Assessing classroom assessment techniques

Victoria Simpson-Beck

Active Learning in Higher Education 2011 12: 125
DOI: 10.1177/1469787411402482

Article

active learning in higher education

Active Learning in Higher Education
12(2) 125–132

DOI: 10.1177/1469787411402482
alh.sagepub.com
SAGE

Assessing classroom assessment techniques

Victoria Simpson-Beck
University of Wisconsin Oshkosh, Wisconsin, USA

Abstract
Classroom assessment techniques (CATs) are teaching strategies that provide formative assessments of student learning. It has been argued that the use of CATs enhances and improves student learning. Although the various types of CATs have been extensively documented and qualitatively studied, there appears to be little quantitative research assessing the effectiveness of these techniques in improving student learning. The purpose of this study was to empirically test the assertion that the Muddiest Point, a specific type of CAT, improves student learning. Study results indicated no significant difference, on average, between the control and experimental groups.

Keywords
CATs, classroom-assessment, formative assessment, Muddiest Point

Classroom assessment techniques (CATs) in context

Classroom assessment techniques include a wide range of activities that may be grouped into the categories of summative or formative. Summative assessment techniques (e.g. testing and student ratings of instruction) are evaluative, occur at the end of learning, and, for the most part, are used to determine the extent to which learning has been retained. Formative assessment is reflective, student-centered, and used as an ongoing process to improve learning. Formative assessment allows for the correction, clarification and adjustment, by both instructor and student, to information prior to summative assessment (Adams, 2004). Formative classroom assessment techniques are linked to cognitive learning theory (Steadman and Svinicki, 1998), which focuses on how information is processed. Since the 1980s, formative classroom assessment techniques have been used by thousands of university instructors, in various disciplines, in the US (Angelo, 1998; Angelo and Cross, 1993). Consequently, much has been written about formative classroom assessment techniques.

In the US, the creation and advancement of formative classroom assessment techniques may be linked to the works of Cross (1987), Cross and Angelo (1988) and Mosteller (1989), who began developing, implementing, and writing about these strategies in the late 1980s. Angelo and Cross (1993) refer to formative classroom assessment techniques as 'CATs', and their writings on CATs, particularly their book *Classroom Assessment Techniques: A Handbook for College Teachers*

Corresponding author:
Victoria Simpson-Beck, Criminal Justice Program, University of Wisconsin Oshkosh, 800 Algoma Boulevard, Oshkosh, Wisconsin 54901-8655, USA.
Email: beckv@uwosh.edu

(1993), have come to be viewed as canonical. The information provided, and the various CATs described, in Angelo and Cross (1993) have been widely cited and are promoted as best practices in teaching on the Center for Teaching websites of many universities and colleges (see e.g. Duke, Vanderbilt, Georgetown, Iowa, Central, Oregon, Texas, Massachusetts, Washington, and Honolulu websites) throughout the US.

According to Cross and Angelo (1993), CATs are a group of specific teaching strategies designed to provide formative assessments of student learning by engaging students in reflective evaluation of course material, and through the systematic collection of student reflections on learning. These student reflections provide the instructor with useful feedback on how much and how well students are learning, which helps to improve the quality of teaching and learning.

> Classroom assessment focuses the primary attention of teacher and students on observing and improving learning, rather than on observing and improving teaching. To improve learning, it may often be more effective to help students change their study habits or develop their metacognitive skills (skills in thinking about their own thinking and learning) than to change the instructor's teaching behavior. In the end, if they are to become independent lifelong learners, students must take full responsibility for their learning. To achieve that end, both teachers and students will need to make adjustments to improve learning. Classroom assessment can provide information to guide them in making those adjustments. (Cross and Angelo, 1993: 3)

The use of CATs assumes that students must receive ungraded feedback, early and often, that students need to assess the quality of their own learning, and that students can assist in improving course instruction. According to Angelo and Cross (1993), the basic steps in the process include:

1. Planning – thinking about what may be gained from implementing CATs and selecting a CAT that meets the needs of the course goals, provides the necessary feedback to improve learning, is consistent with the instructor's teaching style, and can be easily implemented in class.
2. Implementing – explaining the purpose of the activity to the students before conducting the CAT so that they understand the assignment, then collecting responses and immediately analyzing them.
3. Responding – telling the students what has been learned from the assessment and what difference, if any, that information will make.

Angelo and Cross (1993) provide a list of the various CATs, which differ in complexity and the time they take to prepare, administer, and analyze. CATS also vary in use, based on instructional needs; however, a few of the most frequently mentioned CATS in the literature include: *the Minute Paper* – a one-minute writing assignment asking students to write about what they thought was the most important point made in class on a given day; the one-sentence summary; *the Memory Matrix* – an assignment in which students fill in cells of a diagram for which the instructor has provided labels; and *the Muddiest Point* – a short writing assignment asking students to provide information on what they find least clear or most confusing about a particular lesson or topic.

Perhaps not surprisingly, the use of the term 'CATs' in the literature and on university websites appears to have become so much associated with the works of Angelo and Cross (1993) that CATs seem to have become synonymous with the term 'classroom assessment' and, at times, synonymous with the various assessment strategies contained in their book. Given the popularity and advocacy of this approach in academia, it is rather ironic, as has been observed, that there is a notable lack of experimentation and research on CATs (Adams, 2004).

What is missing from the literature on CATs?

Despite the decades of demonstrated popularity in academia, and the associated assertion of the effectiveness of CATs, a review of the literature reveals a dearth of research evidencing quantitative empirical support that CATs improve overall learning. Indeed, there has been much written on this topic that is anecdotal, and most of the research that has been conducted, thus far, is qualitative and/or exploratory. Research on CATS has focused on:

1 the conceptualization and application of CATs (Angelo, 1991a, 1991b; Cottell, 1991; Angelo and Cross, 1993; Cross and Steadman, 1996; Lieberman et al., 2001; Walker, 1991);
2 the documentation of the various types of CATs (Angelo and Cross, 1993);
3 surveys of student and/or faculty opinions assessing the effectiveness of CATs on student learning and/or teaching (Byon, 2005; Kelly, 1991, 2005; Rouseff-Baker and Holm, 2004; Schwarm and VanDeGrift , 2002; Soetaert, 1998; Steadman, 1998);
4 student surveys on how CATs improved student course satisfaction (i.e. affective stance) (Byon, 2005);
5 surveys of faculty attitudes toward the use of CATs in the classroom (Gaddert, 2003).

Only one study was located that actually provided an empirical quantitative analysis measuring the relationship between the use of CATs in the classroom and improvement in learning.

Cottell and Hardwood (1998) employed a quantitative quasi-experimental design testing the effectiveness of several different CATs: background knowledge probe, minute paper, feedback form, directed paraphrasing, pro-and-con grid, what did I learn from the exam, classroom assessment quality circle, group instructional feedback technique, and group-work evaluation form. Despite the use of various CATs, the study results indicated that CATs did not significantly improve student learning. However, Cottell and Hardwood (1998) may have biased their results by implementing course changes during the study, based on information obtained from the use of CATs, to both the treatment and control groups (i.e. contamination bias).

The purpose of this study was to fill a void in the CATs literature by empirically and quantitatively exploring the effectiveness of CATs, while controlling for contamination bias, which prior research has failed to do. In particular, the study was designed to test one of the oldest CATs – the Muddiest Point. The Muddiest Point is an assessment technique which was originally developed by Dr Fredrick Mosteller, a distinguished Harvard professor, while teaching an undergraduate statistics course (Mosteller, 1989). This particular CAT is perhaps one of the most frequently used classroom assessment techniques, as it is easily adaptable. The Muddiest Point was designed to help the instructor understand what students find to be the least clear or most confusing (i.e. the muddiest point) in a particular lesson, assignment, or topic (Cross and Angelo, 1993; Mosteller, 1989). Thus, the current study addressed the following research questions:

RQ1: Does the use of the Muddiest Point CAT improve chapter test scores?
RQ2: Does the use of the Muddiest Point CAT improve cumulative final examination scores?
RQ3: Does the use of the Muddiest Point CAT improve overall course grades?

Methods

To address the above-stated research questions, the study used a pre-/post-test, quasi-experimental design comparing student learning outcomes across two sections of a fall 2009 introductory undergraduate course at a mid-sized northern university in the United States. This convenience sample

of students included 64 students in a Monday, Wednesday, and Friday section of the course, designated the experimental group, and 62 students in the Tuesday and Thursday section of the course, designated the control group. The demographics for the two groups were almost identical. The majority of students in both sections were white (97% and 96%, respectively), male (67% and 62%, respectively), and traditional college-aged (100% and 99%, respectively).

As previously discussed, only one CAT – the Muddiest Point – was examined in this study. The reason for selecting only one classroom assessment technique was to prevent the potential for confounding results: it is possible that one technique may be effective, while another is not. Angelo and Cross (1993) note that, in order for CATs to be effective, they must match the course goals. Since one of the course goals was to build foundational knowledge for upper-level courses in the major, the Muddiest Point was an appropriate choice of assessment technique as it promotes declarative learning (memory learning) and allows for the ongoing assessment of recall and understanding (Angelo and Cross, 1993). Finally, this technique was also selected as it includes some of the most useful aspects of classroom assessment. It is flexible (i.e. can be used on a variety of topics), which is important for an introductory course covering a wide range of topics, it is easy to use and understand, does not require a great deal of class time to administer, and does not require a great deal of time to analyze.

Course requirements and grading were the same across the two sections of the course, except for the Muddiest Point assignment, which was given only to the experimental group. For both sections of the course, chapter quizzes measuring learning outcomes were given at the beginning of the week (Mondays for the M/W/F section and Tuesdays for the T/Th section), and the Muddiest Point assignment was given to the experimental group at the end of the week (Friday). The Muddiest Point task was assigned to the experimental group on Fridays for two reasons: (1) to increase participation in the assignment; and (2) to prevent one potential form of bias. Class attendance was higher on the last day of the week, for both groups, because there were graded in-class critical thinking assignments necessitating attendance to earn the points. Also, having the Muddiest Point assignment on Friday for the experimental group prevented the potential for this assignment to bias information presented to the control group. After the Muddiest Point assignment was completed in the experimental group on Fridays, the instructor did not meet with the control group until subsequent Tuesdays, at which time a new weekly topic/chapter was begun for the control group and that topic was unrelated to the previous Muddiest Point assignment.

On Fridays, at the beginning of class, students in the experimental group were asked to take out a sheet of paper and in a short (1–2 minutes) piece of reflective writing discuss their muddiest point for the week. The assignment was anonymous, to encourage participation, and, although this was an ungraded assignment, students were told they were required to participate. If students did not have a muddiest point for the week, they were asked to write something else. Telling students that they were required to participate and that they must write something, regardless of whether they had a muddiest point, may have resulted in prompting more students to write about a muddiest point. To further increase student participation, the instructor collected the anonymous papers by row. If there were more students in a given row than the number of papers being collected at the end of the row, the instructor mentioned this fact. Then, the instructor advised that row of students that papers were missing and their row would be given more time to complete the assignment while the instructor collected papers from other rows. This approach always resulted in increasing the number of students participating in the assignment (100% of students present in class), but did not necessarily increase the number of student comments focusing on a muddiest point (weekly average was approximately 70%). After the Muddiest Point assignment papers were collected, the instructor reviewed the comments while the students worked on another in-class assignment.

During the last 15 minutes of class, the instructor reviewed and discussed with students the most prevalent muddiest points, and then covered as many additional comments as possible, using a discussion format, until the end of the course period. Typically, there were two or three groupings of muddiest points, which made it possible to cover the majority of comments. In most cases, all muddiest points were covered. On some days, however, only the most prevalent were covered, as time allowed. To prevent students from potentially feeling excluded when class time prevented the instructor from addressing all muddiest points, students were told that the instructor would try to cover all of the muddiest points submitted, but would begin with the most prevalent ones and continue, as time allowed. Then, if a particular muddiest point was not covered during class time, the student was invited to speak with the instructor after class, by email, or during office hours.

To determine the effect of the Muddiest Point CAT (the independent variable) on student learning (the dependent variable), summative learning assessment measures were used in a quasi-experimental pre-/post-test design. The quasi-experimental pre-/post-test research design allows for a straightforward assessment of an intervention by detecting the difference between two points in time. Students in both sections were given a pre-test to measure prior knowledge of course content and a post-test to measure acquisition of learning. The pre-test and the post-test were the same instrument. Repeated-measures t-tests were used to test student acquisition of learning between pre-/post-test scores for the two groups.

Additionally, nine chapter quizzes were given throughout the course to also assess student learning. Independent-samples t-tests were used to test for statistically significant differences, when comparing the nine quizzes and final examination scores between the experimental group and control group.

Results

Since the research questions address differences in learning between the two groups, it was first necessary to determine whether learning took place in either group. In other words, did students in the experimental and control groups learn course content? The results of a paired-samples t-test found a significant difference between the pre-test and post-test scores, with T value 12.058, 118 degrees of freedom, and $p = 0.000$. Students in both groups tested statistically significantly higher, on average, on the post-tests compared with the pre-tests. The test results also indicated that the correlation between the pre-tests and post-tests was strong, positive and statistically significant ($p = 0.001$). Students who did well on the pre-test also did well on the post-test. The balance of this section will focus on the specific research questions.

RQ1: Does the use of the Muddiest Point CAT improve chapter test scores? Independent-samples t-tests were used to compare group means between the experimental and the control group. The test results indicated that there were no statistically significant differences, on average, in the nine chapter test scores between the two groups. Using the Muddiest Point CAT in one section of an introductory criminal justice course did not result in better chapter test scores, compared with the control group.

RQ2: Does the use of the Muddiest Point CAT improve cumulative final examination scores? Independent-samples t-tests were used to compare group means between the experimental and the control group. The test results indicated that there were no statistically significant differences, on average, in the final examination scores. Using the Muddiest Point CAT in one section of an introductory criminal justice course did not result in better cumulative final examination scores for the students in that section of the course, compared with the control group.

RQ3: Does the use of the Muddiest Point CAT improve overall course grades? Again, independent-samples t-tests were used to compare group means between the experimental and the control group.

The test results indicated that there were no statistically significant differences, on average, in final course grade between the two groups. Using the Muddiest Point CAT in one section of an introductory criminal justice course did not result in better course grades for students in the experimental, compared with the control group.

Conclusions

Decades of qualitative literature and anecdotal comments on CATs have asserted the effectiveness of CATs in improving student learning. Only one study, however, had tested that assertion, and the results did not support it (Cottell and Hardwood, 1998). Nonetheless, the findings in Cottell and Harwood (1998) were limited due to contamination bias and a few other methodological problems. The current study sought to expand the literature in this area. Despite the careful planning employed in this study to avoid contamination bias and other methodological problems, the results did not provide evidence that the use of CATs improved learning. Specifically, the Muddiest Point CAT did not result in the treatment group performing, on average, significantly better, academically, than the control group. Cottell and Harwood's (1998) findings combined with the findings in this study suggest that CATs may not play a significant role in improving learning and therefore may not be an essential factor in the process of learning. Of course, academics may, and no doubt will, continue to use CATs as a classroom learning strategy, but evidence is beginning to amass limiting the ability to assert that these strategies improve learning.

Despite the non-significant findings in this study, the results should be interpreted with caution for several reasons. First, this study examined only one type of CAT, the Muddiest Point, and future research may benefit by examining other types of CATs. Second, the sample in this study was a small convenience sample of students at one university, which may limit the generalizability of the findings. Future research should strive to conduct a larger study examining the use of CATs across campuses, courses, and faculty. Finally, it was not possible to randomly assign students to the control and experimental groups, which may have biased the internal validity of the results. Although the pre-/post-test design is, arguably, the strongest of the quasi-experimental designs, quasi-experimental designs do not control threats to internal validity as effectively as true experimental designs employing random assignment to groups. Nonetheless, the characteristics of the students in the two groups were very similar, and social researchers have indicated that 'it is generally less important that a group of experimental subjects be representative of some larger population than that experimental and control groups be similar to each other' (Maxfield and Babbie, 1998: 173).

Some may interpret the findings in this study to mean that teaching has limited bearing on learning. However, such an observation is over-reaching, as the Muddiest Point assignment specifically, and CATs in general, was only one small tool in the pedagogical toolbox for the course. Alternatively, some may interpret the results to mean that the control group (the group that did not have the Muddiest Point assignment) had to work harder to find answers. Given that the control group consisted of freshmen students in a large introductory course, I suspect that they did not work any harder than the experimental group. More plausibly, their 'muddiest points' were clarified by their verbally raising a question, rather than through a writing assignment. It has been my experience (albeit anecdotal) that in the larger classrooms there is a mixture of introverts and extroverts, and questions the introverts may be too shy to ask are more than likely raised by one of the extrovert students – yet all benefit when a question is asked during class. Perhaps the findings in this study point to problems within the literature, rather than problems with the methods in the study.

There does appear to be some confusion regarding the causal relationship between CATS and learning, in the literature, and it is unclear whether CATs are meant to have a direct effect on learning or a direct effect on teaching. For example, Angelo and Cross (1993) explicitly state:

CATs focus 'the primary attention of teacher and students on observing and improving learning, rather than on observing and improving teaching' (p. 3). However, they also state that effective classroom assessment is based on seven assumptions, one of which is: 'The quality of student learning is directly, although not exclusively, related to the quality of teaching. Therefore, one of the most promising ways to improve learning is to improve teaching' (p. 7). Thus, perhaps CATs improve teaching, rather than learning, in which case, it could be that the use of the Muddiest Point CAT did nothing to improve the teaching of a seasoned professor. Finally, the results in this study may also be an artifact of supporters of, and commentators on, CATs confusing student empowerment and feelings of happiness with actual learning outcomes. Therefore, conceivably, the results may indicate that if instructors create a course atmosphere which emboldens students to participate, ask clarifying questions, and be inquisitive, and if instructors provide ongoing formative feedback (in general or specific to a student) based on class interactions, then perhaps additional assignments, such as CATs, become redundant.

Although there have been decades of publications on CATs, we are only beginning to test the assertions contained in that literature. When academics, in particular, assert and promote a specific pedagogical approach as a panacea to improved learning, then it should be incumbent upon academics, in particular, to put that assertion to the test. Thus, there is an explicit need for more research in this area. Future researchers are encouraged to explore the assertions emanating from the literature on CATs and expand the empirical research in this area by continuing to quantitatively study the effect of CATs on learning, and perhaps on teaching, by incorporating larger sample sizes, obtaining samples from various campuses and disciplines, and studying some of the CATs that have not yet been examined.

The lack of support found for the use of CATs to improve learning does not necessarily mean that the practices have no value. Cottell and Harwood (1998) suggested that, rather than be disheartened by non-significant findings, perhaps we should focus more on the role CATs play in the process of learning. As suggested above, it is possible that CATs exert more of an influence on teaching practices (e.g. creating a positive learning environment) than on learning. It is also possible that CATs help students learn how to be better learners, by increasing their ability to, for example, identify where they are missing information or misunderstanding information. Regardless, CATs are a tool that can be used to provide a break from lecture, promote student self-assessment, and provide ungraded, anonymous, formative student feedback.

References

Adams P (2004) Classroom assessment and social welfare policy: Addressing challenges to teaching and learning. *Journal of Social Work Education* 40(1): 121–42.

Angelo T (1991a) Introduction and overview: From classroom assessment to classroom research. *New Directions for Teaching and Learning* 46: 7–15.

Angelo T (1991b) Ten easy pieces: Assessing higher learning in four dimensions. *New Directions for Teaching and Learning* 46: 17–31.

Angelo T (1998) Classroom assessment and research: An update on uses, approaches, and research findings. *New Directions for Teaching and Learning* 75: 1–2.

Angelo T and Cross KP (1998) *Classroom Assessment Techniques: A Handbook for College Teachers*. San Francisco, CA: Jossey-Bass.

Byon AS (2005) Classroom assessment tools and students' affective stances: KFL classroom settings. *Language and Education* 19(3): 173–93.

Cottell P (1991) Classroom research in accounting: Assessing for learning? *New Directions for Teaching and Learning* 46: 43–54.

Cottell P and Harwood E (1998) Do classroom assessment techniques (CATs) improve student learning? *New Directions for Teaching and Learning* 75: 37–46.

Cross KP (1987) The need for classroom research. *To Improve the Academy* 6: 37–45.
Cross KP and Angelo TA (1993) *Classroom Assessment Techniques: A Handbook for College Teachers*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.
Cross KP and Steadman MH (1996) *Classroom Research: Implementing the Scholarship of Teaching*. San Francisco, CA: Jossey-Bass.
Gaddert BK (2003) Improving graduate theological instruction: Using classroom assessment techniques to connect teaching and learning. *Teaching Theology and Religion* 6(1): 48–52.
Kelly DK (1991) *The Effects of Classroom Research by Part-Time Faculty Upon the Retention of Adult Learners*. Saratoga Springs, NY: National Center on Adult Learning, Empire State College, SUNY.
Kelly DK (2005) Do you know what your students are learning? (And do you care?): Emerging issues in the practice of university learning and teaching. All Ireland Society for Higher Education (AISHE), Dublin: http://www.aishe.org/readings/2005-1/toc.html [accessed 7 September 2009].
Lieberman D, Bowers N and Moore DR (2001) Use of electronic tools to enhance student evaluation feedback. *New Directions for Teaching and Learning* 87: 45–54.
Maxfield MG and Babbie E (1998) *Research Methods for Criminal Justice and Criminology*. Belmont, CA: Wadsworth.
Mosteller F (1989) The 'muddiest point in the lecture' as a feedback device. *Journal of the Harvard Danforth Center on Teaching and Learning* 3: 10–21.
Rouseff-Baker F and Holm A (2004) Engaging faculty and students in classroom assessment. *New Directions for Community Colleges* 126: 29–42.
Schwarm S and VanDeGrift T (2002) Using classroom assessment to detect students' misunderstanding and promote metacognitive thinking. In: *Proceedings of the International Conference of the Learning Sciences*: http://www.cs.washington.edu/research/edtech/publications/sv02-icls.pdf [accessed 1 June 2009].
Soetaert E (1998) Quality in the classroom: Classroom assessment techniques as TQM. *New Directions for Teaching and Learning* 75: 23–35.
Steadman M (1998) Using classroom assessment to change both teaching and learning. *New Directions for Teaching and Learning* 75: 23–35.
Steadman M and Svinicki M. (1998) CATs: A student's gateway to better learning. *New Directions for Teaching and Learning* 75: 13–20.
Walker CJ (1991) Classroom research in psychology: Assessment techniques to enhance teaching and learning. *New Directions for Teaching and Learning* 46: 67–78.

Biographical note

Victoria Simpson-Beck is an Associate Professor of Criminal Justice at University of Wisconsin Oshkosh. Her published works have appeared in the *Journal on Excellence in College Teaching*, *Journal of Criminal Justice*, *Juvenile and Family Court Journal*, *Journal of Psychiatry and Law*, *Journal of Violence and Victims*, the book on Youth Violence: *Delinquency Monsters and Myths* and the *International Handbook of Penology and Criminal Justice*. Her research interests include evidence based teaching and learning, juvenile diversion programs, analyses of social policies, and international comparative analysis on criminal punishments. *Address*: Criminal Justice Program, University of Wisconsin Oshkosh, 800 Algoma Boulevard, Oshkosh, Wisconsin 54901-8655, USA. [email: beckv@uwosh.edu]