

Threshold exams to promote learning and assurance of learning

Keith Willey & Anne Gardner.
Faculty of Engineering & IT, University of Technology, Sydney
Keith.Willey@uts.edu.au

BACKGROUND

Formal examinations are often used in engineering classes as the tool to evaluate student learning. These exams are often high stakes assessment tasks and provide no opportunity for feed-forward. Despite academic claims that all topics in their subject are requisite material, students are regularly able to pass these assessment tasks with unsatisfactory, and perhaps even no capacity to demonstrate learning in some topics. Furthermore, while undertaking the exam often highlights to students their learning deficiencies, it typically has no impact on their learning as they rarely receive feedback other than a mark or grade and there is no further opportunity to address these learning gaps. This paper reports on the impact of a two-staged examination process on both student learning and assurance of that learning.

PURPOSE

The aim of the staged examination process was to improve confidence that students had satisfactory knowledge in all requisite subject topics and to test its capacity to be learning-oriented in that it provides improved opportunities for students to learn while simultaneously increasing the level of learning assurance.

DESIGN/METHOD

The first stage of the process was an exam that covered all requisite subject topics. This exam consisted of multiple choice questions set at or just above the level of threshold learning outcomes. Students were required to score 80% on this exam to qualify to undertake the second part of the assessment process at a later date. Students used IFAT (Immediate Feedback Assessment Technique) cards for this stage to facilitate immediate feedback as to their strengths and weaknesses. The time between exams allowed students to review identified areas of weakness before attempting the second stage of the exam. Note: while not contributing to their final grade students who failed the first exam were also permitted to undertake the second exam as an opportunity to learn and as a means of evaluating the impact of the process. The second exam consisted of open-ended questions requiring students to explain their critical thinking and judgement used to arrive at their answer. Evaluation of the effectiveness of this process was based on a student survey, focus group discussions and an analysis of student examination scripts.

RESULTS

The threshold learning outcome exam was effective in improving assurance of learning in that students had to demonstrate satisfactory learning across topics to achieve the 80% required to 'pass' the exam. Furthermore, students reported that they used the opportunity between exams to address identified learning gaps, hence demonstrating the learning orientation and feed-forward capacity of the two stage process. However, the fact that two students who did not achieve the threshold level of 80% in the first exam were able to address their learning gaps and pass the second and harder exam suggests that an alternative to the 80% exclusion criteria should be considered.

CONCLUSIONS

The study demonstrated that a two staged examination process improved confidence in assurance of learning while providing students with an opportunity to first identify and subsequently reduce learning gaps. However, the fact that some students who failed the threshold exam demonstrated significant improvement in their understanding in the second exam suggests that more research is needed to both understand the impact of and improve the benefits from this activity.

KEYWORDS

Assurance of learning, Feed-forward, Assessment design

Introduction

Formal examinations are often used in engineering classes as the tool to 'measure' student learning. These exams are often high stakes assessment tasks that provide no opportunity for feed-forward (apply feedback to subsequent learning). Despite academic claims that all topics in their subject are requisite material, students are often able to pass these assessment tasks with unsatisfactory, and perhaps even no capacity to demonstrate learning in some topics. Furthermore, while undertaking the exam often highlights to students their learning deficiencies, it typically has no impact on their learning as they rarely receive feedback other than a mark or grade and there is no further opportunity to address these learning gaps.

To address these issues we investigated the effectiveness of a two-staged examination process on improving confidence that passing students had satisfactory knowledge in all requisite subject topics while simultaneously being learning-oriented through providing feed-forward to students on their understanding of the subject material. This paper reports on the impact of this two-staged examination process on both student learning and assurance of that learning.

Background

Much of the literature on assurance of learning focuses on the program, institutional or national level. One of the issues to be resolved in a national quality and standards framework is the "...*tension between minimum or threshold standards on the one hand, and excellence...on the other*" (Krausse, Barrie and Scott, 2012). While recognising the value in conversations at the institutional level in regard to 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 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. The authors, in discussions with academics and students at a number of universities, found regular instances of marks being given for activities that did not provide evidence of a subject's learning outcome achievement including:

- revision of pre-requisite material: for example a 20% assessment task being given in week 3 of the semester for students to revise the material in the two pre-requisite subjects,
- frequency of contributions to a discussion forum (without regard to the quality of these contributions),
- participation in an activity, for example self and peer assessment or peer review, without regard to the quality of their contributions.

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 (eg 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).

Rather than discarding all our 'during semester' assessment, Knight (2006) suggests that we design our tasks for learning oriented assessment. Carless (2007) describes learning oriented assessment as assessment designed to meet both certification and learning purposes. He characterizes learning oriented assessment as having three major components:

- Assessment tasks as learning tasks,
- Students involvement in the assessment process (self and peer assessment),
- Feedback as feed forward.

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) they recommend that academics should 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).

Hence in designing a summative examination process our specific aim 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.

In response to the above we developed a two-staged examination process. This paper reports on the impact of this process on both student learning and assurance of that learning.

Design

The two-staged examination process was implemented as follows:

The first stage of the process was an exam covering all requisite subject topics. This exam consisted of multiple choice questions set at or just above the level of threshold learning outcomes. This exam was followed by an exam review where the material was discussed and any common misconceptions were addressed.

Students answered this first stage multiple choice exam using Immediate Feedback Assessment Technique (IF-AT) cards. IF-AT cards (Figure 1) developed by Epstein Educational Enterprises allow students to immediately identify if they have answered multiple-choice questions correctly. These cards require the students to scratch off a covering over the response they think is correct (hence they are

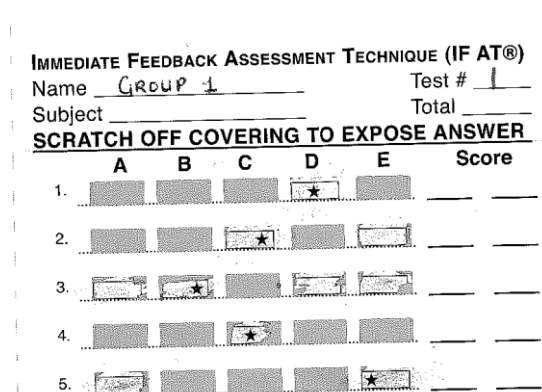


Figure 1: IF-AT card showing a star under the correct answer and multiple attempts to find the answer for Questions 2, 3 & 5.

often referred to as 'scratch cards'). If they have selected the correct response a star is revealed (Figure 1). If they selected incorrectly, they consider the remaining options, and try again.

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).

IF-AT allows students to assess their mastery of the material being assessed. 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).

Students were required to get 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. In reality since the exam covered five broad topic areas it was theoretically possible for a student to master four of the five topics knowing them extremely well (four topics representing 80% of the assessed material) and not know one topic (one topic representing 20% of the assessed material) at all. The four scaffolding recommendations previously discussed were clearly articulated to students. They were told that the intent of the first 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 the review that immediately followed the exam students were able to ask questions to clarify any misunderstandings and/or address any gaps in their learning. The fact that the IF-AT cards, in effect, grade the examination automatically, allowed the instructor to quickly view the cards and pay particular attention in the review to the most common mistakes. These were identified as questions where a cohort of students took multiple attempts to achieve the correct answer. The instructor conducted the review by discussing each question, clarifying any misconceptions and subsequently varying the question to check students' understanding.

Students had their first exam scores multiplied by 0.625 to calculate the contribution to their final examination grade. That is a score of 80% in the first stage exam scaled to 50% of a student's final examination mark, while a score of 100% in the first stage exam scaled to 60% of a student's final examination mark.

Students who received 80% or greater in the first stage exam were invited to sit, at a later date, a second exam consisting of open-ended questions requiring students to explain their critical thinking and judgement used to arrive at their answer. The time between the exams provided students with the opportunity to revise material addressing gaps in their understanding identified through undertaking the first stage exam or during the exam review. This enabled students to feed forward the feedback they received from these activities. The questions in the second stage exam were set at a more 'mastery' level and hence considerably more difficult than those in the first stage exam. This is not unreasonable given that it is the result of the second exam that determines whether students received higher than a passing grade. Again the second exam was followed by an exam review providing students with an opportunity to address knowledge gaps highlighted by the second exam and continue to learn even after the subject's summative assessments had ceased.

Students who failed to achieve 80% in the first exam were given the opportunity to sit a supplementary exam at the end of the semester. Again, undertaking the first exam and participating in the exam review provided them with an opportunity to identify the areas of the subject that they needed to revise.

the second attempt, 0.5 points on the third attempt and 0 points on the fourth and fifth attempts. This exam was followed by a review where the instructor facilitated students discussing their answers to each question. The instructor then clarified any outstanding issues or misconceptions before varying an aspect of each problem for students to first attempt then discuss their answers to check their understanding.

The second examination was held on the following Tuesday evening enabling students to review identified areas of weakness before attempting the second stage exam.

As shown in Figure 2 students who received 80% in the first exam could accept a pass and not undertake the second exam. If the student wanted to try for a higher grade (credit, distinction or high distinction) then they were required to sit for the second exam. In addition, while not contributing to their final grade students who failed the first exam (received less than 80%) were encouraged to undertake the second exam as an opportunity to learn.

Evaluation of the effectiveness of this process was based on a student survey, focus group discussions, structured observations during the examination period and an analysis of student examination scripts. Students were asked to voluntarily identify themselves on the survey to allow the authors to investigate the effect of the two-stage process on students of varying ability, on the understanding that the surveys were sealed and not analysed until after students' final grades had been posted.

Results

All students (n =28) chose to undertake both examinations and attend both review sessions. Furthermore, all students agreed to participate in this research and also chose to voluntarily identify themselves on the survey.

Of the 28 students who sat the first exam 15 (54%) achieved 80% or greater. Of these 5 achieved 100%, 5 achieved $\geq 90\%$ and $< 100\%$ and 5 achieved $\geq 80\%$ and $< 90\%$.

Hence 13 (46%) students sat the first exam and achieved less than 80%. Of these 7 achieved $\geq 70\%$ and $< 80\%$, 3 achieved $\geq 60\%$ and $< 70\%$ and 2 achieved $\geq 50\%$ and $< 60\%$. One student did not receive a grade due to academic misconduct.

The student survey amalgamating the Strongly Disagree (SD) and Disagree (D) and Strongly Agree (SA) and Agree (A) results respectively, are shown in Table 1.

Table 1: Results from the preliminary student survey amalgamating the Strongly Disagree (SD) and Disagree (D) and Strongly Agree (SA) and Agree (A) results respectively.

QUESTION	SD/D	N	A/SA
The requirement to achieve 80% in the first exam of the feedforward summative assessment meant that I tried to understand all topics in the subject.	0%	39%	61%
The Immediate feedback provided by the IFAT cards in the multiple choice exam assisted me to identify areas of the subject material that I did not fully understand?	4%	32%	64%
I revised the subject material to address these indentified gaps before the second exam.	0%	26%	74%
The revision helped me in answering the questions in the second exam.	7%	14%	79%
Having a two staged feedforward summative assessment activity changed the way I approached this subject compared to other subjects.	4%	39%	57%
Having a two staged feedforward summative assessment activity motivated me to learn more than I would typically in other subjects.	4%	29%	68%

Discussion

Arguably, the most significant impact of the two-stage exam process was the improved confidence in the assurance of learning provided to both the instructor and students as a

result of the requirement for students to demonstrate satisfactory learning in each topic to achieve the 80% required to 'pass' the first exam. This requirement also proved to be a significant motivator for student learning with 61% reporting that it motivated them to understand all the topics in the subject (see Table 1).

The immediate feedback provided by the IF-AT cards assisted students to identify gaps in their understanding. Students reported enjoying the opportunity to have a second attempt at a question they got wrong as in reconsidering their answer they had an opportunity to, address a silly mistake, reassess their thinking and/or assess how much they didn't understand. Students commented that taking more than two attempts to answer a question was a good indicator to themselves that they needed to thoroughly revise this topic.

The most frequently reported benefit by students of the two stage process was that the combination of the first exam, exam review and time between exams enabled them to discover, identify and then review areas of the subject that they did not understand. Most students (74%) reported that they used the opportunity between exams to address identified learning gaps, with 79% reporting that this revision helped them to answer the questions in the second exam. Typical student comments were: "*it was good that I found out what I didn't understand*"; the "*review after first exam was good for learning*" and "*it provided good revision before the second exam*"; and the process "*assisted me to identify areas of the subject I didn't understand*".

These results demonstrate the learning orientation and feed-forward capacity of the two stage process. Although one student who received 100% in the first exam commented that the exam was "*too simple*" so it didn't help them "*to identify errors in their understanding*". However, they did comment that the variation provided in the exam review enabled them to identify some gaps in understanding.

57% of students reported that the two-stage feed forward assessment activity changed the way they approach the subject compared to other subjects. Students most commonly reported that the activity gave them the opportunity to find out what they didn't understand "*improve my weaknesses after the first exam*" and allowed them time to revise before sitting the second exam.

It is unclear why students felt this was better than for example a more traditional formative process where students are provided problems to revise before their final assessment tasks. However some comments in the focus group suggests that it was because both parts of the exam were worth marks improving motivation to revise, secondly, it is only under the exam conditions that they actually get to test their understanding and thirdly it was somewhat satisfying or a form of redemption to receive an opportunity and subsequent credit for addressing mistakes they made in the first exam.

68% of students reported that the two-stage feed forward assessment activity motivated them to learn more than they would typically in other subjects. The main motivator for this change appeared to be "*because I needed to understand all the topics in the course to pass the exam*" and that "*I needed to understand more on this subject*" to pass. Other students commented that the activities "*motivated me to learn more on my weak areas of the subject*" and "*it reminds me what areas of topics I not understand so I'll try harder to learn and get higher marks*". One student commented that "*I have a chance to find out my weaknesses of the subject, to encourage me to study again to ensure I will fully understand all the subject in the course*".

The most common complaint in regard to the process reported by students in both the survey and the focus group was that it was too compressed leaving insufficient time for students to rigorously address all learning gaps they had identified. This was articulated in the following free response comments: "*the time is too compressed no time to digest the material*"; "*the learning is very compressed it effected my performance in the exam*"; "*not enough time after the first exam, after finding out what I misunderstood not enough time to*

revise this material". Some students also raised the issue that having to get 80% in the first exam put them under too much pressure and that the requirement was unreasonable. A typical comment being "*the exam is too much pressure*" and "*it is not easy to achieve 80%*".

In the described trial this matter was made worse with approximately 72% of the class enrolling late and subsequently not having access to the pre-work material before the block mode lectures. This provided a very compressed time frame for the students to learn the subject material. Not surprisingly, 93% of the students who failed the first exam were late enrolments.

As previously stated in the results section of this paper all students who failed the first threshold level exam chose to undertake the second and harder exam. This was promoted as a further opportunity to both learn and identify gaps in their learning. Subsequent marking of the second exam revealed that two students who failed the first exam actually passed the second exam demonstrating a significant improvement in their understanding. One student achieved more than 50% in each individual question in the second exam, while the other achieved greater than 50% in all but one question in which they achieved 41%. These students' papers were awarded a grade without having to sit the supplementary exam as they had demonstrated satisfactory knowledge across all topics in the second exam (that was set at above threshold level). This anomaly was somewhat surprising especially given the fact that there was only one day between the exams to revise and that the second exam was significantly more difficult than the first. The authors requested a dialogue with these students to discuss this anomaly. One student agreed. This student described a preference for longer questions which were allocated more time to complete than the multiple-choice format. Arguably more importantly he reported that while he undertook significant study between the two exams focusing on the material he got wrong in the first exam his main reason for failing the first exam (he received 73%) "*was careless reading*" and problems with comprehension (recall the trial was conducted in Hong Kong where all students had English as a second language). This highlights the need when setting multiple-choice exams to pay particular attention to avoiding any ambiguity to reduce the chances of students misinterpreting the question. This is especially important when students are not being examined in their primary language.

To further analyse the effectiveness of the process for assurance of learning the authors examined the number of questions for which students scored at least 30% in the second exam. The 30% level was chosen to roughly correlate with what a student with satisfactory subject knowledge might be expected to achieve given that the second exam questions were set at a higher level with the intention of using it as discriminator for students to receive higher than a passing grade.

The second exam consisted of four questions. Three questions covered a single topic while the fourth question covered two topics. For comparison these results were arranged into two categories: students who passed the two stage exam process and those that were required to sit the supplementary exam even though their combined mark for exam one and two was greater than 50%. That is, the students who under normal circumstances (no requirement to achieve 80% in the requisite material) would have received a passing exam grade.

In interpreting the results shown in Figure 4 the following points should be considered:

No student was required to undertake the second exam. Those that passed the first exam only had to sit the second exam if they wanted to attempt to achieve a credit, distinction or high distinction. Those that failed the first exam had to sit the supplementary exam but were encouraged to do the second exam as a means of revision and further indicating the areas in which they needed to revise. Also only 26% of students agreed that they revised gaps identified by the first exam before undertaking the second exam. Hence student performance in this exam could not be considered an absolute evaluation but is however indicative of a student's knowledge across subject areas.

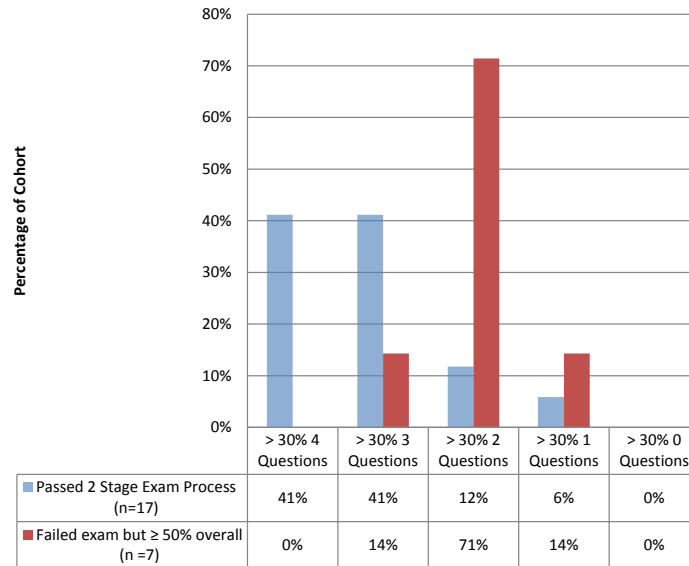


Figure 4: Comparison between the students who passed the two-stage exam process and those that failed but received a combined examination mark of greater than 50%.

The results plotted in Figure 4 show 82% of students (14 students of 17) who passed, achieved greater than 30% in at least three of the four exam questions compared to only 14% of the students (1 student of 7) who were required to sit the supplementary exam. Previously, when the authors used a single examination paper, it typically contains approximately 60% of questions aimed at the pass or satisfactory level, with the remaining 40% of questions used as the discriminators to award students a higher grade. Often this resulted in students who understand one or more topics well, passing the exam without satisfactory knowledge of the remaining requisite material. Indeed, Figure 4 shows that if the two exams were considered as one without the condition to achieve 80% in the requisite material then 6 of the 7 failing students who would have received greater than 50% overall, and hence passed, were unable to demonstrate 30% achievement in at least two topic questions.

This result clearly demonstrates the potential of a threshold exam component to improve confidence in assurance of learning for students who pass a subject. Furthermore, the two-stage exam with the exam reviews, is learning oriented, enabling students to continue to learn even after the summative stage of the examination is finished.

Recommendations

In consideration of keeping the process neutral in regard to academic workload we intend to change the process in the next trial as follows:

Stage 1: All students sit the first stage threshold exam held at the end of the semester and answered using IF-AT cards. This would be followed by an exam review.

Stage 2: One week later (allowing time for students to undertake further revision) students sit a second exam as follows:

- i. Students who received greater than 80% in the first exam would be invited to sit a paper consisting of open-ended high-level questions that require critical evaluation and reasoning to answer for credit towards a higher grade (in the case of UTS credit, distinction or high distinction)
- ii. Students who failed to achieve 80% in the first exam, would sit another threshold level exam, again answered using IF-AT cards. Students would also be required to get 80% in this examination to achieve a pass in the subject.

These exams would again be followed by an exam review undertaken immediately before the start of the next semester.

Conclusions

The study demonstrated the potential of a two stage examination process including exam reviews to improve confidence in assurance of learning while providing students with an opportunity to first identify and subsequently reduce learning gaps. However, the fact that two students who failed the threshold exam demonstrated satisfactory achievement in the mastery exam across topics suggests that more research is needed to improve the design, impact and benefits of this process.

References

- 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.
- Carless, D. (2007) Learning-oriented assessment: conceptual bases and practical implications, *Innovations in Education and Teaching International*, vol. 44: 57-66.
- 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 Educational Enterprises <http://www.epsteineducation.com/home> last viewed 26th May 2012.
- Epstein M. & Brosvic G. Students Prefer the Immediate Feedback Assessment Technique. *Psychological Reports*, 90 (2002), 1136-1138
- 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.
- Knight P. (2006): The local practices of assessment, *Assessment & Evaluation in Higher Education*, 31:4, 435-452.
- Krause, K., Barrie, S. & Scott, G. (2012) Mapping Learning and Teaching Standards in Australian Higher Education: An Issues and Options Paper. *Teaching and Learning Standards: Issues and options*. <http://www.altf.org/category/resources/standards-resources/>
- 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.
- Willey K. & Gardner A. (2012) Collaborative Learning Frameworks to Promote a Positive Learning Culture. Accepted for Frontiers in Education Conference Seattle, Washington, October 3-6, 2012.

Copyright statement

Copyright © 2012 Willey & Gardner: The authors assign to AAEE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive licence to AAEE to publish this document in full on the World Wide Web (prime sites and mirrors), on Memory Sticks, and in printed form within the AAEE 2012 conference proceedings. Any other usage is prohibited without the express permission of the authors.