A DEFINITION FOR EFFECTIVE ASSESSMENT AND IMPLICATIONS ON COMPUTER-AIDED ASSESSMENT PRACTICE

Stephen Broughton, Paul Hernandez-Martinez and Carol L. Robinson

Mathematics Education Centre, Loughborough University

For a decade, computer-aided assessment (CAA) has been used extensively with first-year mathematics and engineering undergraduates studying mathematics modules at the institution under investigation. This project sought to evaluate the effectiveness of CAA. Using assessment literature and activity theory to frame the study, this paper explores the aims of assessment and what it means for assessment to be "effective": it proposes a definition for effective assessment and discusses whether CAA can be considered effective assessment by this definition.

BACKGROUND

This project seeks to evaluate the use of a computer-aided assessment (CAA) system at a higher education establishment in the United Kingdom. The CAA system is used to test mathematics learning in first year mathematics and engineering mathematics modules. It asks questions that are mainly procedural in nature, in multiple choice and numerical input forms. Although lecturers employ the facilities that CAA offers in different ways (Robinson et al., 2012), most students have the opportunity to practice similar questions to the ones they receive in the summative test.

Performing this evaluation required a comparison of the CAA system against an accepted standard for assessment; however, the trend for evaluations of CAA conducted hitherto has been in the form of self-report commentaries of practice. These evaluations lack an objective standard upon which to compare; hence, no precedent has been made and there are calls in the literature for a rigorous review (for example, Bull & McKenna, 2001 and Sangwin, 2003).

Formative assessment appeared to offer such a standard: it has been widely discussed in the literature; it has been argued to be effective (Black & William 1998, for example); and proponents of CAA suggested that it could be adopted with formative intentions (Bull & McKenna, 2001). However, there remain two concerns with using formative assessment as a standard for evaluation.

First, the definition of "formative assessment" is disputed and the term is used inconsistently. For example, there is disagreement whether it is necessary for students to demonstrate improvement as a consequence of the feedback they receive from formative assessment. Sadler (1989 pp.120-121) believed that improvement is necessary, citing Ramaprasad's (1983 p.4) distinction between information about performance and feedback, which requires the student to act upon information about

2013. In xxxx (Eds.). *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. XXX-YYY. Kiel, Germany: PME.

performance. This distinction is not maintained in the most recent definition of formative assessment (Black & Wiliam, 2009).

Black and Wiliam (2009 p.10) believed this requirement was too stringent: determining whether an assessment is formative would require establishing that the assessment has caused an improvement that would not have occurred otherwise. Furthermore, they added, "even the best designed interventions will not always result in better learning for all students" (Black & Wiliam, 2009 p.10, emphasis in original), making it difficult to confidently declare an assessment "formative".

Determining whether practices are formative assessment or not depends on the definition of formative assessment used.

Second, there are gaps in the theoretical underpinning of formative assessment. After Black and Wiliam (2009) developed the theory of formative assessment further, using cultural-historical activity theory for some aspects, Bennett (2011) and Taras (2010) believed that there was scope to develop the theoretical framework of formative assessment still further.

Ultimately, we decided that formative assessment did not provide the standard that we wished to evaluate CAA with. Instead, we would develop a definition for effective assessment that describes the process of assessment, the criteria for success, the roles of the actors in the activity of effective assessment and the outcomes.

THEORETICAL FRAMEWORK

Cultural-historical activity theory (CHAT) offers three ideas in particular that would be necessary in a definition for effective assessment that were not developed fully for formative assessment.

First, there is the role of the community on activity. In CHAT, the community provides resources, shares the activity and imposes rules. For a student, teachers and peers have an impact on the activity of learning; there are rules of the classroom and there are shared responsibilities. For formative assessment, Taras (2010 p.3017) noted, "It is never quite clear who is involved in … the assessment cycle". Black and Wiliam (2009 p.9) suggested they used a framework derived from CHAT to develop themes for their theory of formative assessment; however, while Black and Wiliam discussed the role of the teacher in setting assessment and regulating learning, there remains a desire to explore the roles of peers and individual learners in more detail.

Second, activity theory maintains that all human activity is purposeful — "the expression 'objectless activity' is devoid of any meaning" (Leont'ev, 1978). Therefore, the way in which a student interacts in a learning activity is shaped by his/her own learning goals. When discussing the notion of object-oriented activity in expansive learning, Engeström and Sannino (2010 p.4) pointed out that "motives cannot be taught, they can only be nurtured" over time with students. That is, teachers and peers can influence or dictate students' goals.

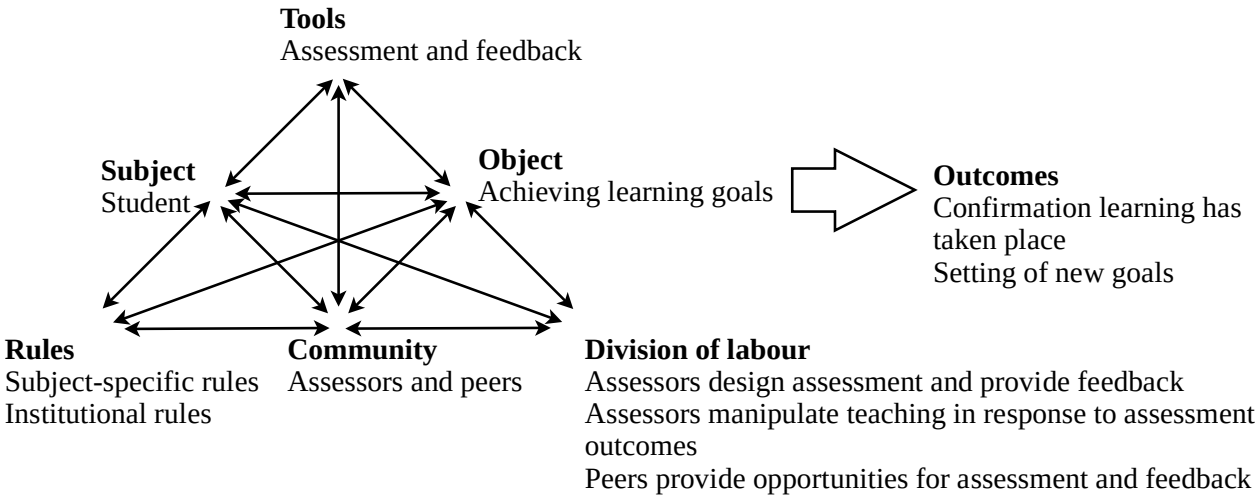

Figure 1: An Engeström (2000) representation of the activity of effective assessment

Black and Wiliam (2009 p.22) believed teachers must have learning intentions and engineer opportunities for learning. There are two problems with this construction.

First, the teacher's learning intentions and the student's learning goals may not always cohere; the formative assessment literature does not consider the outcomes of such a contradiction. Second, many teachers have little, if any, influence on self-assessment, peer assessment or peer support that happens outside the classroom, in which students may set their own learning goals. The outcomes of these assessments are not discussed in the formative assessment literature.

The third CHAT construct that we consider missing from formative assessment arises from the model of expansive learning proposed by Engeström and Sannino (2010 pp.8-9): that learning occurs in cycles through the development of the object of learning activity. That is, the objects of future learning activities are evolved from previously achieved learning goals. This is particularly true for mathematics learning, where learning more advanced concepts quite often demands requisite knowledge from simpler concepts. It is this evolution that permits viewing learning as a cycle and assessment as a process within it: successful learning warrants the setting of new learning goals and assessing new learning.

We propose the following definition of "effective assessment" that incorporates the advice from assessment literature and the constructs of CHAT.

A DEFINITION FOR EFFECTIVE ASSESSMENT

We consider assessment and feedback as tools in the activity of learning, with learning goals forming the object of learning. The student is viewed as the subject of the learning activity (fig. 1): it is the student that performs the assessment and, most often, the aims of the assessment relate to the student's learning and development.

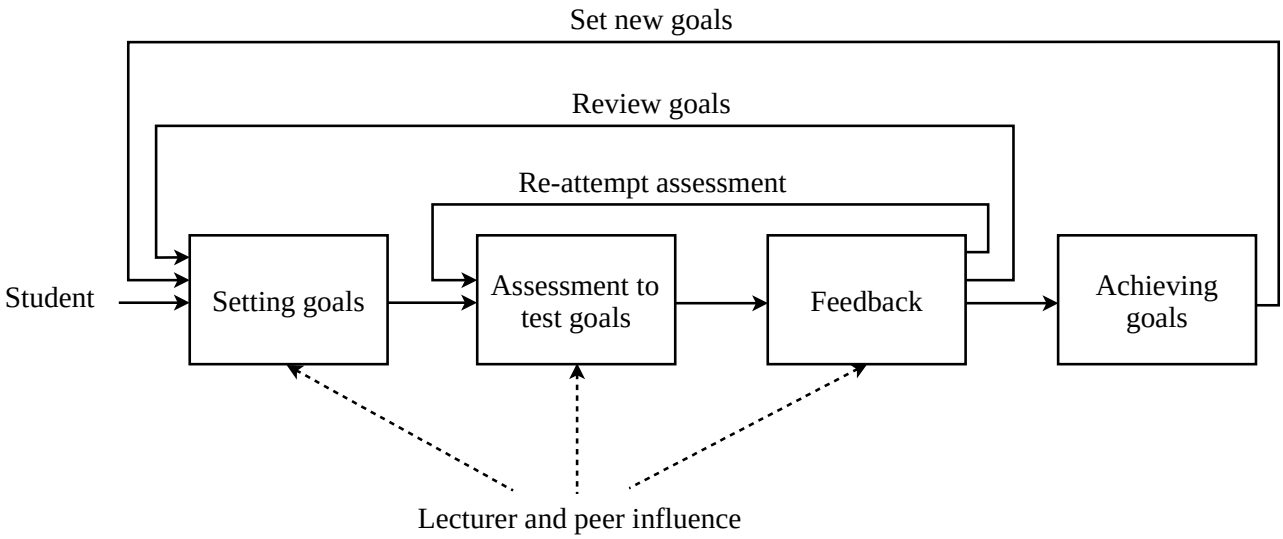

Figure 2: Effective assessment in a model learning cycle and the influence of lecturers and peers on the processes in this cycle

Effective assessment is, in part, defined as an assessment that enables the student to achieve his/her learning goals. However, the student does not act alone in this activity. Assessors and peers have roles as members of the learning community. They provide opportunities to assess learning and provide feedback. They also have an influence on the student when setting learning goals (fig. 2).

As Yorke (2003) suggested for formative assessment, effective assessment should be both a process and part of a cycle. In this cycle, the student possesses initial learning goals. He/she undergoes assessment to test whether those learning goals have been met and receives feedback. On receiving feedback, the student might re-attempt the assessment; or revise his/her learning goals; or, if the student has achieved his/her learning goals, he/she can set more challenging learning goals.

Lecturers and peers have an influence in setting goals, setting assessments and providing feedback. Initially, a novice student may be completely directed by the lecturer: goals may be set (explicitly or implicitly) on behalf of the student; and the assessment and feedback are managed entirely by the lecturer. In effective assessment, the student gains experience, knowledge and understanding so that he/she can take more responsibility for these stages of the learning process.

It could be argued the ultimate aim of learning is that the student "should be able to do unaided what previously needed knowledgeable support" (Yorke, 2003 p.496); an effective assessment should support students in developing self-regulation skills and in setting new personal learning goals autonomously. Therefore, for an assessment to be effective, the impact of lecturers and peers on the activity of learning is reduced.

We define "effective assessment" in the following way:

- An effective assessment must be a purposeful assessment with the aim to test whether the student has achieved his/her learning goals.

- An effective assessment must be part of a wider learning cycle in which the student sets more challenging learning goals with diminishing influence and input from lecturers and peers.
- An effective assessment must give opportunities for the student to receive feedback that is related to his/her performance in relation to his/her learning goals and opportunities for the student to demonstrate that he/she has developed sufficiently to achieve his/her learning goals.

Since each statement of the definition refers to the individual student, it is not possible to separate the effectiveness of an assessment from the individual. That is, if an assessment has been particularly effective for one student, one cannot conclude that it is effective for all.

With these criteria, we can describe the extent to which CAA is effective for each student. Evaluating the effectiveness of CAA with several students according to these criteria yields common strengths and weaknesses.

METHODOLOGY

The definition of effective assessment demands the study of individual students to evaluate effectiveness for his/her circumstances. We adopted a case study approach — limiting the study to one CAA system at one institution — to identify strengths and weaknesses particular to this system. Nine self-selecting first-year undergraduates from four disciplines (four mathematicians, three aeronautical and automotive engineers, one materials engineer and one sports technology engineer) and six lecturers teaching first-year mathematics modules (three in mathematics, two in materials engineering and one in mechanical engineering) were interviewed.

The interview questions related to how students and lecturers use CAA, the influence of peers and lecturers on CAA, and how CAA has helped both students and lecturers. A professional transcriptionist transcribed the audio files and the first author coded the participants' responses according to parts of the definition for effective assessment and other parts of the learning process (fig. 2).

ANALYSIS

The students

In terms of setting learning goals for themselves, many of the students set seemingly superficial short terms goals that related to their long-term aims of career success. Consequently, many of the students' goals when using CAA were expressed in terms of a percentage of marks.

Many of those students believed that achieving high marks in CAA was a demonstration that they had developed the required knowledge and understanding that was expected of them. These students set high percentages as their goals for CAA, with some not happy unless they achieved 100%. To that end, the practice test facility was used extensively to fully prepare for the summative test. They were confident that

CAA helped them to improve in this respect and the feedback was detailed and appropriate enough.

Other students felt that CAA did not test their conceptual knowledge as well as other assessments had. It seems for these students, CAA related poorly to the perceived implicit learning goals. One such student was interested in pursuing an academic career in mathematics research; however, her learning goals were also expressed in terms of marks and did not lead to more challenging learning goals beyond the CAA summative test. Even for these students, there is a dependence on the lecturer to provide implicit learning goals.

These students had developed procedures for approaching and completing the CAA tests; largely comprising a regime of practice test attempts and learning the method offered in the feedback. Although this aided the students in achieving their goals, it appears that it did not inspire them towards further learning.

Although the students were satisfied that they had achieved their learning goals — insomuch that they had achieved the marks they had set for themselves — it would appear that they felt the primary benefit of doing so was the accumulation of marks that contributed to their module grade. Some went further: they wished to accumulate the "easy marks" that CAA offered to lessen the burden of the exam for passing the module. From these comments, CAA has not been effective for these students in terms of encouraging further learning with more challenging learning goals.

Peers have an important role in CAA, with many of the students reporting that they had engaged in collaboration at some point during their first year for CAA. Most of the students had clear, though not always correct, views on when collaboration becomes plagiarism; other students had less clear views and were prepared to engage in practices that could be interpreted as plagiarism. For example, one student believed that helping others in a summative test was satisfactory, since the purpose of the first year is to ensure that all students have developed a common foundation of understanding for subsequent study.

Over the course of the first year, the students became more willing to engage with their peers by offering mutual support. The culture of these student cohorts appears to have had a role in this, since one student commented that she was aware that others on her course were collaborating and later did the same. Another student expressed an initial reservation to collaborate on CAA, but he had recently started to collaborate with others prior to the interview. While the students found such support helpful, it appears to come at the expense of developing self-regulation skills, since the students became less likely to attempt the assessment by themselves first.

Few students referred to the influence of lecturers, particularly in terms of support during assessment. It would appear that they were content with the assessment and feedback they were offered and they possessed little desire to self-assess beyond the compulsory assessment content. It could be argued that this is a culture that they have

become accustomed to: that studying beyond the provided content for which credit can be received has insufficient value.

The lecturers

Although most of the lecturers would have liked CAA to test students' knowledge and understanding more deeply than it currently does, the primary reason for using CAA is the need for regular assessment — for both formative and summative purposes — and CAA provided a means to offer regular assessment to large cohorts without a significant marking burden. Other forms of assessment addressed the need to test students' conceptual knowledge.

The lecturers believed that the students would only be sufficiently motivated to engage with CAA if they were offered module credit. However, they were concerned about the impact of collaboration and the impossibility of eliminating the plagiarism that arises from group-work, since, in some cases, it is not possible to invigilate the entire cohort in one computer laboratory. Hence, lecturers typically offered between 2.5% and 5% of module credit for each CAA summative test.

Some lecturers reported that many students develop a mechanical approach to answering CAA questions. As a result, some students have become quite adept at performing a mathematical procedure without having the flexibility to adapt to different contexts. Often, this problem is not identified until the final exam.

DISCUSSION AND CONCLUSIONS

In terms of satisfying learning goals, CAA is effective to a point: the students were content that the feedback enabled them to demonstrate an improvement; the lecturers are content that students have the opportunity to practise what they have learned in lectures and receive immediate feedback.

A concerning issue is that these students perceived high marks in CAA to be an indication of satisfying implicit learning goals set by the lecturer. The lecturers indicated that CAA might not always be an appropriate test of the knowledge and understanding that they wish to test of students.

The students did not express their aims in terms of the learning that is required; their learning goals were stated in terms of percentages and they often achieved those goals. On the one hand, achieving these learning goals gave the students confidence and reassurance that they had learned the material. On the other hand, students had set superficial goals that did not indicate what had been learned, and perhaps this explains in part why more challenging goals were not set.

Our earlier work with a similar cohort revealed similar findings: students face a contradiction between pursuing more challenging learning goals and concentrating efforts on pursuing marks (Broughton et al., 2011). The culture and history to which these students belong weighs in favour of pursuing marks: past and future examinations are perceived to determine success; and since no marks are offered for

learning beyond summative assessment material, it is perceived to have little value. Hence the learning cycle is broken.

The implication is that CAA is effective for low-level goals, where the depth or breadth of understanding is not important for the student. However, CAA did not inspire the students to continue the learning cycle and explore new learning goals, so there is a point where CAA is no longer effective. The challenge for CAA is to expose students to the value of pursuing further learning.

References

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25.

Black, P., & Wiliam, D. (1998). Inside the Black Box. *Phi Delta Kappan*, *80*(2), 139–148.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Broughton, S., Hernandez-Martinez, P., & Robinson, C.L. (2011). Focus groups to ascertain the presence of formative feedback in CAA. In: C. Smith (Ed.), *Proceedings of the British Society for Research into Learning Mathematics, 31*(2). Leeds, UK: BSRLM.

Bull, J., & McKenna, C. (2001). *Blueprint for Computer-Assisted Assessment*. CAA Centre: Loughborough.

Engeström, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, *43*(7), 960-974.

Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review*, *5*(1), 1–24.

Leont'ev, A. N. (1978). *Activity, Consciousness, and Personality*. Prentice-Hall. Retrieved from http://www.marxists.org/archive/leontev/works/1978/index.htm

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, *28*(1), 4–13.

Robinson, C. L., Hernandez-Martinez, P., & Broughton, S. J. (2012). Mathematics lecturers' practice and perception of computer-aided assessment. In P. Iannone & A. Simpson (Eds.), *Mapping University Mathematics Assessment Practices* (pp. 105–117). University of East Anglia.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144.

Sangwin, C. J. (2003). New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. *International Journal of Mathematical Education in Science and Technology*, *34*(6).

Taras, M. (2010). Assessment for learning: assessing the theory and evidence. *Procedia - Social and Behavioral Sciences*, *2*(2), 3015–3022.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477–501.