Vol. 54, 2, pp.207-231
- Epstein M., Lazarus A., Calvano T., Matthews K., Hendel R., Epstein B & Brosvic G. (2002) Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate First Responses. *The Psychological Record* Vol. 52 pp.187-201.
- Gabriel, Y and Griffiths, D. S. (2002) Emotion, learning and organizing, *Learning Organization*, 9, 5: 214-221.
- Knight P. (2006): The local practices of assessment, *Assessment & Evaluation in Higher Education*, 31:4, 435-452.
- Marton, F., G. Dall'Alba, and E. Beaty. (1993) Conception of learning. *International Journal of Educational Research* 19: 277–300.
- Palethorpe, Rob; Wilson, John P. (2011) Learning in the panic zone: strategies for managing learner anxiety. *Journal of European Industrial Training*; Bradford 35.5 : 420-438.
- Persky A. & Pollack G. (2008) Using Answer-until-correct examinations to provide immediate feedback to students in a Pharmacokinetics course. *American Journal of Pharmaceutical Education* Vol. 72(4) Article 83.
- Price, M., Carroll, J., O'Donovan, B. & Rust, C. (2011) If I was going there I wouldn't start from here: a critical commentary on current assessment practice, *Assessment & Evaluation in Higher Education*, 36:4, 479-492.
- Sadler D. R., (2009) Grade integrity and the representation of academic achievement, *Studies in Higher Education*, 34:7, 807-826.
- Sadler, D. R., (2010) Fidelity as a precondition for integrity in grading academic achievement, *Assessment & Evaluation in Higher Education*, 35(6), 727-743.
- Wiley, K. and Gardner, A. (2012) Collaborative Learning Frameworks to Promote a Positive Learning Culture. In *Proceedings of the Frontiers in Education Conference* Seattle, Washington, October 3-6, 2012.