

Moving Beyond Assessment to Improving Students' Critical Thinking Skills: A Model for Implementing Change

Ada Haynes¹, Elizabeth Lisic², Michele Goltz³, Barry Stein⁴ and Kevin Harris⁵

Abstract: This research examines how the use of the CAT (Critical thinking Assessment Test) and involvement in CAT-Apps (CAT Applications within the discipline) training can serve as an important part of a faculty development model that assists faculty in the assessment of students' critical thinking skills and in the development of these skills within their courses. Seventy-five percent of faculty participating in a CAT scoring workshop at their institution reported greater understanding of students' strengths and weaknesses in critical thinking and 45% reported that CAT scoring had changed their teaching practices and/or assessment. In addition, participants attending a training session on CAT-Apps reported a greater willingness to place more emphasis on critical thinking assessments and less on factual knowledge assessments in their courses as a result of participation in training.

Keywords: critical thinking, assessment, faculty development, teaching, learning

Research shows a consensus in the need for students to develop critical thinking skills. We argue that an important component of improving students' critical thinking skills ultimately involves making faculty aware of student weaknesses and changing how faculty assess student learning. Faculty must be convinced of the need to improve student skills, while simultaneously being provided the tools to help them improve these skills. The current study demonstrates that a faculty driven assessment tool that measures skills that faculty feel are important and lets them examine student work can begin the process for change at an institution. However, many faculty would also benefit from additional training to develop assessments in their own discipline that emphasize these critical thinking skills. This project explores how the NSF funded CAT (Critical thinking Assessment Test) and its new companion CAT-Apps (Application of the CAT critical thinking skills within the discipline) can help close the loop in assessment and improvement of students' critical thinking skills. The current study explores the impact that participation in the scoring of the CAT instrument can have on faculty at one institution and the extent to which it helps faculty understand student weaknesses in critical thinking and motivates faculty to change their instructional methods. The current study will also explore how additional training in developing CAT-Apps in their discipline can impact faculty willingness to change their instructional methods

¹ Department of Sociology, Tennessee Technological University, Box 5052, Cookeville TN 38505, ahaynes@tntech.edu

² Center for Assessment and Improvement of Learning, Tennessee Technological University, Box 5031, Cookeville TN 38505, elisic@tntech.edu

³ Institutional Effectiveness, Eastern Kentucky University, Coates CPO 10, Richmond, KY 40475, michele.goltz@eku.edu

⁴ Department of Counseling & Psychology, Center for Assessment and Improvement of Learning, Tennessee Technological University, Box 5031, Cookeville TN 38505, bstein@tntech.edu

⁵ Center for Assessment and Improvement of Learning, Tennessee Technological University, Box 5031, Cookeville TN 38505, kharris@tntech.edu

and the potential value of such training for institutional change. This study will explore faculty attitudes toward the CAT and how participating in CAT scoring and CAT-App training has served as an impetus toward course changes.

Importance of Critical Thinking

Various constituent groups in our society are in widespread agreement about the importance of critical thinking. For instance, the Higher Education Research Institute (HERI) found that over 99% of faculty across the United States felt that teaching critical thinking is “essential” or “very important” (DeAngelo et al., 2009). Similarly, Derek Bok, president emeritus of Harvard University, notes national studies have found that more than 90% of faculty members in the United States consider critical thinking *the* most important goal of an undergraduate education (Bok, 2006).

Employers also recognize the importance of critical thinking and problem solving skills. A recent survey by the American Association of Colleges and Universities (AACU) found that 75% of employers want colleges to place more emphasis on critical thinking, real world problem solving, communication, and creativity, and 93% of these employers felt that these skills were more important than major (AACU, 2013). Similarly, a recent national survey released by Northeastern University (2013) found that both the general public and employers thought colleges should focus more efforts on improving critical/creative thinking skills and communication.

The need for critical thinking skills is not limited to the United States. In a 2012 survey conducted by the American Management Association of international managers and executives, critical thinking skills were ranked as the most important skills related to growth (AMA, 2012). These same managers and executives say that the need for critical thinking skills is only going to increase.

Teaching and Assessing Critical Thinking

There are considerable resources available to faculty about high impact educational practices that involve students in active learning and that can contribute to gains in critical thinking (e.g., real world problem solving, involving students in original research, use of case studies, simulations, service learning, team based learning). Unfortunately, even when high impact practices are used there is frequently a disconnect between the skills faculty want to develop using these activities and the way students are assessed in those courses. Higher education courses have a pervasive tendency to emphasize the rote retention of factual information (Stein et al., 2009). The impact of using rote retention assessments across a college career can lead to undesirable outcomes and dissatisfaction with programs of study. Unsurprisingly, some educators have been very critical of higher education for doing little to improve students’ critical thinking skills (Arum et al., 2011; Bok, 2006; Pascarella et.al, 2011).

One reason for the disconnect between what faculty think is important and what appears to be emphasized in their courses is the lack of good assessments to measure the critical thinking skills that are deemed so important. Constructing a test to measure students’ rote retention of factual information is much easier than designing an assessment that evaluates critical thinking skills. This is true within courses and across programs of study in higher education. Unfortunately, when faculty use rote retention tests to assess student performance, they are inadvertently encouraging students to devote most of their time and energy to memorizing information. The

impact of these assessments across a college career can lead to undesirable outcomes and dissatisfaction with programs of study. Secondly, it is rare for valid and reliable nationally normed outcome assessments to be directly linked to classroom assessments and instruction (Wilson, 2009).

Assessments of critical thinking are important at the institution and department level. Valid and appropriate assessment tools are an essential element in any effort to evaluate the effectiveness of programs in higher education and to evaluate the potential benefits of new approaches that aim to improve student learning. Without appropriate assessment tools, gauging progress and determining if efforts to improve critical thinking have been successful is problematic (Vogler, 2002). Indeed, many institutions may be hesitant to invest institutional resources in improving critical thinking skills if they cannot assess the effects of expenditures and be accountable to governing boards and accrediting bodies. A report by the Spellings Commission (2006) noted a lack of quality assurance in higher education that stems from the failure to adequately assess student progress in the outcomes that are deemed most important. Many accrediting organizations (e.g., ABET, SACS-COC) are now demanding systematic plans to improve student learning and are requiring assessment evidence to substantiate progress. Yet, few faculty-driven, valid and authentic national outcome assessments of students' critical thinking that involves faculty in scoring are available.

Objective tests of factual information are also the primary type of assessment in higher education classrooms (Kvale, 2007). In terms of Bloom's taxonomy, Anderson and Sosniak (1994) reported that 60% of test questions were at the Knowledge level, 20% at the Comprehension level, and 15% at the Application level.

Assessment often becomes an afterthought to faculty due to their lack of training and time constraints (Gayton, 2007). Faculty have little or no training in developing classroom assessments that promote the development of critical thinking skills and most faculty have little or no training in developing effective course assessments (Hutchings, 2010; Petress, 2007), but the classroom assessments faculty use to assign grades ultimately determine how students will invest their efforts to learn (Gibbs and Simpson, 2004-5). If assessments emphasize the rote retention of factual information, students will spend much of their time memorizing information. Boud and Falchikov (2007) argue that how faculty assess is more important than how faculty teach on impacting student learning. Assessment not only prioritizes what is important but incentivizes students' studying.

Faculty Role in Teaching and Assessment

Faculty are a key to any successful initiative to improve student learning. Two distinct movements have developed to engage faculty to improve student learning, the assessment movement and the faculty teaching and learning movement. Many campuses have offices for both assessment and teaching and learning centers. We argue that the assessment of students' critical thinking skills using an "authentic" faculty driven assessment where faculty can see student responses and simultaneously providing faculty development support can motivate faculty to focus more on the improvement of students' critical thinking skills. Thus, an authentic, faculty driven and faculty scored assessment can serve as a link between the teaching and learning and assessment communities on college campuses. Effective faculty development is key to helping faculty identify student strengths and weakness, introducing new pedagogies, and implementing these changes in pedagogy.

Obtaining faculty involvement and support for assessment activities in higher education is difficult (Hutchings, 2010; Ikenberry, 2009; Kinzie, 2010; Kuh and Ikenberry, 2009). Faculty attitudes toward assessment are a major challenge for improving outcomes assessment in higher education (Kinzie, 2010; Kuh and Ikenberry, 2009). This study will explore faculty attitudes toward the CAT and how participating in CAT scoring and CAT-App training has served as an impetus toward course changes.

The focus and quality of teaching methods, specifically those related to practices involving active learning, are vital to the improvement of student learning (Estepp, Roberts and Carter, 2012). Faculty professional development programs are designed to equip faculty with the skills to implement these practices in their courses. Behar-Horenstein, Schneider-Mitchell, and Graff (2009) studied the effectiveness of faculty development in participants' ability to infuse critical thinking skills into their instructional practices. After participation in in-depth faculty development seminars, the researchers found that across all emerging themes participants' discussed a heightened awareness of skills involving student discovery, seeking alternative explanations, and questioning student perceptions in contrast to only teaching content.

There are multiple methods of faculty development that are believed to be effective. Ash, Brown, Kluger-Bell, and Hunter (2009) suggest a connection between professional community, faculty development and critical thinking instruction. Professional collaboration is an integral part of the faculty development process. The development and refinement of new instructional strategies is most effectively done as part of a professional community (Eddy and Garza Mitchell, 2012). Collaboration between faculty members is often described as innovative and energizing (Hill, Kim, and Lagueux, 2007; Eddy and Graza Mitchell, 2012; Stevenson, Duran, Barret and Colarulli, 2005). Faculty report that preparing for courses can be an isolating experience (Stevenson et al, 2005; Thomson, 2007), the process of interdisciplinary collaboration not only provides a sense of community but also offers a fresh perspective.

While faculty development training around improving student critical thinking skills is prevalent, little research has been conducted on the impact of using training around a valid, reliable assessment tool which has been developed with input from multiple disciplines and multiple institutions as a central component of a faculty development model. Huber (2015) suggests, in general, a gap in the literature about the evidence that teaching and learning training programs work and are transferable across multiple institutions. An early research project on the effectiveness of faculty development programs on stimulating critical thinking conducted by Gibbs, Browne, & Keeley (1988) did not find the development program effective. Obstacles included the reluctance of faculty to admit skill deficiencies, difficulty of getting agreement about critical thinking across the disciplines, the time needed to make changes, there is a tension for faculty between teaching discipline material and critical thinking skills, and limitations of the existing assessment tools at the time. Huber (2015) conducted research at Carleton University and Washington State University on a faculty development model centered on the Washington State University's (WSU's) Guide to Rating Critical Thinking. Huber found that well designed faculty development training on critical thinking can impact faculty teaching. However, the rubrics developed at WSU to evaluate student learning were not as sensitive of a measure for Carleton University. This study will explore whether the CAT-App development training which is based upon the CAT which has established face validity for faculty across multiple institutions can be used as a component of a successful faculty development model.

Using the CAT Instrument to Drive Change in Teaching and Assessment

While there is widespread agreement among faculty about the importance of critical thinking, the definition of critical thinking is much more controversial. The classic definition of critical thinking is examining logical fallacies or formal reasoning (Drefus & Jungwirth, 1980; Garnett & Tobin, 1984; Bailin, 2002). This classic definition of critical thinking can be traced back to Plato and Aristotle (Hughes, Lavery, and Doran, 2014). More recently though the definition of critical thinking has been broadened to items such as problem solving and creativity (Halpern, 1999). Many of the current definitions of critical thinking (Ennis, 1962; Facione, 1990; Halpern, 1997) define critical thinking in terms of skills.

Critical thinking in this article is based upon a contemporary, inclusive, skill based approach. The CAT instrument was developed by an interdisciplinary team of faculty at Tennessee Technological University to assess the critical thinking skills (see table 1) that faculty felt their students needed the most ten years after they graduated in order to be successful in their careers. The National Science Foundation (NSF) provided funds to nationally refine and disseminate the instrument.

Table 1. Critical Thinking Skills Associated with the CAT Instrument

Skill Areas
Separate factual information from inferences that might be used to interpret those facts.
Identify inappropriate conclusions.
Understand the limitations of correlational data.
Identify evidence that might support or contradict a hypothesis
Identify new information that is needed to draw conclusions.
Separate relevant from irrelevant information when solving a problem.
Learn and understand complex relationships in an unfamiliar domain.
Interpret numerical relationships in graphs and separate those relationships from inferences.
Use mathematical skills in the context of solving a larger real-world problem.
Analyze and integrate information from separate sources to solve a complex problem.
Recognize how new information might change the solution to a problem.
Communicate critical analyses and problem solutions effectively.

The CAT instrument is faculty driven - it was designed by faculty to assess the skills which faculty think are most important (Stein and Haynes, 2011). The skills assessed by the CAT instrument provide “a de facto definition of the aims of college” (Miller, 2012) and promote faculty involvement in assessment (Ewell, 2010; Ewell, 2014). The CAT instrument directly involves faculty in the scoring of student essay responses so faculty can see, first hand, their students’ strengths and weaknesses. Although the need to involve faculty in the scoring of the CAT instrument is considered a burden by some, research conducted by the National Institute for Learning Outcomes Assessment (NILOA) indicates faculty prefer assessments that allow them to see and talk about student work (Kinzie, 2010). Directly experiencing student weaknesses and

having the opportunity to discuss effective practices for addressing the weaknesses are two essential elements that underlie the effective use of assessment for the improvement of student learning (Wolcott and Gray, 2003).

The CAT instrument is relatively unique because it is one of only a few critical thinking assessment tools available that uses primarily short answer, essay responses to assess critical thinking. Short answer essay tests to assess critical thinking have several advantages. Specifically, many authentic real world situations that require critical thinking and problem solving do not have a simple answer or a simple set of alternatives from which to choose the best answer. Providing such alternatives on a test can dramatically alter the original problem and greatly simplify the complexity of the task. Additionally, communication and creativity skills cannot be evaluated with a multiple-choice test. Many would argue that the reasons given to support a specific answer are as important as the answer itself. The short answer essay format allows the CAT instrument to examine such reasoning and thought patterns. The CAT has been used at over 200 colleges and universities across the mainland United States, Alaska, Hawaii, and Guam. It has also been used in Australia, Japan, Palestine, and Canada. The CAT has not been found to have cultural bias when used at institutions in the United States. Cultural bias for outside the United States has not yet been fully assessed since these studies are still under way. It has been used across a diverse range of institutions from community college to Ivy League institutions and has been used in both STEM (Science, Technology, Engineering, and Mathematics) and non-STEM disciplines.

The CAT instrument incorporates a process known as “dynamic assessment” (e.g., Campione and Brown, 1990; Feuerstein, 1979; Lidz, 1987; Samuels, 2000; Sternberg and Grigorenko, 2002). A series of increasingly deeper and more explicit question prompts are used to engage students’ critical thinking skills to measure the extent to which people can understand and evaluate new information and apply that information to a novel situation. Deeper prompting frequently helps reveal critical thinking skills that might not otherwise be observed and provides a more valid measure of the students’ potential to think critically. The CAT instrument uses dynamic assessment coupled with problems that are intrinsically interesting to students and representative of real-world problems. These features contribute to student motivation to perform well on the test and improve the validity of the instrument.

The CAT instrument allows faculty to directly experience student weaknesses when approaching real-world problems. In addition to obtaining quantitative information about students’ performances, faculty are able to directly experience the strengths and weaknesses of their students’ ability to perform various types of critical thinking tasks. Many institutions have found high levels of faculty interest in participating in scoring workshops and increased interest in having assessments of their own students (including R1 institutions). These observations are particularly important in the light of a recent national survey that revealed faculty attitudes toward assessment are a major challenge for improving outcomes assessment in higher education (Kinzie, 2010; Kuh and Ikenberry, 2009).

While the CAT instrument was initially developed as an interdisciplinary tool to assess students’ critical thinking, it can also serve as a model for helping faculty develop better discipline specific assessments of critical thinking. In fact, some faculty are able to immediately see how they can transform their courses using assessments modeled on the CAT but tailored to their own discipline content, but most faculty seem to require more training and resources to develop these CAT-Apps.

The realization that more support may be needed to help faculty develop better discipline assessments led to efforts to develop CAT-App training. The goal of this training is to provide

individual or groups of faculty with more guidance in the design of classroom level assessments in their discipline to improve critical thinking skills. This training is based upon principles of cognitive psychology such as that presented in the sentinel work of Bransford, Brown, and Cocking, *How People Learn*. Faculty in CAT-App training learn to develop course assessments based upon the IDEAL framework of problem solving of Bransford and Stein (1993). A key to this process is that it takes a constructive alignment approach to course development where the student learning goals are aligned with teaching and classroom assessments (Biggs, 1996). It is also consistent with threshold concepts (Meyer and Land, 2003) as faculty identify key concepts to teach and assess in their courses. The alignment of this training with these established concepts in cognitive psychology that have been used both across institutions and across disciplines increases the probability of the training to be effective and portable across institutions and disciplines. The training also seeks to establish a national teaching and learning community where faculty work together to improve course assessments and thus student learning.

The current study explores the impact that participation in the scoring of the CAT instrument can have on faculty at one institution and the extent to which it helps faculty understand student weaknesses in critical thinking and motivates faculty to change their instructional methods. The current study will also explore how additional training in developing CAT-Apps in their discipline can impact faculty willingness to change their instructional methods and the potential value of such training for institutional change. This study will explore faculty attitudes toward the CAT and how participating in CAT scoring and CAT-App training has served as an impetus toward course changes.

Methodology and Findings

This study examines converging evidence of the CAT as a faculty development tool to both motivate faculty and to enhance their teaching of critical thinking skills in the classroom. We will examine the results of a focus group of CAT scorers at a regional university, Eastern Kentucky University (EKU), as well as a survey of CAT users attending special training sessions to develop CAT-Apps to see the potential impact of the CAT and corresponding CAT-Apps.

Eastern Kentucky University (EKU) Focus Group

An informal qualitative focus group was conducted at Eastern Kentucky University (EKU), a mid-sized, Southeastern university in the United States to discuss faculty reactions to the CAT scoring process as related to faculty development. The institution had been utilizing the CAT instrument for over five years and had over one hundred experienced faculty scorers. All faculty members who had scored the CAT instrument at this institution during the previous five years were invited to attend. Participants were not compensated for participation in this focus group. Prior to the session, the coordinators at EKU prepared five question prompts for faculty attendees to encourage discussion in relation to changes that they had made in their courses as a result of participating in CAT scoring workshops. Question prompts included:

1. Have you added content to your courses?
2. Have you modified your teaching approach?
3. Have you modified lesson plans or assignments?
4. Have you modified grading or feedback approaches?
5. Have you added or modified student-learning outcomes?

The sessions were held over two days with 20 faculty members in attendance. A faculty member from this institution facilitated the session with the assistance of personnel from the university assessment center.

Following the focus group meeting, a survey was distributed to all participants to obtain additional feedback about the session. All 20 faculty participants responded to the survey, which included both open and close-ended questions.

Focus Group Discussion Twenty participants in the university assessment focus group discussed the five questions related to changes that had been made in their teaching or assessment practices since participating in CAT scoring. Researchers hand-recorded participant comments related to each of the five questions. These comments were later reviewed and common statements were withdrawn from the researchers' notes. Faculty discussion indicated that change had been made in each of the areas reviewed. Participants indicated that since scoring the CAT instrument they were more selective on the content used in both class examples and exams. Faculty stated that they are focused on going "beyond 'tell me numbers and trends.'"

These changes extended beyond content choices with faculty indicating changes in both teaching approach and assignments. After scoring student responses to the CAT and gaining a greater understanding of students' strengths and weaknesses in the skill areas assessed, faculty indicated a greater focus on not just evaluating information, but separating relevant from irrelevant information. Participants indicated that they have begun presenting both types of information to students in order to more effectively model a real-world scenario inside the classroom. Many faculty members indicated an increased effort to go beyond just teaching facts and to, more specifically, help students creatively think by developing and evaluating alternative explanations to problems.

Respondents indicated a greater focus on integrating critical thinking skills into discipline specific content. One faculty member specifically noted a shift in their theoretical understanding of teaching critical thinking from a focus on independently teaching skills to an approach where these skills are infused into the course content and students are given an opportunity to apply these skills.

Faculty participants also indicated shifts in assessment practices. Multiple respondents noted a change in question prompts to mirror the model of dynamic assessment utilized in the CAT instrument. Faculty referred to this as "progressive prompting" or a "step by step approach" and indicated that it allowed students to understand the process of problem solving.

Focus Group Participant Survey Focus group participants were invited by the EKU assessment center to participate in a short survey about their experiences with CAT scoring. Of the twenty faculty members that responded to the five-question instrument, 95% stated that they had changed the way they presented information to their students based on their participation in CAT scoring. Respondents were asked in an open-ended question to share what they would tell a colleague about participating in CAT scoring. Responses to this question were coded and the following results were identified. Survey responses were read and 12 initial codes used a total of 38 times were recorded. These codes were grouped into five primary categories related to student learning and teaching change. The in-vivo code "eye-opening experience" emerged from the data as this sentiment was embedded in multiple faculty responses. The frequency of occurrence for each category was counted and the percent of inclusionary responses was calculated (see table 2).

Table 2. Case Study: Impact of the CAT

Category	Percent of respondents including this code (n)
Greater understanding of student learning	75% (15)
Teaching changes/impact	45% (9)
Eye-opening experience	25% (5)
Qualities of faculty development	10% (2)
CAT-App Development	10% (2)

Seventy-five percent of faculty respondents indicated that their experience scoring the CAT led them to be more aware of their students' strengths and weaknesses in regards to critical thinking. Participants indicated that reading, scoring, and discussing student responses to the CAT questions allowed them to "think critically about...their students' thinking," as well as "learn so much about what students know and don't know." Multiple faculty members recognized this opportunity to "gain insight" to both how students think and how they apply what they learn.

This greater understanding of student strengths and weaknesses led many faculty participants to acknowledge changes that could be made in their own courses. Faculty noted the importance of thinking critically about their pedagogy and what they want their students to learn. Forty-five percent of respondents accredited this experience as an opportunity to modify teaching approaches in an effort to best teach critical thinking skills. One faculty member described this as "an eye opening experience that can highlight your strengths and weaknesses as a teacher." Variations of the phrase "eye-opening experience" were used by 25% of participants when asked what they would tell a colleague about participating in a CAT scoring session. While 75% of respondents reported better understanding of students' strengths and weaknesses and 45% of faculty reported changing how they taught, only 10% reported the development of CAT-Apps.

As with any *ex post facto* approach to data collection, certain limitations were encountered. The participants in the sample were self-selected by their response to the invitation. This could bias the sample toward faculty that are more intrinsically motivated or are more interested in this process.

CAT App Training

The CAT-App training is a workshop focusing on assisting faculty in the development of discipline specific assessments (CAT-Apps) that target similar critical thinking skills as those measured by the CAT test. These CAT-Apps will allow students to practice the skills that the CAT assesses within the content of a course. By infusing these activities within their courses, faculty will merge the assessment of both discipline specific content and critical thinking.

We have conducted three pilot workshops to assess the potential benefits of these workshops. These pilot workshops were conducted at Tennessee Technological University, Charleston, South Carolina, and Knoxville, Tennessee. We had a total of 35 attendees with faculty and administrators from 11 states and Australia. These faculty were from a variety of institutions such as community colleges, regional colleges, research universities, and private, elite institutions. Participants were not compensated for participation in this survey.

Most participants at these sessions had participated in either a previous CAT Train-the-Trainer workshop or a scoring session at their home institution. At the conclusion of each of the

pilot CAT-App trainings, an online survey was disseminated to participants via an anonymous link to an instrument housed within Qualtrics Survey Software. This survey focused on participants' experiences in the training workshop, their prior experience scoring the CAT, changes they planned to make to their assessment and instructional practices, and factors that motivate and restrict faculty change.

CAT-App Training Survey At the conclusion of each CAT-App training, participants were invited to participate in an online survey regarding their experiences with the training, along with questions related to faculty motivation and barriers to change. Thirty participants fully completed the survey. Respondents self-identified their predominant role at their institution. The frequency distribution is listed below (see table 3) – the cumulative percentage is greater than 100 because some respondents selected more than one category.

Table 3. Respondents Predominant Role

Predominant role at home institution	Percent of respondents (n)
Teaching Faculty	43% (15)
Research Faculty	10% (4)
Administrator/Staff	30% (11)
Other	20% (7)

As stated previously, these pilot sessions were used to gauge faculty interest in a National Teaching and Learning Community focused on the development of CAT-Apps. As part of this National Teaching and Learning Community participants would develop and enter CAT-Apps into a searchable, online database that would be shared with a national community. Participants confirmed this interest with 96.7% of respondents agreeing that an online database of CAT-Apps would be beneficial for faculty.

The CAT-App training workshop focuses on assisting participants with the development of multiple discipline-specific assessments that they can take back and use in their courses. This training was developed after noting two major difficulties of faculty spontaneously developing CAT-Apps as witnessed by only 10% of the faculty in our case study developing CAT-Apps: (1) lack of training and (2) lack of time. This CAT-App training, along with participation in CAT scoring, enables faculty to identify the areas within their courses that can be modified to better focus on the skills associated with critical thinking improvement. Respondents were surveyed about their instructional practices before the CAT-App training, as well as their plans following the training. A paired-samples t-test was conducted to compare the percentage of class time devoted to lecture-based instruction before and after training participation. There was a significant difference in the responses, $t(21) = 4.11$, $p = .001$, before training ($M = 39.14$, $SD = 21.43$) and anticipated change after training ($M = 31.73$, $SD = 18.76$). This indicates that participants plan to devote significantly less time to lecture-based instruction after participation in CAT-App training (see figure 1). These faculty began with less time spent lecturing than is reported in other studies such as the study by Watts and Schaur (2011) that reported the median time in Principles of Economics classes spent lecturing was 83% of course time. This data was collected immediately after the training, so a follow up study could be conducted to assess if these changes were put into practice in participants' courses. This would further confirm the evidence discussed previously

from the EKU case study which included faculty up to five years after participation in CAT scoring.

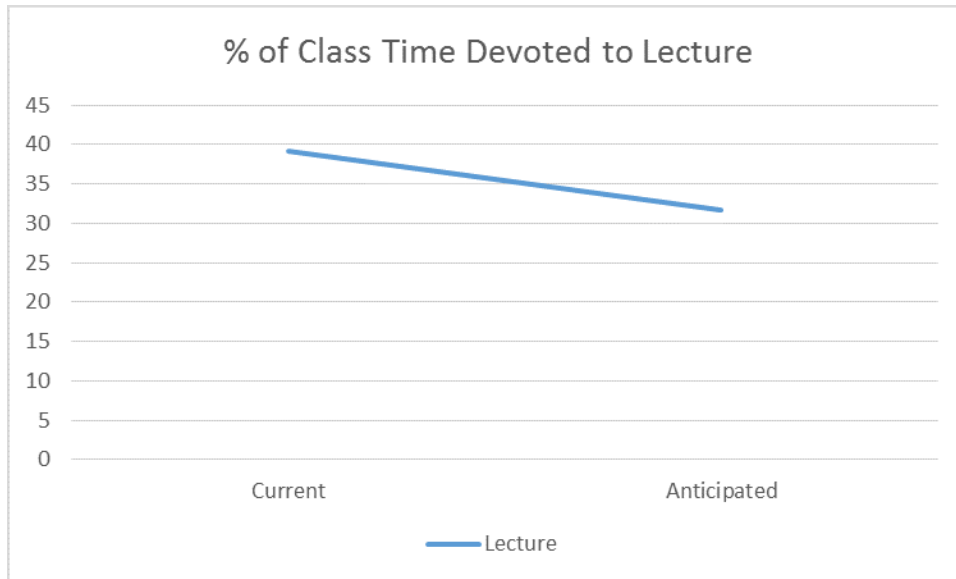


Figure 1. Percent of Class Time Devoted to Lecture

Additionally, participants were asked to identify what percentage of students’ final grade is determined by factual-knowledge tests, as opposed to critical thinking assessments. A paired-samples t-test was again conducted on this data to compare the scope of use of each type of assessment before and after participation in CAT-App training. As displayed in Table 4, there were significant differences, at the .01 level, in participant responses regarding their assessment practices before and plans for after the training. Results indicated that respondents anticipate the percentage of students’ grades based upon factual-knowledge tests to decrease and the percentage of students’ grades based upon critical thinking assessments to increase after their participation in the CAT-App training. Figure 2 illustrates this difference and depicts the divergence in assessment practices following the CAT-App training.

Table 4. Changes in Types of Assessments

	Before Training		After Training		<i>df</i>	<i>t</i>	<i>r</i>	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Factual-knowledge assessments	48.91	29.51	40	24.13	21	3.58	0.93	0.002
Critical thinking assessments	55.32	28.01	64.6	21.77	21	-3.8	0.93	0.001

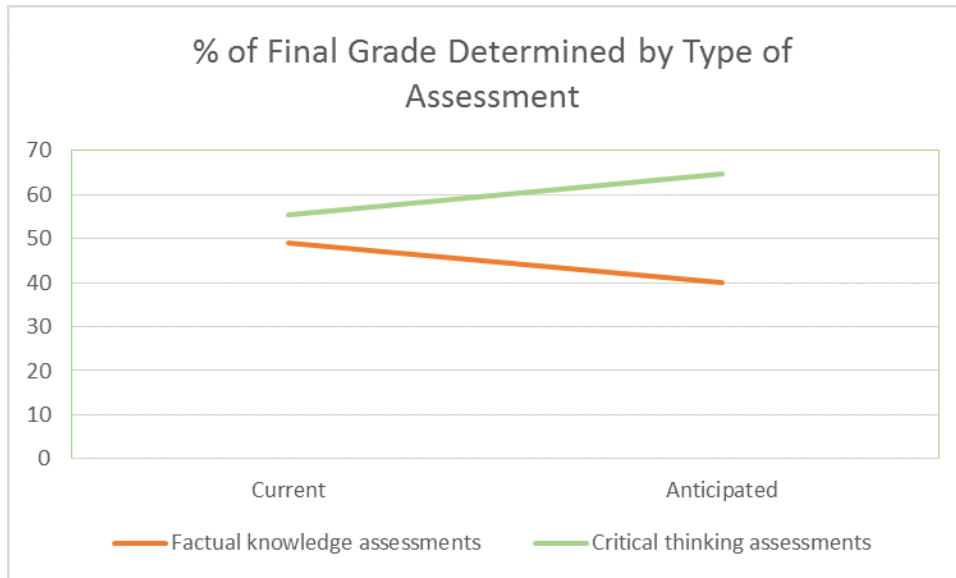


Figure 2. Percent of Final Grade Determined by Type of Assessment

Survey respondents also answered a series of questions related to hurdles they face when making changes in their courses to improve critical thinking. Table 5 includes a list of the potential barriers included in the survey. Respondents were presented with each statement and asked to express their level of agreement as to whether or not it was a hurdle to making changes in their courses to improve critical thinking. Level of agreement was expressed via responses to a six-point Likert scale with response options ranging from strongly disagree to strongly agree. A bivariate correlation matrix was computed to examine the relationship between each of these hurdles and the assessment and instructional practices of participants prior to their participation in CAT-App training. There was a positive correlation between respondents reporting a lack of personal motivation to make changes and those devoting more time to lecture-based instruction, $r(21) = +.43, p = .038$. Respondents that found a lack of personal motivation to be a barrier to change were more likely to use a higher percentage of class time devoted to lecture to teach students in their courses. Conversely, there was a negative correlation between respondents reporting a lack of support from administrators and those devoting more time to active learning, $r(21) = -.42, p = .047$. This indicates that participants that receive greater levels of support from administration at their institution are more likely to use active learning strategies in their courses.

Table 5. Barriers to Faculty Change

Potential Barrier	Mean
Finding the time to modify your course	.43
Finding the time to implement changes in your course	.08
Lack of support from other faculty	.63
Lack of support from administrators	.73
Lack of interest or push-back from students	.7
Lack of personal motivation to make changes	.7
Lack of training	.7
My discipline does not lend itself to these skills	.4

Additionally, significant negative correlations were found between respondents disagreeing that a lack of training was a barrier to change and those reporting a higher percentage of class time devoted to active learning, $r(21) = -.41$, $p = .05$, and those reporting that a greater percentage of students' final grade is determined by critical thinking assessments, $r(21) = -.61$, $p = .003$. These results indicate that faculty that receive more training are more likely to use active learning strategies in their courses and are also more likely to use critical thinking and real-world problem solving assessments to determine their students' final grades.

Respondents also acknowledged the integrated components of faculty development within the scoring process of the CAT test. One hundred percent of respondents that felt able to evaluate the features of the CAT stated that it was valuable to have faculty involved in the scoring process. Of those respondents, 95.6% stated that the involvement of faculty scorers was a unique property of the CAT. Faculty answered a series of questions related to instructional changes made since participation in CAT scoring. Of the faculty that had participated in CAT scoring, 75% stated that they had modified course activities based on this experience. It would stand to reason that the observation of student strengths and weaknesses in regards to these critical thinking skills motivates faculty to make changes in their instructional practices.

Conclusion

There are acknowledged limitations to these studies discussed in this article. The participants of both studies are based upon their self-selection. Both groups of faculty volunteered their time to either score the CAT test or participate in CAT training so they had some interest in critical thinking. Both studies were based upon faculty self-reported data. Longitudinal studies are needed about the long lasting impact of these CAT-App training workshops on faculty changes. Neither of these studies actually measures whether there were gains in student learning as a result of faculty changes. Additional studies are needed that measure the impact of the CAT-Apps developed by faculty on student learning. However, despite these limitations, these studies provide important insights for the scholarship of teaching and learning.

Scoring workshops on college campuses are helping build teaching communities that are highly receptive to new ideas for improving student learning in STEM and non-STEM disciplines.

As a consequence of this increased faculty interest, many institutions are investing additional resources in broad efforts to improve students' critical thinking skills.

The case study research reported here confirms that these investments have the potential to improve teaching and learning at these institutions. The evidence shows that participating in CAT scoring workshops motivate faculty to make changes in their classes. However, often these scoring workshops alone do not provide enough training and support for faculty to develop CAT-Apps.

The survey from the CAT-App training shows that it provides additional motivation for faculty to improve their teaching practices such as spending less time lecturing. It also demonstrates that faculty feel there is a need for a database of CAT-Apps. This is consistent with the early research on faculty development around critical thinking conducted by Gibbs, Browne, and Keeley (1988) that faculty motivation to teach and assess critical thinking improves when they are trained to simultaneously assess both critical thinking and discipline specific material therefore reducing the tension between the two. Also, faculty would appreciate materials such as a CAT-Apps database that would reduce the additional time and effort required to produce critical thinking assessments for their classrooms. Finally, it demonstrates that the participation in CAT-App training shifts faculty toward more of a focus on critical thinking assessments and less time on lecturing in their classes. Overall, it shows that faculty development training on developing effective classroom assessments such as CAT-Apps based upon the principles of cognitive science such as constructive alignment and threshold concepts can be an important component of a faculty development model. Having an assessment that has face validity for faculty across institutions and disciplines where faculty can see first-hand student weaknesses provides a motivation for faculty to change. The findings also indicate faculty interest in being part of a national teaching and learning community centered on critical thinking.

The findings of this study are consistent with other research on the impact of faculty development programs centered around critical thinking. Like most of the other studies in this area (Ash, Brown, Kluger-Bell, & Hunter, 2009; Behar-Horenstein, Schneider-Mitchell, & Graff, 2009; Huber, 2015), this study found that faculty development around critical thinking can be an effective way to engage faculty in improving teaching and learning. This study goes beyond the previous research though to explore how participation in training to developing classroom assessments of critical thinking around the skills on the CAT, CAT-Apps, can reduce the amount of time spent lecturing and the type of overall assessment approach of a course across multiple institutions. This study also explores impediments to change. Understanding impediments to change is essential if we desire to implement classroom changes that improve student learning.

Acknowledgements

Partial support for this work was provided by the National Science Foundation's TUES Program (formerly CCLI Program) under grant 1022789. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Association of American Colleges and Universities. (2013). *It takes more than a major: Employer priorities for college learning and student success*. Retrieved from <http://www.aacu.org>
- American Management Association. (2012). *AMA 2012 critical skills survey*. Retrieved from <http://www.amanet.org>
- Anderson, L., & Sosniak, L. (1994). *Bloom's taxonomy: A forty-year retrospective*. Chicago, IL: University of Illinois Press.
- Arum, R., Roksa, J., & Cho, E. (2011) *Improving undergraduate learning: Findings and policy recommendations from the SSRC-CLA Longitudinal Project*. Retrieved from <http://www.ssrc.org>
- Ash, D., Brown, C., Kluger-Bell, B., & Hunter, L. (2009). Creating hybrid communities using inquiry as professional development for college science faculty. *Journal of College Science Teaching*, 38(6), 68–76. Retrieved from <http://www.nsta.org/college>
- Bailin, S. (2002). Critical thinking and science education. *Science and Education*, 11, 361-375.
- Behar-Horenstein, L. S., Schneider-Mitchell, G., & Graff, R. (2009). Promoting the teaching of critical thinking skills through faculty development. *Journal of Dental Education*, 73(6), 665–675. Retrieved from <http://www.jdentaled.org/>
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364.
- Bok, D. (2006). *Our underachieving colleges: A candid look at how much students learn and why they should be learning more*. Princeton, NJ: Princeton University Press.
- Boud, D., & Falchikov, N. (2007.) Introduction: Assessment for the longer term. In D. Boud & N. Falchikov (Eds.), *Rethinking Assessment in Higher Education: Learning for the Longer Term* (pp. 3-13). New York, NY: Routledge.
- Bransford, J. Brown, A. & Cocking (2000). *How people learn*. Washington, D.C.: National Academy Press.
- Bransford, J. & Stein, B. (1993). *The IDEAL problem solver*. New York: Freeman.
- Campione, J. C., & Brown, A. L. (1990). Guided learning and transfer: Implications for approaches to assessment. *Diagnostic monitoring of skill and knowledge acquisition*, 141-172.
- DeAngelo, L., Hurtado, S., Pryor, J. H., Kelly, K. R., Santos, J. L., & Korn, W. S. (2009). *The American college teacher: National norms for the 2007–2008 HERI faculty survey*. Retrieved from <http://www.heri.ucla.edu>

Dreyfus, A., & Jungwirth, E. (1980). Students' perception of the logical structure of curricular as compared with everyday contexts—study of critical thinking skills. *Science education*, 64(3), 309-321.

Eddy, P. L., & Garza Mitchell, R. L. (2012). Faculty as learners: Developing thinking communities. *Innovative Higher Education*, 37(4), 283–296. doi:10.1007/s10755-011-9202-z

Ennis, R. (1962). A concept of critical thinking. *Harvard Educational Review*, 32, 81-111.

Estep, C. M., Roberts, T. G., & Carter, H. S. (2012). An experiential learning model of faculty development to improve teaching. *NACTA Journal*, 56(1), 79–86. Retrieved from <http://www.nactateachers.org/journal.html>

Ewell, P. (2010). Foreword to *Opening doors to faculty involvement in assessment*. National Institute for Learning Outcomes Assessment. Champaign, IL: University of Illinois. Retrieved from: <http://www.learningoutcomeassessment.org>

Ewell, P. (2014). The “quality agenda.” An overview of current efforts to examine quality higher education. A discussion paper prepared for the American Council on Education (ACE). Retrieved from <http://www.acenet.edu/>

Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations.

Feuerstein, R., Rand, Y. A., & Hoffman, M. B. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Scott Foresman & Co.

Garnett, P. J., & Tobin, K. (1984). Reasoning patterns of preservice elementary and middle school science teachers. *Science Education*, 68(5), 621-631.

Gayton, E. (2007.) “Introduction: Why these assessment opportunities make sense in a world where assessment of factual knowledge has taken hold.” In R. J. Mazeske & B. A. Mazeske (Eds.). *Beyond tests and quizzes: Creative assessments in the college classroom* (pp. xvi-xxiv). San Francisco, CA: Jossey-Bass.

Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students’ learning. *Learning and teaching in higher education*, 1(1), 3–31. Retrieved from <http://www.glos.ac.uk/about/teachinglearning/Pages/lathe.aspx>

Gibbs, L., Browne, M. & Keeley, S. (1988). Stimulating critical thinking through faculty development: Design, evaluation, and problems. Washington, D.C.: American Association of State Colleges and Universities.

Halpern, D. (1997). *Critical thinking across the curriculum: A brief edition of thought and knowledge*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Halpern, D. F. (1999). Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New directions for teaching and learning*, 1999(80), 69-74.

Hill, L., Kim, S. L., & Lagueux, R. (2007). Faculty collaboration as faculty development. *Peer Review*, 9(4), 17. Retrieved from <https://www.aacu.org/peerreview>

Huber, M. T. (2015) Faculty from faculty learning to student learning In Condon, et.al. (Eds.) *Faculty development and student learning* Bloomington: Indiana University Press.

Hughes, W. Lavery, J. and Doran, K. (2014). *Critical thinking: An introduction to the basic skills*. Calgary: Broadview Press.

Hutchings, P. (2010). *Opening doors to faculty involvement in assessment*. National Institute for Learning Outcomes Assessment. Champaign, IL: University of Illinois.

Ikenberry, S. (2009). The national institute for learning outcomes assessment. Paper presented at the meeting of the Southern Association of Colleges and Schools Commission on Colleges, Atlanta, GA.

Kinzie, J. (2010). *Perspectives from campus leaders on the current state of student learning outcomes assessment: NILOA focus group summary 2009-2010*. National Institute for Learning Outcomes Assessment. Champaign, IL: University of Illinois.

Kuh, G., & Ikenberry, S. (2009). *More than you think, less than we need: Learning outcomes assessment in American higher education*. National Institute for Learning Outcomes Assessment. Champaign, IL: University of Illinois.

Kvale, S. (2007). Contradictions of assessment for learning in institutions of higher learning. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term*, (pp. 3–13). New York, NY: Routledge.

Lidz, C. S. (1987). *Dynamic assessment: An interactional approach to valuating learning potential*. New York: Guilford Press.

Meyer, J. & Land, R. (2003). Threshold concepts and troublesome knowledge. In Rust C. (ed.) *Improving student learning-ten years on*. Oxford: OCSLD.

Miller, M. (2012). From denial to acceptance: The stages of assessment. *National Institute for Learning Outcomes Assessment*. Champaign: University of Illinois.

National Survey of Student Engagement. (2013). *Annual Results*. Retrieved from http://nsse.iub.edu/html/annual_results.cfm

Northeastern University. (2013). Key Findings. *Innovation imperative: Enhancing higher education outcomes*. Retrieved from <http://www.northeastern.edu>

Pascarella, E.T., Blaich, C., Martin, G.L., & Hanson, J.M. (2011). How robust are the findings of *Academically Adrift?* *Change: The Magazine of Higher Learning*, 43(3), 20–24. Retrieved from <http://www.changemag.org>

Petress, K. (2004). Critical thinking: An extended definition. *Education*, 124(3), 461–467. Retrieved from <http://www.projectinnovation.biz/education>

Samuels, M.T. (2000). Assessment of post-secondary students with learning difficulties: Using dynamic assessment in a problem-solving process.” In C.S. Lidz & J.G. Elliott (Eds.), *Dynamic Assessment: Prevailing Models and Applications*, (pp. 521–542). Amsterdam:JAI/Elsevier Science

Sternberg, R.J. and Grigorenko, E.L. (2002). *Dynamic Testing: The Nature and Measurement of Learning Potential*. Cambridge, England: University of Cambridge.

Stevenson, C. B., Duran, R. L., Barrett, K. A., & Colarulli, G. C. (2005). Fostering faculty collaboration in learning communities: A developmental approach. *Innovative Higher Education*, 30(1), 23–36. doi:10.1007/s10755-005-3293-3

Spellings, M. (2006). A test of leadership: Charting the future of U.S. higher education. Retrieved from <http://www.ed.gov>

Stein, B., & Haynes, A. (2011). Engaging faculty in the assessment and improvement of students’ critical thinking using the CAT. *Change: The Magazine of Higher Learning*, 43, 44–49. doi:10.1080/00091383.2011.550254

Stein, B., Haynes, A., & Harris K. (2009). Assessing and improving critical thinking using the CAT instrument, *ABET Best Assessment Processes Symposium XI*. Symposium conducted at the meeting of Accreditation Board of Engineering and Technology, Indianapolis, IN. Retrieved from <http://www.tntech.edu/cat>

Thomson, A. (2007). Faculty collaboration: Creating multidisciplinary learning communities. *The Community College Enterprise*, 13(2), 27–38.

Vogler, K. E. (2002). The impact of high-stakes, state-mandated student performance assessment on teachers’ instructional practices, *Education* 123(1), 39–56.

Wolcott, S.K., & Gray, C.J. (2003). Assessing and developing critical thinking skills. Paper presented at the 2003 Assessment Institute. Purdue University, Indianapolis, IN.

Watts, M. and Schaur, G. (2011). Teaching and assessment methods in undergraduate economics: A fourth national quinquennial survey. *The Journal of Economic Education*, 42(3), 294-309.

Wilson, M. (2009). Assessment from the ground up. *Phi Delta Kappan*, 91(1), 68-71.