

The assessment cycle: A model for learning through peer assessment

Daniel Lee Reinholz

Center for STEM Learning, University of Colorado, Boulder, Colorado, USA

393 UCB Boulder, Boulder, CO 80309, daniel.reinholz@colorado.edu

Daniel Reinholz is a Research Associate at the Center for STEM Learning at the University of Colorado, Boulder. He completed his PhD in mathematics and science education at the University of California Berkeley, advised by Alan Schoenfeld. His research focuses on the use of reflection to promote deep learning. His dissertation focused on the use of reflection tools in introductory college calculus.

The assessment cycle: A model for learning through peer assessment

This paper advances a model describing how peer assessment supports self-assessment. Although prior research demonstrates that peer assessment promotes self-assessment, the connection between these two activities is underspecified.

This model, the assessment cycle, draws from theories of self-assessment to elaborate how learning takes place through peer assessment. The model is applied to three activity structures described in the literature to analyse their potential to support learning by promoting self-assessment. Broadly speaking, the model can be used to understand learning that takes place in a variety of peer assessment activities: marking/grading, analysis, feedback, conferencing, and revision. This approach contrasts most studies on peer assessment, which have focused on calibration of instructor and peer grades, rather than learning opportunities.

Keywords: peer assessment, peer evaluation, peer learning, formative assessment

Introduction

This paper advances a theoretical model of how peer assessment activities support self-assessment. Although prior studies demonstrate the learning potential of peer assessment, there is still a lack of common terminology and widely accepted theoretical model of *how* such activities support learning (Kollar and Fischer 2010). In particular, peer assessment is argued to support self-assessment (Black, Harrison, and Lee 2003), but the mechanisms through which it does so are not well defined. Drawing upon Sadler's (1989) model of self-assessment, this paper aims to better establish this connection.

In this paper, peer assessment is defined as: a set of activities through which individuals make judgments about the work of others. Individuals may make judgments about real or hypothetical work; when students assess real work, it generally comes from other students in the same course (Topping 1998). This distinguishes peer

assessment from other activities (e.g., peer tutoring), which often involve students from different courses or grade levels. Beyond making judgments, students may provide feedback and conference about the work they analyse; peer assessment is an umbrella term, encapsulating a number of related activities.

Assessment can be used both to evaluate student outcomes and to support student learning. Most studies of peer assessment have focused on its role in evaluation, particularly the calibration of student and instructor grades; these studies show that peer grades generally agree with instructor grades (Falchikov and Goldfinch 2000). This paper focuses on the second, less studied purpose of peer assessment: its role in learning (Stefani 1998). Although students assigning summative grades could result in learning, assessment for learning activities are typically designed differently from evaluation activities, because they serve a different purpose.

This paper is organized as follows: (1) a literature review connects peer and self-assessment, (2) a model of learning through peer assessment is introduced, and (3) the model is applied to examples from the literature.

Assessment for learning

Assessment for learning, or formative assessment, is concerned with how to *evoke* information about learning and use it to *modify* teaching and learning activities (Black, Harrison, and Lee 2003). Formative assessment has a significant positive impact on student learning (Black, Harrison, and Lee 2003; Black and Wiliam 1998). Although eliciting and using information about student understandings has traditionally been seen as an instructor's role, peer and self-assessment are increasingly viewed as important components of formative assessment (Black and Wiliam 2009).

Peer assessment provides students with opportunities to reflect upon their own understandings, build on prior knowledge, generate inferences, integrate ideas, repair misunderstandings, and explain and communicate their understandings (Roscoe and Chi 2007). Many of these activities, such as explaining ideas (Chi et al. 1994), also deepen students' content knowledge. Thus, peer assessment can have a variety of benefits for students, such as improved: conceptual understanding, communication skills, and self-assessment skills (Black, Harrison, and Lee 2003; Falchikov 2005). This paper focuses on how peer assessment supports self-assessment, recognizing that the same learning mechanisms (e.g., explanation, collaboration) also support content understanding.

A number of reviews (e.g., Dochy, Segers, and Sluijsmans 1999; Topping 2003) provide insight into important aspects of peer and self-assessment, such as: calibration, bias, training, and learning. These reviews highlight commonalities between peer and self-assessment, such as the need for clear assessment criteria (Stefani 1998). Nevertheless, when the learning impacts of peer and self-assessment are considered, these activities are generally considered independently (e.g., Sadler and Good 2006). As a result, the theoretical mechanisms through which peer assessment supports self-assessment are still underspecified.

Self-assessment involves an individual comparing his or her performance to a desired goal to adjust and improve his or her practice. Self-assessment is closely related to self-regulation, which is widely recognized as a hallmark of competent disciplinary practice (e.g., Zimmerman 2002). Self-regulation is concerned with individuals' abilities to set goals, adopt strategies for meeting goals, and monitor progress towards goals (Boekaerts and Corno 2005). There is general consensus that self-assessment plays an important role in self-regulation (Panadero and Alonso-Tapia 2013), such as in monitoring one's performance on a task (Zimmerman 2002). Given their close

relationship, a number of authors have proposed models that synthesize self-assessment and self-regulation (Andrade 2010; Nicol and Macfarlane-Dick 2006).

The connection between self-assessment and self-regulation highlights two complementary notions of self-assessment: (1) as an instructional strategy in which students engage with their own work, and (2) as a mechanism for students to guide and regulate their own learning (Panadero and Alonso-Tapia 2013). While asking students to engage with their own work may result in them using such assessments to guide future learning, it is not necessarily the case. Rather than asking students to directly engage with their work, this paper focuses on how asking students to engage with the work of others can help students learn to regulate their own learning. Not only do such activities support self-assessment, but they may also support other types of learning through peer interactions, such as improved communication skills.

Self-assessment requires an individual to: (1) accurately conceive the desired performance, (2) accurately conceive their actual performance, and (3) act to close the gap between desired and actual performance (Sadler 1989). Henceforth, these three components of self-assessment are referred to as: (1) *goal awareness*, (2) *performance awareness*, and (3) *gap closure*. Awareness along each of these dimensions is a spectrum; as an individual improves along (1)-(3), it generally improves his or her self-assessment abilities. Thus, Sadler's (1989) three components provide focal areas for how peer assessment can help students develop self-assessment skills.

Goal Awareness

Goal awareness is an understanding of what one is trying to achieve. For complex activities (e.g., designing or teaching), the goal state is often not well understood (i.e. an individual may not have an accurate conception of highly skilled

teaching). As a result, individuals struggle to self-assess because they do not know what competent performance looks like, let alone how to produce it.

Sadler (1989) argues that direct assessment experience is one of the most effective means to improve one's goal awareness. When students analyse the work of others, they have access to a variety of examples that help them better see gradations in quality. Moreover, because the students did not create the work themselves, it is easier to analyse, because they can view it from a more distanced perspective (Black, Harrison, and Lee 2003). Thus, peer assessment seems to promote self-assessment by making otherwise invisible assessment processes more explicit and transparent. This process of analysing peer work is often scaffolded through instructor modelling, class discussions, and feedback on students' peer assessments. This scaffolding is crucial, because peer assessment is a novel activity for students (Min 2006; Smith, Cooper, and Lancaster 2002; Topping 2009).

Self-Awareness

Self-awareness is the ability to accurately judge the quality of one's own work. Making detailed judgments about one's own work is often more difficult than judging the work of others, because the assessor is 'too close' to the work, lacking a distanced, objective perspective (Black, Harrison, and Lee 2003). For instance, in writing, it is often difficult for novice (and even relatively skilled) writers to detect areas where their writing is unclear, even though it is easy for them to recognize when others' writing is unclear.

Individuals generally do a poor job of assessing their own understanding (Dunlosky and Lipko 2007), systematically overestimating their abilities (Dunning, Heath, and Suls 2004); even with limited information, peers are often able to assess the skills and characteristics of others better than of themselves. When individuals are asked

to determine whether or not they know something, they often rely on a notion of *cognitive ease* (Kahneman 2011); rather than trying to actually recall or generate the required information, they rely on a visceral feeling of how well they think they know something, which results in inaccurate judgments. Nevertheless, by explicitly prompting individuals to recall the relevant information, rather than just determine if they think they know it, accuracy can be increased (Dunlosky and Lipko 2007).

For promoting self-awareness for more complex concepts, not just recall, explanation appears to be a promising technique. When learners are forced to explain their reasoning, it provides them with opportunities to see gaps in their logic (Chi et al. 1994). In contrast, if learners are never forced to make such reasoning explicit, then they are much less likely to see the weaknesses in their reasoning. Thus, a cycle of peer assessment activities that requires students to explain their own reasoning, not just the reasoning of others, should be more likely to promote self-awareness.

Gap Closure

Gap closure is achieved by reducing the discrepancies between actual and desired performance. Attempts to reduce such discrepancies may include: increasing effort, motivation, and engagement; seeking additional information from a teacher, peers, or other source of knowledge; or revisiting initial ideas and pursuing new strategies (Hattie and Timperley 2007). Gap closure is an ongoing, piecemeal process, requiring continuous monitoring and self-regulation of one's performance (Butler and Winne 1995); individuals are unlikely to achieve gap closure through a single action. As individuals adjust their performance, they must readjust their understandings of the goal state, current state, and how to reduce the gap between them; in practice, there is a constant interplay between all three components of self-assessment. This interplay is evident in complex activities such as mathematical problem solving (Schoenfeld 1985).

The Assessment Cycle

Although peer assessment is valued for its ability to promote self-assessment, there is still no theoretical model connecting the two processes (Kollar and Fischer 2010). To address this need, the assessment cycle builds on Kollar and Fischer's (2010) framework, centred around four phases: (1) task performance, (2) feedback provision, (3) feedback reception, and (4) revision. The framework is extended to include peer analysis and peer conferencing, and also emphasize the roles of learning processes, not just learning products in assessment.

The assessment cycle aims to make the connection between peer and self-assessment in a domain-general way, contrasting domain-specific models (e.g., Pulman 2009). Moreover, while the assessment cycle draws from and is inspired by broader models of learning, it is distinct because it is narrowly focused on the learning that takes place in peer assessment. For instance, the experiential learning notions of reflecting on experiences to make abstractions (Kolb and Kolb 2012) are closely related to the idea that peer assessment helps students develop objective lenses for self-assessment. Because the assessment cycle narrows in such connections, it is well suited for deeper analyses of these specific types of activities. To provide a theoretical basis for the learning that takes place, Sadler's (1989) theory of self-assessment is used.

Figure 1. The Assessment Cycle

INSERT FIGURE 1

Peer assessment may include all six of the activities shown in Figure 1, or only a subset of them, and the order of activities may vary (e.g., feedback reception and provision might occur in opposite order). Nevertheless, it is useful to see these activities as a

cycle, because their combined use changes their demand characteristics (e.g., receiving feedback with no opportunity for revision limits the value of the feedback).

Task Engagement

Students generally begin by engaging with a task (e.g., solving a problem, writing an essay) similar to the one they will assess work from. This allows students to support their peer analyses with insights from working on the task itself. Open-ended tasks with multiple solutions support learning through analysis, because they give students more chances to compare, contrast, and connect their solutions (Schoenfeld 1991). Self-assessment can also be furthered by explicit prompts for explanation, which help students reflect on the quality of their work, developing greater *performance awareness*.

Explanation supports self-assessment by making gaps in one's knowledge more evident (Chi et al. 1994; Lombrozo 2006; Wong, Lawson, and Keeves 2002). For instance, Schoenfeld (1987) prompted his students to explain their thought processes in problem solving through metacognitive questions (e.g., Why are you doing that? How will it help you?). Over time, his students incorporated the questions into their regular practice, using them to guide their problem solving. This provided opportunities for students to work towards *gap closure* during problem solving, because they were better able to interrogate the quality of their problem solving process on an ongoing basis. In this way, students learned to self-assess their engagement processes, not just products, further developing their self-regulation skills (Panadero and Alonso-Tapia 2013).

Peer Analysis

Peer analysis involves any attempt to make judgments about the quality of a work, to assign a grade, to give constructive criticism, or to generate feedback. This

practical analytic experience helps individuals develop a sense of distanced objectivity that they can apply to their own work (Black, Harrison, and Lee 2003). Imagine that a student provides feedback to her peers on their writing, and she begins to notice common errors, such as: insufficient ‘sign posting’ in the introductions of their work and the use of unnecessarily long sentences. Because she is removed from the particular work she is analysing, she is better able to see such flaws in the writing. If these observations are made consistently, they become a lens that she can later apply to her own work. However, without the opportunity to see these flaws in her peers’ work, she may never have developed the appropriate lens for identifying them in her own work.

In analysing peer work, students are exposed to a variety of examples, which helps them see gradations in quality (Sadler 1989). In contrast, exposing students to model solutions alone may make it difficult for them to determine what makes the solutions good (limiting *goal awareness*). When students are able to compare different solutions to the same problem, it is easier to see the strengths and flaws in the solutions. Such experience, even with hypothetical work, can help students develop deeper conceptual understanding (Swan 2006).

Feedback Provision

Feedback provision involves students describing their analyses to their peers. By making students accountable for the learning of their peers, they are more likely to engage meaningfully with peer assessment (Featherstone et al. 2011). Students can provide written and verbal feedback. Written feedback may be provided efficiently from a number of students, but it also limits the personal connections students make. Verbal communication highlights the social aspects of the interaction, but also requires additional class time. Verbal discussions provide students with opportunities to practice

explaining their ideas in a way that other students can understand and allow students to receive immediate feedback if their ideas are understood or not (supporting *performance awareness*). Close peer interactions also push students to be constructive, focusing on how to improve work, and not just critique it (supporting *gap closure*).

Feedback Reception

When students receive feedback, it allows them to see their work from another's perspective. Over time, feedback helps individuals focus on the aspects of their work that seem to be problematic, forming an objective lens for self-assessment (and promoting *performance awareness*). Suppose a mathematics student consistently receives feedback that it is difficult to follow his calculations, because he skips steps and provides no explanations. Integrating these lenses into his repertoire of self-assessment, after solving a problem, the student may ask: did I sufficiently explain my calculations?

Not all feedback is equally useful (Hattie and Timperley 2007; Shute 2008). Feedback that helps students learn to independently analyse, critique, and improve their own work is the most beneficial (Hattie and Timperley 2007). Simply receiving feedback that tells the correct answer is less likely to be of value (Alevin et al. 2004). Even worse, praise focuses individuals on themselves rather than the task at hand, actually resulting in reduced performance (Mueller and Dweck 1998). Similarly, grades can distract students from using elaborated feedback (Butler 1988). Thus, students need guidance on what types of feedback they should provide to one another.

Peer Conferencing

Peer conferencing allows students to discuss their feedback and analyses. This can support *goal awareness*, *performance awareness*, and *gap closure* by enhancing the benefits of analysis, feedback provision, and feedback reception. Through conferences, students can explain their ideas verbally and also discuss problems more broadly. Because both students have spent time thinking about the problem before the conference, even a short discussion can be productive. Conferencing also allows students to provide more affective support, as compared to distanced feedback, which may be perceived as critical or insensitive (Patton 2012; Wilson, Diao, and Huang 2014).

Revision

Revision allows individuals to close the feedback cycle (Sadler 1989). When students receive feedback after an assignment is already completed, they have no opportunity to actually use the feedback to revise their work. This is detrimental, because students miss out on the learning involved in revision (related to *gap closure*). In contrast to instructor comments that accompany grades, peer assessment can often be conducted formatively, and as a result, students can revise their work before turning in a finished product. Moreover, when students know they will be expected to revise, it may influence the feedback they give and how they perceive feedback they receive.

Applying the Model

The assessment cycle describes how peer assessment supports self-assessment. Simply implementing all six activities does not guarantee learning; the way in which activities are implemented is equally important. Each activity and how it supports Sadler's (1989) three components of self-assessment are summarized in Table 1. This

model is now applied to three examples from the literature, each involving students in their first year of university studies.

Table 1. Key Aspects of Peer Assessment

Component	Examples of how it supports self-assessment
Task Engagement	<i>Performance awareness</i> : students explain their ideas <i>Gap closure</i> : revisions during engagement / problem solving
Peer Analysis	<i>Goal awareness</i> : experience analysing a variety of examples
Feedback Provision	<i>Performance awareness</i> : explaining ideas and receiving feedback on explanations <i>Gap closure</i> : developing constructive feedback to improve work, not just critique it
Feedback Reception	<i>Performance awareness</i> : students are able to view their own work from another's perspective
Peer Conferencing	Opportunities to discuss analyses and feedback can increase the impact of peer analysis, feedback provision, and feedback reception.
Revision	<i>Gap closure</i> : students use analyses and feedback to improve their work

Peer-Assisted Reflection

Method

Peer-Assisted Reflection (PAR; Reinholz, 2013) was developed during three iterations of a design-based research study (Cobb et al. 2003). The first iteration took place in a community college algebra classroom, while the final two iterations took place in introductory college calculus. The purpose of the first iteration was to develop tools whose impact could be tested in the final two iterations of the study. PAR was designed to support deeper conceptual understandings of calculus by having students: (1) work on a difficult problem, (2) self-reflect, (3) analyse a peer's work and exchange peer feedback, and (4) revise before turning in a final solution. Students completed these

activities each week with one special homework problem that was designated as a PAR problem. Students were pushed to work with different students each week to ensure that they were exposed to a variety of examples of work. During the final iteration of PAR, students also participated in a weekly training exercise in which they were given three sample solutions of varying levels of quality. Students silently analysed the solutions and then had a short whole class discussion in which they were exposed to the analyses of their peers and their instructor.

Each semester approximately 400 students register for the calculus course, spread across 10 sections that each have an average of 30-40 students enrolled; a few larger sections have 50-90 students. All sections are taught with a common curriculum and common examinations. During the second iteration of study, a cooperating instructor (Michelle) taught two sections of the course, and used PAR in only one of them so that a comparison could be made to account for teacher effects. During the third iteration of study, there was a single experimental section and the rest were considered comparison sections. A variety of data were collected, including: video observations, copies of student work and examinations, audio records of student conversations, student surveys, and interviews of students regarding their experiences with PAR (Reinholz, forthcoming). All examinations were randomized and graded blindly by the group of course instructors to ensure objectivity.

Results

This paper focuses on the quantitative results for the purpose of demonstrating PAR's impact; further analyses will be presented elsewhere (Reinholz, forthcoming). To account for non-random assignment of groups, students' examination 1 scores were used as a pre-test during both iterations of the study. In both cases, there were no

significant differences between students in the experimental sections and comparisons section. Additionally, prior achievement data (e.g., high school grade point average (GPA), American College Test (ACT) scores) were collected during the third iteration, and no significant differences were found between sections. These analyses supported the validity of the quasi-experimental comparison.

To measure the impact of PAR on examination scores, a two-level random effects hierarchical linear model (HLM) was created. The null model included class section and examination number as second-level variables, while the alternative model added the use of PAR as a fixed effect. A comparison of the two models, using the anova package in R, showed that PAR had a significant impact on examination scores $\chi^2(1) = 3.9635, p = 0.0465^*$. Michelle's comparison section (M = 66.84%, SD = 23.47) performed similarly to the rest of the comparison sections (M = 67.32%, SD = 19.37), while her experimental section performed better (M = 73.03%, SD = 18.37); thus, it is unlikely that improvements can be attributed to teacher effects. Moreover, the pass rates in the course (passing with an A, B, or C) were 13% higher in Michelle's experimental section than in the other sections.

The results of the second iteration were replicated and improved during the third iteration. Once again, HLM models were created that showed the impact of PAR was significant $\chi^2(1) = 8.6565, p = 0.00325^{**}$. During this iteration, the average examination scores in the experimental section (M = 75.2%, SD = 18.6) were much higher than in the comparison sections (M = 64.1%, SD = 21.1). This was also evident with pass rates being 23% higher in the experimental section than the other comparison sections.

Applying the Assessment Cycle

PAR was designed on the same principles as the assessment cycle, so it included all components of the cycle. Students developed *goal awareness* by regularly analysing the work of their peers and also by analysing three sample solutions through a weekly training exercise. This training was only present during the third iteration of the study, and likely accounts for some of the greater impacts of the intervention (e.g., 23% improvement in pass rates compared to 13%) during iteration three. Students developed *performance awareness* by reflecting on their own work each week and by receiving and incorporating peer feedback on their work; they were also prompted to explain their work in the tasks and through peer conferences. Finally, students had opportunities to practice *gap closure* as they revised their work each week. In a random sample of 122 PAR assignments, only three assignments did not include revisions. PAR was aligned well with the theoretical learning model suggested by the assessment cycle, which likely describes why student learning was supported so effectively by PAR.

Calibrated Peer Review

Description of the Study

Calibrated Peer Review (CPR) is an online tool designed to improve students' reading and writing in science (Robinson 2001). The CPR process involves students: (1) writing and submitting an essay, (2) assessing three calibration essays using a rubric, and (3) using the same rubric to assign grades to three anonymous peer essays (University of California 2012). Student assessments of the calibration essays are compared to a set of standards defined by the instructor, to determine how well the students are calibrated with the assessment criteria. If students are poorly calibrated, their scores for peer essays are weighted less heavily than the scores assigned by other students. Students are

able to see the reviews submitted by the two other reviewers of each essay they review, to help them develop a better sense of the quality of their own assessments.

The impact of CPR on student performance was investigated in an introductory physiology course, with 40 students enrolled (Pelaez 2002). Each topic in the course was assigned to either the experimental or traditional condition, so comparisons were made by topic rather than by students. To assess the impact of the teaching methods, student performance on examination problems taught using the experimental and traditional approaches was compared. The researcher attempted to match topics that were taught using the experimental and traditional approaches to make an even comparison.

The experimental topics were taught using resource materials such as research abstracts and textbooks assignments combined with CPR. Students wrote an essay about each of these topics, and engaged in the CPR process to assess their classmates' essays. Afterwards, students discussed the topics as a class. The control condition involved didactic lectures followed by groupwork on textbook problems.

Results

Final examination scores on experimental questions were significantly higher than traditional questions, for multiple choice (76.9% vs. 65.1) and essay (81.2% vs. 76.9%) questions. However, it is difficult to assess the impact of peer assessment, because when students engaged with CPR they also had to generate essays, unlike the traditional lecture format, and topics were taught using resource materials. Nevertheless, the study does provide some evidence that, combined with other aspects of active learning, CPR helped improve conceptual understanding. Another limitation was the assignment of topics to conditions; there was no randomization, and because the

researchers constructed the examination questions, there may be some bias in the examination procedures.

Applying the Assessment Cycle

Students engaged in all components of the assessment cycle except for peer conferencing and revision. Peer conferencing was not possible because CPR is anonymous and revision was not possible because CPR was used summatively to assign grades. Through CPR, students were exposed to three calibration essays for each assignment and three peer essays. This practical experience analysing a large variety of examples likely helped students develop *goal awareness*. Students also applied the same rubric to their own work, which helped them develop *performance awareness*. However, students were not able to revise, limiting opportunities for *gap closure*. Also, while students were asked to justify their scores, they were not pressed to generate further feedback to support their peers. As a result, the quality of student feedback was deeply dependent on the quality of the assessment rubrics. In sum, the activities in this study had some alignment with the assessment cycle, with some areas of discrepancy. This may help account for the fact that some learning gains were witnessed, but not as much as the researchers had hoped for.

Development of Understanding of Assessment for Learning (DUAL)

Description of the Study

The DUAL program was designed to help first-year biology students transition to the university by helping them develop a better understanding of assessment criteria (Yucel et al. 2014). Through exemplar marking and peer review, students were intended to better understand the expectations for report writing and also to give and receive useful

feedback to improve the quality of their reports. The exemplar marking activity involved students scoring two exemplar laboratory reports using a standard rubric for scoring first-year biology reports. After marking the reports, students had an opportunity to discuss their assigned marks in groups. For the peer review exercise, students brought a draft of their first report to class a week before it was due. Students were asked to mark another report anonymously using the same rubric as from the marking exercise. Students were encouraged to seek clarification on the scores from their demonstrators, but did not speak with peers directly.

Two cohorts of students studying Animal Evolution and Diversity, from 2009 and 2010, were involved in the study; there were approximately 400 students in each cohort. Australian Tertiary Admissions Rank (ATAR) scores were compared between cohorts to confirm that they were comparable groups. Student scores on lab reports were compared between the 2010 cohort (who participated in DUAL) and the 2009 cohort (who did not participate in DUAL). Also, a survey of the 2011 cohort was conducted to better understand student perceptions of DUAL.

Results

Students in the DUAL program achieved significantly lower scores on their laboratory reports than the cohort that did not participate in DUAL. However, the differences were small, with the comparison cohort scoring 4% higher on the first report and only 2% higher on the second report. Despite lack of improvement, the 96% of students surveyed in 2011 agreed that the exemplar marking and discussion exercises helped them better understand the assessment criteria. In contrast, only 65% of students agreed that the peer review exercise helped them improve their reports before

submission. None of the students commented about the benefits of *providing* feedback, not just receiving it.

Applying the Model

Yucel et al. (2014) offer some possible explanations for the lack of positive impact from DUAL, such as differences in cohorts or the demonstrators achieving greater understanding of the assessment criteria for the second cohort; while these explanations may be true, the assessment cycle supports an alternative interpretation. Students in the program engaged with all components of the assessment cycle other than peer conferencing, but with limited frequency. Although students scored two model reports, they only provided peer feedback once during the semester; this limited their exposure to the work of others and likely limited their development of *goal awareness*. Also, students only received feedback once, and did not practice applying the assessment criteria to their own work on a regular basis, so they had few opportunities to develop *performance awareness*. Although students had a chance to revise, it only took place once, so there were also limited opportunities to practice *gap closure*.

The above analyses indicate that one of the primary limitations of the DUAL program may be that students did not have sufficient opportunities to participate in the assessment activities to develop expertise. Another possible limitation is the assessment rubrics. Although the researchers did not provide the actual rubrics, given that the rubrics were used to assign grades, it may have focused students more on grades and less on supportive feedback, limiting the impact of their assessments (Butler 1988). Finally, some students reported a negative experience or spoke poorly of their peers (e.g., “The peer who corrected mine was an idiot...”; cf. Yucel et al., 2014, 12), which may have been mitigated if students had conferenced with their peers and developed

personal connections. In this case, the assessment cycle highlights a number of areas where the learning activities could have been improved, which would have likely resulted in greater student learning.

Instructional Implications

Although the assessment cycle focuses on self-assessment, peer assessment can help learners develop a variety of skills, including: collaboration, communication, conceptual understanding, and problem solving skills (Falchikov 2005). The assessment cycle provides some insight into how such skills may be developed. For instance, activities that do not include peer conferencing are less likely to help students develop collaboration skills because they provide fewer opportunities for student interactions. Similarly, providing feedback can help improve students' communication more than only analysing work, because students practice communicating their insights to other students. In this way, the desired learning outcomes may influence which aspects of the assessment cycle are emphasized most in instruction.

While the specifics of particular instructional strategies are beyond the scope of this paper, one promising area of research is noted: the use of rubrics. Rubrics can help students develop a better sense of assessment criteria (goal awareness), as long as students are actively involved in applying the rubrics (i.e. rubrics cannot simply be handed out; cf. Andrade and Valtcheva 2009). Moreover, allowing students to participate in the creation of rubrics can help increase their learning impact, improving understanding, increasing self-regulation, reducing avoidance goals, and increasing the accuracy of self-assessments (Panadero and Romero 2014).

Summary and Conclusion

The assessment cycle provides a theoretical framework for understanding the learning that takes place through peer assessment by connecting peer and self-assessment. There are six components to the assessment cycle, each describing a different aspect of peer assessment with different learning potential. A given activity structure may include all of these components or only a subset of them. Analysing such activities both in terms of which components of the assessment cycle they include and how the components are implemented provides the basis for understanding how the learning activity may support self-assessment. The potential learning benefits of such activities are grounded in Sadler's (1989) three criteria for self-assessment: *goal awareness*, *performance awareness*, and *gap closure*.

The assessment cycle was applied to three examples of peer assessment activities in undergraduate education: PAR, CPR, and DUAL. Students who engaged with PAR in introductory calculus (Reinholz, forthcoming) significantly improved their course performance, improving success rates by 23% during the third iteration. The use of CPR in physiology was also shown to improve student performance, with multiple choice problem scores on the final examination 11% higher for questions that were based on topics taught with CPR. Benefits for essay questions were lower (4% improvement); this may be due to some of the limitations to the activity structure, such as no opportunities to discuss feedback or revise essays. As a result, students were unable to close the feedback cycle, limiting the impact of the feedback they received (Sadler 1989). The final example, the DUAL Programme, did not show any improvement in student understanding. This was most likely because students had limited opportunities to engage in peer assessment, and thus did not develop the required goal awareness, performance awareness, or ability for gap closure.

In each of these examples, the actual learning benefits were consistent with what the assessment cycle would predict; the more the activity structure supported goal awareness, performance awareness, and gap closure, the more that students learned. Nevertheless, these results should be interpreted with caution, because the model was only applied to a limited number of cases, and was done so only *after* the studies were already conducted. A next step for research is to use the cycle to design new activity structures and use it as a guide for future research studies, to test and validate the model.

References

- Aleven, V., A. Ogan, O. Popescu, C. Torrey, and K. Koedinger. 2004. "Evaluating the Effectiveness of a Tutorial Dialogue System for Self-Explanation." In *Intelligent Tutoring Systems*, 29–33.
- Andrade, H. L. 2010. "Students as the Definitive Source of Formative Assessment: Academic Self-Assessment and the Self-Regulation of Learning." In *NERA Conference Proceedings 2010*. Paper 25.
- Andrade, H. L., and A. Valtcheva. 2009. "Promoting Learning and Achievement through Self-Assessment." *Theory Into Practice* 48 (1): 8.
- Black, P., C. Harrison, and C. Lee. 2003. *Assessment for Learning: Putting It into Practice*. Berkshire, England: Open University Press.
- Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education* 5 (1): 7–74.
- . 2009. "Developing the Theory of Formative Assessment." *Educational Assessment, Evaluation and Accountability* 21 (1): 5–31.
- Boekaerts, M., and L. Corno. 2005. "Self-Regulation in the Classroom: A Perspective on Assessment and Intervention." *Applied Psychology* 54 (2): 199–231.
- Butler, D. L., and P. H. Winne. 1995. "Feedback and Self-Regulated Learning: A Theoretical Synthesis." *Review of Educational Research* 65 (3): 245–81.
- Butler, R. 1988. "Enhancing and Undermining Intrinsic Motivation: The Effects of Task-Involving and Ego-Involving Evaluation on Interest and Performance." *British Journal of Educational Psychology* 58 (1): 1–14.
- Chi, M. T. H., N. De Leeuw, M. H. Chiu, and C. LaVancher. 1994. "Eliciting Self-Explanations Improves Understanding." *Cognitive Science* 18 (3): 439–77. doi:10.1016/0364-0213(94)90016-7.
- Cobb, P., J. Confrey, A. Disessa, R. Lehrer, and L. Schauble. 2003. "Design Experiments in Educational Research." *Educational Researcher* 32 (1): 9–13. doi:10.3102/0013189X032001009.
- 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): 331–50. doi:10.1080/03075079912331379935.
- Dunlosky, J., and A. Lipko. 2007. "Metacomprehension: A Brief History and How to Improve Its Accuracy." *Current Directions In Psychological Science* 16 (4): 228–32.

- Dunning, D., C. Heath, and J. M. Suls. 2004. "Flawed Self-Assessment." *Psychological Science in the Public Interest* 5 (3): 69.
- Falchikov, N. 2005. *Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education*. New York, NY: Routledge.
- 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.
- Featherstone, H., S. Crespo, L. M. Jilk, J. A. Oslund, A. N. Parks, and M. B. Wood. 2011. *Smarter Together! Collaboration and Equity in the Elementary Math Classroom*. Reston, VA: National Council of Teachers of Mathematics.
- Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.
- Kolb, A. Y., and D. A. Kolb. 2012. "Experiential Learning Theory." In *Encyclopedia of the Sciences of Learning*, 1215–19. Springer.
- Kollar, I., and F. Fischer. 2010. "Peer Assessment as Collaborative Learning: A Cognitive Perspective." *Learning and Instruction, Unravelling Peer Assessment*, 20 (4): 344–48.
- Lombrozo, T. 2006. "The Structure and Function of Explanations." *Trends in Cognitive Sciences* 10 (10): 464–70. doi:doi:10.1016/j.tics.2006.08.004.
- Min, H. 2006. "The Effects of Trained Peer Review on EFL Students' Revision Types and Writing Quality." *Journal of Second Language Writing* 15 (2): 118–41.
- Mueller, C. M., and C. S. Dweck. 1998. "Praise for Intelligence Can Undermine Children's Motivation and Performance." *Journal of Personality and Social Psychology* 75 (1): 33–52.
- Nicol, D. J., and D. Macfarlane-Dick. 2006. "Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice." *Studies in Higher Education* 31 (2): 199–218.
- Panadero, E., and J. Alonso-Tapia. 2013. "Self-Assessment: Theoretical and Practical Connotations. When It Happens, How Is It Acquired and What to Do to Develop It in Our Students." *Electronic Journal of Research in Educational Psychology* 11 (2): 551–76.
- Panadero, E., and M. Romero. 2014. "To Rubric or Not to Rubric? The Effects of Self-Assessment on Self-Regulation, Performance and Self-Efficacy." *Assessment in Education: Principles, Policy & Practice* 21 (2): 133–48. doi:10.1080/0969594X.2013.877872.
- Patton, C. 2012. "'Some Kind of Weird, Evil Experiment': Student Perceptions of Peer Assessment." *Assessment & Evaluation in Higher Education* 37 (6): 719–31.
- Pelaez, N. J. 2002. "Problem-Based Writing with Peer Review Improves Academic Performance in Physiology." *Advances in Physiology Education* 26 (3): 174–84.
- Pulman, M. 2009. "Seeing Yourself as Others See You: Developing Personal Attributes in the Group Rehearsal." *British Journal of Music Education* 26 (02): 117–35.
- Reinholz, D. L. 2013. *PAR for the Course: Developing Mathematical Authority*. Paper presented at the annual conference on Research in Undergraduate Mathematics Education, Denver, CO, USA, February 21-23.
- . Forthcoming. "Peer-Assisted Reflection: A Design-Based Intervention for Improving Success in Calculus." Manuscript under review.

- Robinson, R. 2001. "Calibrated Peer Review™." *The American Biology Teacher* 63 (7): 474–80.
- Roscoe, R. D., and M. T. H. Chi. 2007. "Understanding Tutor Learning: Knowledge-Building and Knowledge-Telling in Peer Tutors' Explanations and Questions." *Review of Educational Research* 77 (4): 534–74.
- Sadler, D. R. 1989. "Formative Assessment and the Design of Instructional Systems." *Instructional Science* 18 (2): 119–44. doi:10.1007/BF00117714.
- Sadler, P. M., and E. Good. 2006. "The Impact of Self-and Peer-Grading on Student Learning." *Educational Assessment* 11 (1): 1–31.
- Schoenfeld, A. H. 1985. *Mathematical Problem Solving*. New York: Academy Press.
- . 1987. "What's All the Fuss about Metacognition?" In *Cognitive Science and Mathematics Education*, 189–215. Hillsdale, NJ: Lawrence Erlbaum Associates.
- . 1991. "What's All the Fuss about Problem Solving." *ZDM: The International Journal on Mathematics Education* 91 (1): 4–8.
- Shute, V. J. 2008. "Focus on Formative Feedback." *Review of Educational Research* 78 (1): 153–89. doi:10.3102/0034654307313795.
- Smith, H., A. Cooper, and L. Lancaster. 2002. "Improving the Quality of Undergraduate Peer Assessment: A Case for Student and Staff Development." *Innovations in Education and Teaching International* 39 (1): 71–81.
- Stefani, L. A. J. 1998. "Assessment in Partnership with Learners." *Assessment & Evaluation in Higher Education* 23 (4): 339–50.
- Swan, M. 2006. *Collaborative Learning in Mathematics: A Challenge to Our Beliefs and Practices*. London: National Institute for Advanced and Continuing Education (NIACE) for the National Research and Development Centre for Adult Literacy and Numeracy (NRDC).
- Topping, K. J. 1998. "Peer Assessment between Students in Colleges and Universities." *Review of Educational Research* 68 (3): 249.
- . 2003. "Self and Peer Assessment in School and University: Reliability, Validity and Utility." *Optimising New Modes of Assessment: In Search of Qualities and Standards*, 55–87.
- . 2009. "Peer Assessment." *Theory Into Practice* 48 (1): 20–27.
- University of California. 2012. "Calibrated Peer Review: Web-Based Writing and Peer Review." <http://cpr.molsci.ucla.edu/Overview.aspx>.
- Wilson, M. J., M. M. Diao, and L. Huang. 2014. "'I'm Not Here to Learn How to Mark Someone Else's Stuff': An Investigation of an Online Peer-to-Peer Review Workshop Tool." *Assessment & Evaluation in Higher Education*.
- Wong, R. M. F., M. J. Lawson, and J. Keeves. 2002. "The Effects of Self-Explanation Training on Students' Problem Solving in High-School Mathematics." *Learning and Instruction* 12 (2): 233–62.
- Yucel, R., F. L. Bird, J. Young, and T. Blanksby. 2014. "The Road to Self-Assessment: Exemplar Marking before Peer Review Develops First-Year Students' Capacity to Judge the Quality of a Scientific Report." *Assessment & Evaluation in Higher Education*, 1–16.
- Zimmerman, B. J. 2002. "Becoming a Self-Regulated Learner: An Overview." *Theory into Practice* 41 (2): 64–70. doi:10.1207/s15430421tip4102_2.