

Using self- and peer-assessments for summative purposes: analysing the relative validity of the AASL (Authentic Assessment for Sustainable Learning) model

Sean Kearney*, Timothy Perkins and Shannon Kennedy-Clark

School of Education, University of Notre Dame, Sydney, Australia

The purpose of this paper is to provide a proof of concept of a collaborative peer-, self- and lecturer assessment processes. The research presented here is part of an ongoing study on self- and peer assessments in higher education. The authentic assessment for sustainable learning (AASL) model is evaluated in terms of the correlations between sets of marks. The article provides an explanation of the assessment process, and analyses sets of marks as a means of justifying the validity of the process. The results suggest that students, even those with no prior experience in peer- or self-evaluation, in their first year of tertiary study, under the right conditions, are able to accurately judge their own work and make reasonably accurate judgements of the work of their peers. While previous studies have expounded the benefits of self- and peer assessments in tertiary study, undertaking a prescribed process, such as AASL, has a further implication in allowing others to replicate the process with reasonable assuredness of the validity of the process across various fields of study.

Keywords: assessment; self-assessment; summative; validity; ASPAL

Introduction

The role, value and importance of assessment in higher education have been extensively researched and discussed in the literature (Bloxham and Boyd 2007; Boud and Associates 2010; James, McInnis, and Devlin 2002; Shepard 2000). The result is a general consensus that assessment is a driving force in the learning process of students in higher education (Boud 1990; Joughin 2010). The link between assessment and student learning is such that assessment can be utilised not only to foster student learning, but also to enhance specific aspects of learning that are deemed to be the most beneficial and useful to students both in their courses and throughout their lives. Changes in the assessment of students in Australian universities began over a decade ago (James, McInnis, and Devlin 2002), in recognition of the critical role of assessment in the teaching and learning continuum. These changes have seen a shift from measuring an individual's discrete knowledge of a subject to a far more open, collaborative and authentic approach to assessing student learning.

Questions on why and how to assess student learning have garnered increasing attention, which in turn has contributed to a renewed interest in the use of self- and peer assessments to enhance students' learning in higher education. This current interest is in response to the recognition that improved assessment practices are

*Corresponding author. Email: sean.kearney@nd.edu.au

needed to replace forms of assessment that focus on superficial and or discreet representations of knowledge (e.g. multiple-choice examinations and quizzes), which are seen as separate from learning (Shepard 2000). Although there were early concerns about using self- and peer assessments for summative purposes (Boud 1989), more recent studies have shown that self-assessment is reliable, and can and should be used in higher education (Falchikov 2005; Hafner and Hafner 2003; Stefani 1994). This article consolidates earlier research by the authors (Kearney 2013; Kearney and Perkins 2014), and reports on the correlations between self-, peer- and lecturer assessment marks utilising the authentic assessment for sustainable learning (AASL) model. We also discuss the process by which these procedures occur, using the authentic self- and peer assessment for learning (ASPAL) framework.

In our earlier research, we examined the development of AASL and ASPAL (Kearney 2013; Kearney and Perkins 2010), and the level to which ASPAL is successful in improving the engagement of students who undertake the process (Kearney and Perkins 2014). In this paper, we put forward a proof of concept that the use of AASL, in particular, is a valid assessment strategy that demonstrates that self- and peer assessments can be used for summative purposes without sacrificing the academic integrity of the course in which it is used. A summary of results from one cohort of first-year education students is provided to demonstrate the application of the AASL and ASPAL models.

The issue of validity and reliability has been raised by Falchikov and Goldfinch (2000), where they noted that validity in assessment is the ‘degree of agreement between [the teachers’] marks and those awarded by their students’ (288), whereas reliability is about agreement in ratings between and amongst peers. This is an important distinction in that the current paper utilises the correlation between self-, peer- and lecturer marks of a particular authentic assessment task to evaluate the model’s validity. The paper also seeks to verify whether students of lower abilities overmark themselves and students of higher ability undermark themselves during self-assessment, as has been found in previous research (Boud and Falchikov 1989; Dochy, Segers, and Sluijsmans 1999). Lastly, the paper evaluates the usefulness of the AASL and ASPAL models to be used both for formative learning and a summative assessment mark with reasonable validity.

Background

Self- and peer assessments in higher education

The use of self- and peer assessments has been widely researched and reported on in the literature for over 20 years (Bloxham and West 2004; Dochy, Segers, and Sluijsmans 1999; Falchikov and Goldfinch 2000; Goldfinch and Raeside 1990; Hanrahan and Isaacs 2001; Thomas, Martin, and Pleasants 2011). The literature on self-assessment is overwhelmingly positive; some of the benefits that have been associated with self-assessment include: improvements in the quality of learning (Hassmen, Sams, and Hunt 1996); student awareness of their strengths and weaknesses (McNamara and Deane 1995); improvements in capacities of self-awareness and monitoring learning (Falchikov and Boud 1989); improvements in motivation (Norton 2004); improvements in learning autonomy (Peckham and Sutherland 2000); improved academic engagement (Kearney and Perkins 2014); and the promotion of lifelong learning (McAlpine 2000). Other studies have focused on student

expectations, and identified that students want expeditious feedback and that feedback is only valued if delivered in a timely manner (Hattie 2003; James, McInnis, and Devlin 2002; Zariski 1996). Other studies have consolidated research in the field by putting forward generalised conclusions. For example, Dochy, Segers, and Sluijsmans's (1999) literature review, which covered over two decades of research on self-, peer- and co-assessments, established that, overwhelmingly, the findings concerning the use of self-assessment were positive.

There are a number of studies on peer assessment that expound the benefits of its implementation (Bostock 2000; Falchikov 2005; Topping 2000). These benefits include a better sense of responsibility, accountability and motivation (Black et al. 2003). More specifically, peer assessment has been linked to an increased understanding of content, standards and students' achievement (Bostock 2000). While there has been some concern with regard to the reliability and accuracy of peer assessment (see Boud, Cohen, and Sampson 1999; Taylor 2008), these concerns have focused on the evaluation of students rather than their learning (Boud, Cohen, and Sampson 1999) and the first-year experience (Taylor 2008). These concerns are not generally supported by empirical research, which finds satisfactory levels of agreement when students are working with criteria (Falchikov and Goldfinch 2000). It has also been put forward in the research that inaccuracies in peer assessment can also be further mitigated when students receive appropriate training in assessing peers (Falchikov 2005). Providing training prior to marking and the provision of marking criteria are both prominent features of the ASPAL model.

Authentic self- and peer assessments

Although self- and peer assessments are commonly used terms in higher education and are easily identifiable, the notion of authentic assessment or authenticity in learning is more difficult to define or identify. A conventional understanding of authentic assessment is one in which there is a strengthened link with the world outside the classroom, thus enhancing learning by making it relevant to another context (e.g. developing work place readiness and professional skills) (Darling-Hammond, Aness, and Falk 1995; Wehlage, Newmann, and Secada 1996; Wiggins 1998). Research has shown strong links between the implementation of authentic assessment, as defined here, and high-quality learning (Brown, Collins, and Duguid 1989; Darling-Hammond and Snyder 2000; Ridley and Stern 1998).

The definition of authentic assessment used for the development and understanding of the AASL model and ASPAL is that the task must have relevance outside of formal education; it must allow for students to reflect and learn from one another collaboratively to extend their thinking and learning beyond what they could accomplish on their own. Sustainable assessment or learning has garnered less attention in the literature, but draws on the principles of authentic assessment. The term is used by Boud, who argues that sustainable assessment draws attention to the 'knowledge, skills and predispositions that underpin lifelong learning activities' (2000, 151). It assists students in establishing a base for lifelong learning and in developing the potential to engage, as intrinsically motivated learners, in reflective practice; and, also encourages autonomous self-sustained learning and critical thinking by developing capacity, influence and metacognition (Tait-McCutcheon and Sherley 2006). The implementation of sustainable assessment endeavours to 'meet the needs of the present without compromising the ability of students to meet their own future learning

needs' (Boud 2000, 151). The ideas of authentic and sustainable assessment are such that they meet the needs of students and foster the skills required for success outside of the classroom (Boud 2000; Vu and Dall'Alba 2008).

The literature is clear about the value of assessment for learning, feedback and the importance of authenticity; however, this is problematic in terms of being translated into action in university courses. Although the foundations for reform in assessment are present (Boud and Associates 2010), in many university courses, forms of assessment are too entrenched to facilitate genuine change. This issue is further compounded by the difficulty that many educators face in making changes to assessments and unit outlines. University processes can inhibit change in assessment practices. Hence, there seems to be a perpetuation of the dichotomy between conventional approaches of assessment and the ideals of critical thinking, autonomy and thoughtfulness in education (Robinson, cited in Azzam 2009), which are foundational aspects of the AASL and ASPAL models.

AASL and ASPAL

The AASL and ASPAL models were developed within the context of the Australian Government's call for reform in higher education, the launch of the Assessment Futures Project at the University of Technology, Sydney, and the publication of the Assessment 2020 paper (Boud and Associates 2010), which called for reform in the nature of assessment in higher education. The conceptual foundation of the models and the results of the ASPAL process have been previously published (Kearney 2013; Kearney and Perkins 2014). This paper analyses the AASL model in the context of its validity in providing a means of summative assessment, and functions as a proof of concept. AASL and ASPAL work together as a complete assessment framework.

An important facet of this research is its reliance on a specific model that can be easily replicated. That is, the model can be applied across a range of learning contexts in higher education from undergraduate to postgraduate units. The ASPAL model is the delivery method for AASL and involves, prior to the submission of the task: surveying students about their experience with self- and peer assessments; collaboratively developing marking criteria for the authentic assessment students carry out; undertaking a pilot marking session, in which students are given the opportunity to practice mark similar assignments to the ones they will be completing. At this point in the process, students submit their task, and then the AASL process of self- and peer markings occurs. Once all the marking has been completed a debriefing session is held in which students receive their feedback from peers and the lecturer. The final stage in the ASPAL process is another survey to elicit students' feedback about the process.

In the AASL process, lecturer assessment, self-assessment and peer assessment are combined to produce a summative grade for the student. The application of the model is fairly structured in order to provide clarity regarding the task, and clear guidelines regarding the evaluation of assessments for the students. In this model, the allocation of the marks are weighted so that the lecturer's assessment of student work accounts for 40% of the overall mark, acting as a moderator for the self- and peer assessment marks. Peer assessment accounts for 30% of the assessment marks. In AASL, two peers collaboratively mark another student's anonymous task. In this process, the peers collaborate during the process, with verbal communication being

an integral aspect of the evaluation; however, they do not have to agree on the mark given. Each peers' mark accounts for 15% of the overall mark. In the final stage of the assessment evaluation, a student will mark their own task against the criteria, which they had a part in creating. The self-assessment accounts for the final 30% of the overall mark, thus providing students with a significant influence on their own marks. This empowers students to critically reflect on their assessment in relation to their peers, and provides a valuable space for workplace readiness skills, such as self-reflection and evaluation. The model, as outlined in the process, allows for timely and triangulated feedback and results for students.

Research design

In the case presented in this article as a proof of concept, approximately 280 students in an undergraduate Bachelor of Education degree (primary and birth to 12 years), in their second semester of their first year of study, undertook this process as an assignment within a unit of study on mathematics. The University of Notre Dame, Sydney, has a 13-week semester and all units are undertaken face-to-face.

The assessment priming and process was conducted over the duration of the unit (see Figure 1). In the second week of the semester, the students were presented with the idea of using self- and peer assessments to grade their assignment, which was the creation of a lesson plan for primary mathematics. As per the ASPAL process, all students were surveyed in week 3 before the start of the process. The surveys sought to ascertain their previous experience with self- and peer assessments, their overall feelings about assessment throughout their course, and their current levels of engagement in their course.

The next step in the process was the development of the marking criteria, which took place during week 4. Although university regulations stipulate that marking criteria rubrics must be published in unit outlines before the start of each semester, the students and the lecturer were able to discuss the marking criteria and ensure that the criteria were worded in a way that made them clear and transparent to all students. The essence of the criteria remained unchanged, in accordance with university regulations; however, the criteria descriptors were rewritten by the students, with lecturer approval. In week 5, a pilot marking session was held. Students were given sample lesson plans to mark, using the criteria descriptors they had written the

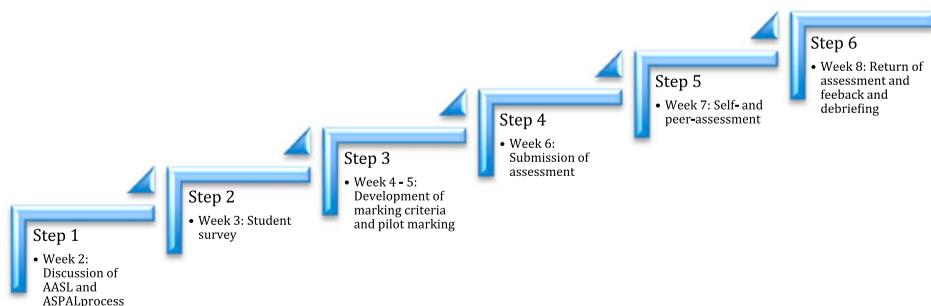


Figure 1. ASPAL weekly process.

previous week and practised marking the samples with the rubric; two lesson plans were marked during the session. The assignment was due in week 6.

The assignment that utilised the AASL and ASPAL framework was submitted in week 6 of the semester, and the self- and peer marking processes took place in week 7. The students were required to submit three copies of their assignment, one with a standard university cover sheet with their student number and two with no designation; the lecturer coded the assignments without student identification numbers to ensure anonymity. In the lecture in week 7, students were randomly paired and each pair of students received two copies of two different, anonymous assignments to peer mark. At the conclusion of the peer marking, students marked their own assignment and handed in all the grades.

In week 8, students received their marks and all of their feedback from their peers and the lecturers, and a debriefing session was held; this allowed the lecturers two weeks to mark and moderate the assignments. Between weeks 6 and 7, the assignments were marked and coded, and between weeks 7 and 8, they were moderated and weighted. The debriefing session gave students an opportunity to talk about the process and convey what they thought, and it also allowed the lecturers to communicate to the students their observations and some of the trends with regard to the marks. The final step of the process was the final survey, given to students to complete, which sought their feedback about the entire process.

In this unit, attendance at lectures was compulsory and, although full attendance each week was not reached, average attendance at lectures throughout the semester was 93%. At the lecture at which the self- and peer assessments took place, there was only one absence; the student later withdrew from the unit for unrelated reasons. All marking was out of 100, and students were notified at the start of the process that a ± 15 marks acceptance range from the lecturer's mark would be prescribed to ensure there was no inflation of marks, specifically in regards to the self-assessment marks. Any mark that fell outside of the range would not be included as part of the AASL average as these were deemed to be outliers. In this application of the process, there were only five such instances, or less than 2% of the sample.

Results

The data gathered were analysed to determine the variation in results by comparing the lecturer assessment mark, the self-assessment mark, the peer assessment mark and final assessment results. The data were also used to test previous studies' contentions that students of higher ability undermark themselves and students of lower ability overmark themselves (Boud and Falchikov 1989; Dochy, Segers, and Sluijsmans 1999). Overall, 284 students participated in self- and peer assessments and 83% of all marks were within a 5% variance (five marks). Out of the remaining, 17% that did not fall within that five-mark variance, 42% of those had the lecturer mark and the final mark within five marks. Therefore, less than 10% of all marks (lecturer, self and peer) had more than a 5% variance between marks.

Variance in the marking process

In order to eliminate outliers, the researchers considered any two marks that were within five percentage points an accurate assessment of that particular task. Marking

assessment tasks at the tertiary level can be quite subjective, and both internal moderation of assessments and external benchmarking exercises between institutions have shown that a five-mark discrepancy is an allowable margin of error (Rone-Adams and Naylor 2009). Therefore, with regard to the results of this process 50% of students self-assessed with reasonable consistency with the lecturer and 45% of students assessed their peers (when averaged between two peers) within 5% of the lecturer's assessment.

Table 1 presents both the averages and standard deviations of the lecturer-, self- and peer assessments and the final mark for the assignment. The results indicate that the average marks for the self- and peer markings were both within two percentage points of the lecturer mark, and that the lecturer mark, which accounts for 40% of the total final mark, worked as a moderator for the self- and peer marks. The mean final mark and the lecturer marks were almost identical, with a 0.01 mark variance, indicating a very strong case for the validity of the AASL percentages.

Table 2 illustrates the self-assessment marks. Although 23% of students marked themselves 10 marks or more outside that of the lecturer, 24% of students gave themselves marks that were less than three marks outside the lecturer's mark.

In the case of peer vs. lecturer marks, 26% of students marked their peers more than 10 marks outside that of the lecturer, while 23% of students marked their peers assignments less than three marks difference from the lecturer (see Table 3).

As can be seen in Table 4, the variance between the lecturer mark and the final mark did not fall outside of the 15-mark margin set as the top end of the scale of acceptable discrepancy (Kearney 2013). However, 85% of all marks fall within the five-mark range. Hence, though there was variation between the peer and lecturer

Table 1. Comparison of means and standard deviation for lecturer, self, peer and final.

	<i>N</i>	Mean	Standard deviation
Lecturer mark	284	65.78	9.00
Self mark	284	67.60	6.91
Peer mark	284	64.13	9.53
Final mark	284	65.79	6.84

Table 2. Comparison of lecturer marks and self-evaluation marks.

Percent variance (%)	Percent occurrence (%)
0	8
1	9
2	7
3	10
4	9
5	6
6	7
7	8
8	5
9	3
10	4
11–15	16
More than 15	7

Table 3. Comparison of lecturer marks and peer evaluation marks.

Percent variance (%)	Percent occurrence (%)
0	6
1	10
2	7
3	10.5
4	6
5	6
6	5
7	6
8	4
9	3
10	7
11–15	16
More than 15	10

Table 4. Comparison of lecturer marks and final marks.

Percent variance (%)	Percent occurrence (%)
0	23
1	12
2	10
3	12.5
4	13
5	5
6	6
7	5
8	4
9	3
10	1.5
11–15	4
More than 15	0

marks, the combination of the three marks resulted in a fairly even distribution of marks.

Table 5 shows the raw results of the process and indicates that nearly 50% of all students had at least one of the four marks (lecturer, peer1, peer2 and self) that fell more than 10% outside the final mark, which indicates some inconsistency with student marking.

Student self-evaluation

As seen in Table 6, the relative reliabilities of the top quartile to more accurately self-assess are found to be moderate, with a difference between self-marks and lecturer marks in this quartile to be -6 ; whereas the bottom quartile overscored themselves by an average of almost 10 marks ($+10$). The most accurate of the self-assessment occurred in the second quartile where students scored themselves less than one mark difference on average than the lecturer (-0.7). While the results here are significant, in that they confirm the fears expressed by Boud and Falchikov (1989) and later dispelled by Stefani (1994), they lend further credence to utilising

Table 5. Comparison of lecturer marks, self-evaluation marks, peer evaluation marks and final marks.

Percent variance (%)	Percent occurrence (%)
0	3
1	<1
2	5
3	6
4	3
5	12
6	1
7	5.5
8	5.5
9	3.5
10	5.5
11–15	27
16–20	13
More than 21	8.5

Table 6. Lecturer vs. self-assessment: differences in means based on performance quartiles.

Quartile group (lecturer marks)	Number in quartile	Self mark (mean)	Lecturer mark (mean)	Difference of means
75–90	59	72	78	–6
66–74	69	68.7	69.4	–0.7
59–65	61	66.2	62.5	+3.7
40–58	65	64	54.6	+9.6

the AASL percentages as part of the ASPAL process to better mediate the moderate inaccuracies encountered in self- and peer markings.

Discussion

The results presented here complement the literature on the professional competence of undergraduate students and their ability to engage in self- and peer assessments with reasonable capability, if they are properly acculturated (provided with training and background information) to the process. The results also point to the need for moderation of the inaccuracies of self- and peer assessments, which can be countered with a simple process of weighting the lecturer assessment slightly higher than the self-assessment, and encouraging collaboration between peers when awarding peer marks.

The minor variance in the marks exceeded all of the researchers' expectations with regard to the process, and highlights a noteworthy factor about students' judgement of their own and others' work when properly prepared. Students found their formal preparation and collaboration with fellow students to be essential in the success of the process (Kearney and Perkins 2014). Specifically, 82% of students found seeing others' work in the process of the pilot marking and peer marking especially beneficial to their learning (Kearney and Perkins 2014). This is consistent with other

research in the field (Dochy, Segers, and Sluijsmans 1999; Hanrahan and Isaacs 2001). While the contention that the ASPAL process of preparing students for self- and peer assessments was responsible for the results cannot be definitive, student responses in the post-survey suggest that it plays a significant role (Kearney and Perkins 2014). Although the peer marks were not as accurate as the self-assessment marks, the researchers would hypothesise that with repeated exposure to the ASPAL process these would improve.

Table 4 presents the most significant results, in illustrating how the AASL and ASPAL models can be used in higher education contexts as a reliable form of self- and peer assessments. By combining the self-, peer- and lecturer marks at a percentage rate of 30–30–40, enough emphasis was placed on the process of self- and peer marking to engage the students in the process, and the inconsistencies of individuals' judgements were successfully moderated. The process of preparing students through ASPAL, specifically the pilot marking, seemed to have the most significant effect on the confidence of students going into the self- and peer marking processes.

The process was successful at engaging students with near full attendance throughout the semester; however, we cannot definitively say that students were academically engaged simply because they attended. Students' feedbacks, both positive and negative, indicated that they took it seriously, due, at least in part, to the summative nature of the process. The ASPAL process encourages continuous engagement by extending the assessment process from an arbitrary due date to a continuous process over the course of seven weeks. The process of developing marking criteria, pilot marking and assessing others' work is all part of an authentic assessment process. As pre-service teachers, this process imitates the professional work of teachers in assessing students' work; therefore, despite the summative nature of the assessment task itself, the entire process is formative authentic learning for the pre-service teacher.

It was shown here that students were able to self-assess and assess their peers with reasonable accuracy. In the study, the use of the AASL percentages is highlighted as a process to bring the marks back to a normal distribution. While establishing a normal distribution was not a goal of AASL, it helps to validate the use of the percentages to ensure the academic integrity of the self- and peer marking processes. While the use of self- and peer assessments may be gaining traction in the literature, institutions are concerned with maintaining academic integrity (Bloxham and Boyd 2007) and questions have been raised as to whether self- and peer summative marking processes can be a valid form of assessment.

In regards to discrepancies between markers, there were considerable inconsistencies in self-, peer- and lecturer markings; however, the biggest discrepancies were found in the peer marking. The encouragement of collaboration and discussion amongst peers, limiting each peer's weighting to 15%, and applying the AASL percentages can mitigate the inconsistencies in the peer marking. In addition to moderating the peers' marks in this way, the 40% weighting of the lecturer's mark can be seen as a successful moderator of the other marks, illustrated by the 0.01 mark variance between the means of the lecturer marks and the final marks of the assignment. The distribution of marks for this assignment is similar to both past and recent grade distributions for similar assignments, which provides further evidence of the validation of the process.

A final point of discussion adds to the literature on self-regulation and whether higher ability students were undermarking themselves and students of lower ability

overmarking themselves. When looking at the reliability of different ability levels to self-assess accurately, congruence was found with previous studies about the under-scoring of higher achievers and overscoring of lower achievers. Although there was inaccurate scoring, more so in the peer marking, this does not outweigh the significance of such a process and the impact it has on students' collaboration, engagement and learning (Kearney and Perkins 2014). Over 25 years ago, Cowan (1988) remarked that, despite the risks of the differences in marking between lecturers and students, the benefits of self-assessment are such that students should be relied upon to conduct themselves appropriately throughout the process. Through an analysis of the raw marks, it was found that very little of the discrepant self-marking was a deliberate attempt to inflate students' marks, but this was rather due to inaccurate judgement, which we believe will improve through subsequent iterations of self- and peer assessment processes. What this indicates is that the rubric and piloting phases of the process can develop students' skills and understanding of the task and evaluation.

With regard to the validity of the marks, the following summary can be made: 30% of all marks were within a five-percentage point margin; 82% of all marks had at least two of the four marks (self, peer, lecturer, final) within five percentage points. When the AASL percentages were applied to the raw marks given by the students, the difference between the means of the lecturer marks and the final marks was 0.01, and 75% of the all marks were within five percentage points. What we are trying to demonstrate here is not that the AASL and ASPAL frameworks are without issues, but that the framework offers educators a scaffold for self- and peer assessments that can enable a reliable process for developing students' skills to evaluate and assess their own work and the work of others. These skills are essential in a number of professions, such as teaching, and in this respect the combination of AASL and ASPAL can provide an authentic assessment experience for students that has workplace ready relevance.

Conclusion

The findings about the validity of self- and peer assessments complement our previous findings about the impact of this process on student learning, collaboration and engagement (Kearney and Perkins 2014) and supplement the growing literature expounding the benefits of the use of self- and peer assessments to enhance the student experience (Falchikov and Boud 1989; Hassmen, Sams, and Hunt 1996; McNamara and Deane 1995; Norton 2004; Peckham and Sutherland 2000).

Further research is necessary in the area of self- and peer assessments for summative purposes. While self- and peer assessments are a popular topic, the majority of literature expounds the benefits from a formative point of view. While this is an important aspect of self- and peer assessments, previous research indicates that, unless the assessment process is summative in nature, students will not engage at the same level they otherwise would (Boud 1990). Ongoing research utilising these models continues, specifically with reference to whether students' judgement improves over time, and whether or not participation in self- and peer assessments has an effect on learning efficacy over time.

The current paper has examined the validity of self- and peer markings using the AASL percentages as part of the ASPAL process to further substantiate and validate the use of self- and peer assessments in tertiary learning environments, while

ensuring academic integrity is maintained. The results revealed that students from an early stage in their tertiary education were able to judge their own work as well as peers' work with reasonable accuracy. These findings have implications for expanding both self- and peer assessments in tertiary education with reasonable assurance that the academic integrity of the subject, and consequently the institution, will be retained.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Sean Kearney is associate dean in the School of Education, Sydney, at the University of Notre Dame Australia. He lectures in educational psychology and general pedagogical methods and has been nationally recognised for innovative methods to enhance student learning. His research includes cross-cultural service learning, assessment reform in higher education and beginning teacher attrition and induction.

Timothy Perkins is a senior lecturer in the School of Education, Sydney, at the University of Notre Dame Australia. He lectures in mathematics education and has been nationally recognised for innovative methods to enhance student learning. He is currently completing his PhD on mathematics anxiety in pre-service primary teachers.

Shannon Kennedy-Clark is a senior lecturer and secondary course coordinator at the University of Notre Dame Australia. She is an active educational researcher, and she has worked independently and in collaboration with both Australian and international colleagues on a wide variety of educational research projects. Current research activities comprise the analysis of individual and group problem solving with computer-supported learning, design-based learning and long-term design projects, pre-service teacher development, measuring creativity in assessment and academic literacy.

References

- Azzam, A. M. 2009. "Why Creativity Now? A Conversation with Sir Ken Robinson." *Educational Leadership* 67 (1): 22–26.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. 2003. *Assessment for Learning: Putting It All into Practice*. Maidenhead: Open University Press.
- Bloxham, S., and P. Boyd. 2007. *Developing Effective Assessment in Higher Education: A Practical Guide*. Maidenhead: Open University Press.
- Bloxham, S., and A. West. 2004. "Understanding the Rules of the Game: Marking Peer Assessment as a Medium for Developing Students' Conceptions of Assessment." *Assessment and Evaluation in Higher Education* 29 (6): 721–733.
- Bostock, S. 2000. "Student Peer-assessment." Accessed March 2. http://www.reading.ac.uk/web/FILES/engageinassessment/Student_peer_assessment_Stephen_Bostock.pdf
- Boud, D. 1989. "The Role of Self-assessment in Student Grading." *Assessment & Evaluation in Higher Education* 14 (1): 20–30.
- Boud, D. 1990. "Assessment and the Promotion of Academic Values." *Studies in Higher Education* 15 (1): 1–13.
- Boud, D. 2000. "Sustainable Assessment: Rethinking Assessment for the Learning Society." *Studies in Continuing Education* 22 (2): 151–167.
- Boud, D., R. Cohen, and J. Sampson. 1999. "Peer Learning and Assessment." *Assessment & Evaluation in Higher Education* 24 (4): 413–426.
- Boud, D., and N. Falchikov. 1989. "Quantitative Studies of Student Self-assessment in Higher Education: A Critical Analysis of Findings." *Higher Education* 18: 529–549.

- Boud, D., R. Sadler, G. Joughin, R. James, M. Freeman, S. Kift, F. Dochy, et al. 2010. *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education*. Sydney: Australian Learning and Teaching Council.
- Brown, J. S., A. Collins, and P. Duguid. 1989. "Situated Cognition and the Culture of Learning." *Educational Researcher* 18 (1): 32–42.
- Cowan, J. 1988. "Struggling with Students' Self-assessment." In *Developing Student Autonomy in Learning*, 2nd ed., edited by D. J. Boud, 192–210. London: Kogan.
- Darling-Hammond, L., J. Ancess, and B. Falk. 1995. *Authentic Assessment in Action: Studies of Schools and Students at Work*. New York: Teachers College Press.
- Darling-Hammond, L., and J. Snyder. 2000. "Authentic Assessment of Teaching in Context." *Teaching and Teacher Education* 16: 523–545.
- Dochy, F., M. Segers, and D. Sluijsmans. 1999. "The Use of Self, Peer and Co-assessment in Higher Education: A Review." *Studies in Higher Education* 24 (3): 331–350.
- Falchikov, N. 2005. *Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. New York: Routledge Falmer.
- Falchikov, N., and D. Boud. 1989. "Student Self-assessment in Higher Education: A Meta-Analysis." *Review of Educational Research* 59 (4): 395–430.
- Falchikov, N., and J. Goldfinch. 2000. "Student Peer Assessment in Higher Education: A Meta-analysis Comparing Peer and Teacher Marks." *Review of Educational Research* 70 (3): 287–322.
- Goldfinch, J., and R. Raeside. 1990. "Development of a Peer Assessment Technique for Obtaining Individual Marks on a Group Project." *Assessment and Evaluation in Higher Education* 15 (3): 210–231.
- Hafner, J., and P. Hafner. 2003. "Quantitative Analysis of the Rubric as an Assessment Tool: An Empirical Study of Student Peer-group Rating." *International Journal of Science Education* 25 (12): 1509–1528.
- Hanrahan, S. J., and G. Isaacs. 2001. "Assessing Self- and Peer-assessment: The Students' Views." *Higher Education Research & Development* 20 (1): 53–70.
- Hassmen, P., M. R. Sams, and D. P. Hunt. 1996. "Self-assessment Responding and Testing Methods: Effects on Performers and Observers." *Perceptual and Motor Skills* 83: 1091–1104.
- Hattie, J. 2003. "Teachers Make a Difference: What is the Research Evidence?" Paper presented at the Australian Council for Educational Research annual Conference on building teacher quality, Melbourne.
- James, R., C. McInnis, and M. Devlin. 2002. *Assessing Learning in Australian Universities*. Melbourne: Centre for the Study of Higher Education, University of Melbourne.
- Joughin, G. 2010. "The Hidden Curriculum Revisited: A Critical Review of Research into the Influence of Summative Assessment on Learning." *Assessment & Evaluation in Higher Education* 35 (3): 335–345.
- Kearney, S. 2013. "Improving Engagement: The Use of 'Authentic Self-and Peer-assessment for Learning' to Enhance the Student Learning Experience." *Assessment & Evaluation in Higher Education* 38 (7): 875–891.
- Kearney, S. P., and T. Perkins. 2010. "Developing Students' Capacity for Innovation, Creativity and Critical Thinking through Contemporary Forms of Assessment: A Case Study in Progress." Paper presented at the ATN assessment Conference, University of Technology, Sydney.
- Kearney, S. P., and T. Perkins. 2014. "Engaging Students through Assessment: The Success and Limitations of the ASPAL (Authentic Self and Peer-assessment For Learning) Model." *Journal of University Teaching and Learning Practice* 11 (3): 1–13.
- McAlpine, D. 2000. "Assessment and the Gifted." *Tall Poppies* 25 (1). Accessed November. http://www.tki.org.nz/r/gifted/pedagogy/tallpoppies_e.php
- McNamara, M. J., and D. Deane. 1995. "Self-assessment Activities: Toward Language Autonomy in Language Learning." *TESOL* 5: 17–21.
- Norton, L. 2004. "Using Assessment Criteria as Learning Criteria: A Case Study in Psychology." *Assessment & Evaluation in Higher Education* 29 (6): 687–702.
- Peckham, G., and L. Sutherland. 2000. "The Role of Self-assessment in Moderating Students' Expectations." *South African Journal for Higher Education* 14 (2): 67–72.

- Ridley, K. L., and B. S. Stern. 1998. "Using Authentic Assessment and Qualitative Methodology to Bridge Theory and Practice." *The Educational Forum* 62 (2): 178–185.
- Rone-Adams, S., and S. Naylor. 2009. "Examination of the Inter-rater Agreement among Faculty Marking a Research Proposal on an Undergraduate Health Course." *The Internet Journal of Allied Health Science and Practice* 7 (4): 1–11.
- Shepard, L. A. 2000. "The Role of Assessment in a Learning Culture." *Educational Researcher* 29 (7): 4–14.
- Stefani, L. A. J. 1994. "Peer, Self and Tutor Assessment: Relative Reliabilities." *Studies in Higher Education* 19 (1): 69–75.
- Tait-McCutcheon, S., and B. Sherley. 2006. "In the Hands of the Learner: The Impact of Self-Assessment on Teacher Education." Mathematics Education Research Group of Australasia 2006 Conference Proceedings, 353–359, Canberra.
- Taylor, J. A. 2008. "Assessment in First Year University: A Model to Manage Transition." *Journal of University Teaching and Learning Practice* 5 (1): 19–33.
- Thomas, G., D. Martin, and K. Pleasants. 2011. "Using Self and Peer-assessment to Enhance Students' Future Learning in Higher Education." *Journal of University Teaching and Learning Practice* 8 (1): 1–17.
- Topping, K. J. 2000. *Peer Assisted Learning: A Practical Guide for Teachers*. Cambridge, MA: Brookline Books.
- Vu, T. T., and G. Dall'Alba. 2008. "Exploring an Authentic Approach to Assessment to Enhance Student Learning. In Re-imagining Higher Education Pedagogies." Proceedings of the Australian Association for Research in Education Conference, Brisbane, Australia.
- Wehlage, G., F. Newmann, and W. Secada. 1996. "Standards for Authentic Achievement and Pedagogy." In *Authentic Achievement: Restructuring Schools*, edited by F. Newmann and Associates, 21–29. San Francisco, CA: Jossey-Bass.
- Wiggins, G. 1998. *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass.
- Zariski, A. 1996. "Student Peer-assessment in Tertiary Education: Promise, Perils and Practice." In *Teaching and Learning within and across Disciplines*, edited by J. Abbott and L. Willcoxson, 189–200. Proceedings of the 5th Annual Teaching and Learning Forum, Murdoch University, Perth.

Copyright of Assessment & Evaluation in Higher Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.