AC 2011-1015: SIMPLIFYING ASSESSMENT USING DIRECT MEASURES

Ronald W. Welch, University of Texas, Tyler

Ron Welch is Professor and Head, Department of Civil Engineering at The University of Texas at Tyler. He is a registered Professional Engineer in Virginia. Until 2 Jan 2007, Ron was an Academy Professor at the United States Military Academy (USMA). Ron received a BS degree in Engineering Mechanics from the USMA in 1982 and MS and Ph.D. degrees in Civil Engineering from the University of Illinois at Urbana-Champaign in 1990 and 1999, respectively. Ronald.Welch@uttyler.edu.
Simplifying Assessment Using Direct Measures

Abstract

Almost everyone has heard or read somewhere that to complete quality assessment a program must use multiple assessment methods, and furthermore that indirect measures such as surveys cannot be the only source of assessment. These types of issues bring up challenging questions. How many assessment methods are enough? What is the required depth for each assessment method? What direct measures are best for your program? This paper will provide some insight into choosing direct measures and how they have been used by one program.

There are many reasons for a program to use multiple assessment methods. One important aspect is the ability to provide multiple view points for each program outcome which assist in determining trends in performance that might necessitate program changes. Although surveys are an easy assessment method, their low rates of return and many times poor quality of questions result in them not particularly being effective instruments to drive program improvement. Most programs have turned to multiple direct measures to ensure they can properly demonstrate their student’s accomplishment of an outcome. The direct measures presented in this paper are: embedded indicators, capstone (point allocation), external exam (Fundamentals Engineering Exam, FE), and internal exam (Gateway). This paper will also present the lessons learned by one program through their four years of using these direct measures and the program’s success in reducing the work and data collected as well as the superb ABET visit results based on using these methods (no deficiencies or weaknesses at the end of the on-campus ABET visit). Additionally, this paper will also present the current program efforts to further reduce the assessment data collected and time required to the irreducible minimum.

1.0 Introduction

1.1 Assessment Measures

Assessment using a single data point to make a major decision about changes to anything, but especially a program would be absurd. Therefore, programs select multiple assessment methods to provide varying views and a collection of data points to assist in determining whether an outcome has been met or not. Some analogies to the need for multiple assessment measures would be the need for multiple camera views during the making of a movie or multiple camera views during instant replay calls in sports or multiple identical tests during research. Each view or test provides critical insight into the conclusion before making the decision for a call reversal, research publication, or changes to a program.

Some of the common assessment methods used around the country are: External Exams, Internal Exams, Capstones, Surveys, Embedded Indicators, Course Assessments, and Industrial or External Advisory Committees. External Exams are the most desired since they provide an assessment of your students against state or national averages. Many
states have developed countless exams to determine the demonstrated capabilities of their students in response to “No Child Left Behind” legislation (TAKS test in Texas). Engineering uses the Fundamental Engineering Exam (FE) for not only the first step in professional licensure, but also to demonstrate the ability of graduates based on the percentage of their students who actually take the exam and represent the student body demographics. Some programs use Internal Exams (Gateway Exams) to determine if a student is prepared to graduate or enter the next phase of the program such as the junior year in nursing or to enter into an engineering program. ABET accreditation requires a major design experience or Capstone where students use most of their skills to solve some multi-disciplinary problem using knowledge and skills developed within the curriculum. Some programs are using the grading of portions of the Capstone to demonstrate some or all of their outcomes. Surveys of students, faculty, alumni, and employers provide insight into how the program is performing while the students are in school as well as after they have graduated in comparison to the desired capabilities of established program outcomes and program educational objectives. Embedded Indicators have become very popular in recent years as an effective method of demonstrating student accomplishment for an outcome. Since the faculty are already assessing individual assignments and exams as part of their normal course activities, then the use of portions or all of an assignment or exam as a demonstration of accomplishment of an outcome seems rather obvious. Course assessments are used in a number of programs to assess each course’s accomplishment of course objectives that are tied directly to certain program outcomes. The process also ensures any changes in one course are discussed by the entire faculty to ensure the effect on follow-on courses is understood or minimized. Industrial or External Advisory Committees are industry partners, both locally and nationally, that provide direction, focus, and assessment of the program based on their companies experiences and needs from the next generation of engineers.

Most programs start off using surveys since they are so easy to build and send out, but the return rates are many times extremely low. Even though surveys do produce useful data points if the return rates are acceptable, the results are subjective based on the responder’s bias unless the questions are crystal clear. However, low survey response rates appear to be acceptable when the group is composed of homogeneous populations with strong attitudes, perspectives due to similar age, sex, location, etc. With such a wide variety of career paths for CE graduates, low return rates can be problematic. Anyone can use SurveyMonkey to generate results, but the crafting of survey questions is an art and properly developed survey questions are necessary if the results are to be useful. ABET recognizes the issue and conducts webinars to help administrators to develop proper surveys. Surveys are critical with collecting data for program educational objectives – what graduates are accomplishing four to eight years after graduation. Key data sources are alumni and employers which are difficult to gain access to without surveys. Using surveys of seniors and faculty as the sole source of data collection to determine if students can demonstrate an outcome is not particularly desirable since sometimes faculty are too close to the solution to see the issues when students do not demonstrate outcomes and seniors are not particularly skilled at assessing how well they demonstrate outcomes. Therefore, programs are encouraged to use these
types of surveys to provide additional data points on outcomes and key data for program educational objectives, but not as the sole source of data collection.

How many assessment methods are enough? Multiple means more than one. Is that one assessment method beyond using surveys? The Civil Engineering program at the University of Texas at Tyler (UT Tyler) decided to use all of the assessment methods previously mentioned to ensure multiple view points for each outcome. Since the program leadership was comfortable in the use of these assessment methods during ABET evaluations as Program Evaluator’s (PEV) (one is currently an ABET Commissioner), they decided to implement all with a focus on limiting the faculty assessment requirements long term. The key to accomplishing the limitation of faculty time in assessment is direct assessments. The answer to the question posed depends on the program and whether they are getting multiple view points from the assessment methods they have chosen. If not, then they will need to possibly add another or improve the data collection from the assessment methods they are currently using. Some outcomes will require particular assessment methods to gather the desired data.

1.2 Why Direct Measures?

Direct measures are quickly becoming the assessment methods of choice because the data most directly relates to the outcome. Many of the best practices from ABET\textsuperscript{13} point toward direct measures. The most recent Southern Association of Colleges and Schools (SACS) visit further highlighted the importance of external exams because they provide national comparison of key indicators for demonstration of outcomes. However, the sole use of external exams is not possible since the current available external exams for engineers (FE) do not provide demonstration of each outcome or demonstration by each student since the exams results are program averages. Therefore, programs like UT Tyler are using additional direct measures such as the internal exams to provide results for each student, but again they might not provide coverage of every outcome since they are conducted early within the program at the end of the sophomore and/or junior years. The capstone rubrics at UT Tyler provide direct assessment of almost every outcome, but the results are for team projects. Therefore, the use of embedded indicators (direct measures) provide the best direct measurement of student demonstration of each outcome by collecting results for the best student, the average student, and the worst student performance leading to a better collection of data representing students demonstration of each outcome.

2.0 Direct Assessment Process at UT Tyler

Generally the assessment process is a lonely job performed by one or maybe two faculty in a program. Many faculty do not know who is doing the work and they really do not care. The person in charge (even if the chair) begs peers to submit their assigned input in a timely fashion and then tries to assess the data (if it is assessed at all) that is collected -- data that is not always complete and organized in a useful way. The UT Tyler program developed an assessment process that not only provides a robust assessment of the program, but also engages every member of the faculty in the accreditation process.
performing “Course Focused” assignments that naturally align with normal daily teaching (and research) activities.

The Civil Engineering Program at UT Tyler faced a difficult challenge and a very unique opportunity. The program was established in 2005, but no assessment processes had been put into place when the current chair arrived in January 2007. The self study was scheduled to be mailed the end of June 2008 with the site visit the fall of 2008. How to establish a robust assessment program in only 18 months, produce a detailed self study and organize the site visit to ensure ABET accreditation, while in-processing two new faculty, hiring the last authorized faculty member, and teaching all of the senior courses for the first time within the ABET record year?

The new department chair who has also been an ABET program evaluator decided to establish a robust assessment program that included multiple assessment methods as well as all members of the faculty providing input and assessment of the collected data. Unfortunately the new faculty team being assembled was new untenured, tenured track faculty. They knew nothing about assessment. Fortunately the new faculty team being assembled was new untenured, tenured track faculty. They knew nothing about assessment. The chair considered and used these two contradictory facts as he developed the assessment process and developed assessment skills within the new faculty.

The following assessment methods chosen to ensure adequate data collection and assessment of each outcome from multiple viewpoints were:

- Embedded Indicators
- External Exams
- Internal Exams
- Capstones
- Surveys
- Course Assessments
- Industrial or External Advisory Committee

The sheer number of assessment methods may appear daunting for a new program to consider employing, but the chair had seen good applications of each during his time as a PEV and as a faculty member at the United States Military Academy. Since he was also teaching two to three courses each semester during the preparation for the visit, the chair performed each task before demonstrating their use to the faculty team and within the assessment process while also guiding their efforts within assessment tasks that mirror normal teaching tasks.

Obviously, a plan had to be established as to who accomplished what. The chair desired to minimize the requirements of the faculty within the process since they are untenured, tenured track faculty who must establish a research program while also including them within each assessment activity. The ultimate goal is a faculty team where assessment is second nature and it is obvious that the faculty understand the assessment process being employed within the program. The chair focused on the faculty completing activities...
directly associated with the teaching and assessing of activities within the courses they taught – something they should be naturally doing, but even many seasoned professors do not. The direct measures employed were embedded indicators as the primary assessment methods and external and internal exams and capstones to provide additional view points.

2.1 Embedded Indicators

First, the chair required each faculty member to collect data on each course assignment for their spring 2007 courses just as they would need to do for ABET Criteria 5 during their record year: Fall 2007 – Spring 2008. The exercise completed by all including the chair quickly highlighted the pitfalls and personal routines needed to collect the assignment, the solution, the grading rubric, the assessment of the assignment (both for the course and the outcome if assigned as an embedded indicator for an outcome), and a sample of the best, average, and worst student work. The entire faculty (including the two new faculty that arrived in early August 2007) assessed the spring 2007 results and established protocols that the entire faculty team used for the record year. The process led faculty towards more focused assessment methodologies overall once they started using thoughtful assessments of each of their own assignments. The goal was to determine whether the assignment met the desired level of student performance against course objectives and where to make adjustments within the course to improve student performance during the next offering of the course. The same process is used with assessing a portion or the entire assignment or exam against a program outcome. The ABET PEV during the Fall 2008 visit noted that the course notebooks were about the most complete and consistent he had seen across a program.

Second, the chair took the nine program outcomes that included the new civil engineering program criteria outcomes that would be assessed for the first time in the fall 2008 visits and broke them into manageable sized program outcomes that included an action verb to determine the level of demonstration by students. Example: Outcome 1 (Can apply knowledge of traditional mathematics, science, and engineering skills, and use modern engineering tools to solve problems) became outcome 1a-d (1a. Can apply knowledge of traditional mathematics to solve problems; 1b. Can apply knowledge of traditional science (calculus-based physics, chemistry, additional science) to solve problems; 1c. Can apply knowledge of traditional engineering skills to solve problems; 1d. Can use modern engineering tools to solve problems). This ensured that as an embedded indicator (a portion of an exam or homework or a portion of the capstone experience) was being assessed, it was clear as to which part of the outcome was being considered. Similar efforts for focused assessment of the outcome can be observed within the ASCE BOK II document where there are now 24 outcomes separating some of the ABET 3 a-k outcomes into more focused outcomes.

Third, the chair assembled the faculty available prior to the start of the spring 2007 semester and listed all of the outcomes on a whiteboard, and then asked each to assess their courses that they would teach and determine whether their course could provide demonstration of the outcome. At the end of the spring semester, the embedded indicators collected within the outcomes notebooks were assessed as to whether the chosen embedded indicator was a good indicator of student demonstration of the outcome.
Unfortunately, some of the data collected was poor. However, the review of the available data in the trial outcome notebooks as well as trial course notebooks showed that some of the assignments were great indicators of demonstration of an outcome while others had only a portion of the assignment or exam that directly related to an outcome. This allowed for the entire faculty team to assess the quality and determine procedures to improve the collection of data (similar to the results on the course notebooks). The faculty (including the two new faculty) once again went through the list of outcomes and listed courses beside each outcome where an embedded indicator was possible. This process allowed the faculty to determine if they felt they could collect an embedded indicator rather than being directed by the chair or program assessment guru. Since no minimum number of indicators was listed for each embedded indicator, some outcomes had over 20 embedded indicators collected during the record year while others had only three. Additionally, some courses provided over 20 embedded indicators while others only provided one or two.

In May 2008, two faculty were assigned (not the same pair) to each outcome. Their mission was to assess the embedded indicators collected to determine whether they were a good demonstration of the outcome, whether the embedded indicator should be kept for the next year, whether it should be adjusted to improve its demonstration of the outcome, or whether it should be removed or replaced. The goal was six embedded indicators for each outcome for the next year of embedded indicator data collection while evenly spreading the collection across all of the courses within the program. The ABET PEV in Fall 2008 did note that many embedded indicators were not good demonstrations of the outcome, but once the assessment of the process was provided (not really sure how to provide this in the self-study), he noted that there were enough (how many required?) good indicators for each outcome and that the program was demonstrating assessment of the assessment process! Something he had not seen before. Currently, the annual process of assessing the embedded indicators collected for each outcome has narrowed the set of excellent demonstrations of each outcome down to 3-4 as well as limited the number of embedded indicators collected from each course. Additionally, the assessment skills of the six faculty are steadily improving each year. The larger the program the more faculty to assign to an outcome and the fewer number of outcomes that each faculty member must assess and the fewer number of embedded indicators to collect within a course.

2.2 Capstones

Since the senior design (capstone) is already being taught, assignments being assessed, and it is the one course that should bring all of the students skills to bear on a given multi-discipline problem, the faculty decided that the grading rubric should be adjusted to ensure that each outcome to include professional outcomes is being assessed within the capstone course.21,22,23,24 There were over 17,000 search results for capstones being used to provide direct results.25 Each capstone project requires coverage of the seven traditional sub-disciplines of civil engineering. Therefore, the capstone course is bigger than one faculty to assess properly, so each faculty member assesses portions of the submissions that fall into their areas of expertise. The entire faculty team has been involved in the adjustment of the focus within the capstone through the course.
assessment process mentioned earlier (another course focused activity with direct ties to the embedded indicators) to ensure a proper balance of effort within the capstone course aligned with the level of demonstration established by the action verb within the outcome.

2.3 Fundamentals of Engineering (FE) Exam

The rest of the assessment process (external and internal exams, surveys, external advisory committee) basically falls on the chair’s shoulders with input from the faculty when they review the results. The other two direct measures used by the program are the external and internal exams. Since the external results can be used as a direct measure of the program’s demonstration of an outcome, the number of students each year who actually take the exam is an issue. (Over 129,000 search results for using the FE exam as a direct measure to include a white paper from NCEES)\textsuperscript{26} The UT Tyler program has the requirement that every student must take the FE Exam to graduate. The students are directed to take the civil engineering portion in the afternoon so that the results can be used as valid direct assessment. The chair receives the overall program results and since every student takes the exam, the results can be used as an external exam indicator for some of the outcomes. If a majority of the students do not take the exam, the program would need to demonstrate how those that did take the exam are an accurate representation of the student body demographics before using the results.

The results are presented to the faculty as one data point to possibly direct changes to course content within the program. Currently the program has used the results to slightly modify content and the current results demonstrate the students are performing at or above the national level on the afternoon portion focused on civil engineering topics.

2.4 Gateway Exams

The internal exams are used by some programs to allow admittance into the program or to continue into the junior year of the program (nursing and some engineering programs).\textsuperscript{4} The civil engineering program at UT Tyler decided to apply the concept, but instead of passage of the exam being required to continue in the program, the faculty use the results to provide input into the assessment process without needing to wait for the FE results (taken senior year on topics that could be covered sophomore year) for the given group. Additionally, so many of the program’s students are transfers, so FE results for sophomore level courses could be obscured by where the student actually took the course. The UT Tyler civil engineering program decided to not only use the results to determine adjustments to the curriculum and assignments, but also to assign those students performing 10 percentage points below the average in any subject area a homework on that subject that must be completed within the first week of academics in the fall. The faculty developed the homework assignment to best review the key concepts tested on the exam. The goal is to improve skills and retention of skills before the students are too far into the fall semester. The faculty were asked to develop 10 multiple choice exam questions from their engineering science (statics, mechanics of materials, etc.) and civil engineering specific courses during the sophomore and junior years. The chair assembled the Gateway exam that is administered late in the spring semester. Each year the faculty
are asked to review the performance of students on each portion of the exam and
determine adjustments to courses, exam questions, and/or assigned homework. Currently
the student’s performance mirrors closely with the performance level on the FE exam.

3.0 Where the Program Is Now

Not every student passes the FE exam on the first try (7 out of 10 in 2008, 7 out of 11 in
2009, 9 out of 14 in 2009, 17 out of 22 in 2010), but most eventually pass the exam
because of personal motivation developed through the program to achieve the first major
step towards becoming a Professional Engineer. However, the program currently has FE
results that show the students performing at or above the national average for all CE
topics areas: surveying, environmental engineering, structural analysis, transportation,
soil mechanics and foundations, hydraulics and hydraulic systems, structural design,
construction management, and materials. Since the program does not teach toward the
exam nor has had anyone who has worked on development of the exam, these results
demonstrate course assignments and exams that not only meet course objectives but lead
to students meeting program outcomes.

The Gateway exam results are currently improving and mirroring the FE exam results
through the development of better worded questions that focus on the key concepts from
each course. Poor performing students receive homework based on the exam results on
the first day of class during the fall semester that supports their development of better
skills. The consistent improvement in FE results to be currently performing at or above
the national average for all CE topics areas provides input that the resulting course
adjustments and the developmental homework is working.

The faculty provides the students the 10%, 35%, 65%, and 100% capstone grading
rubrics to ensure they experience application of skills surrounding each outcome while
completing the two-semester senior design. Proper allocation of the points on rubrics
provided to the students drive each team to wrestle with each outcome during the
development of their design.

The assessment of embedded indicator outcomes notebooks at the end of each spring
semester has decreased the collection of embedded indicators to 4 or 5 per outcome
spread across sophomore to senior courses minus the capstone that feeds into its own
direct measure above.

4.0 Lessons Learned

The “Course Focused” assessment process ensures each member of the faculty is an
integral part of program assessment through the use of direct measures – embedded
indicators and the capstone. Since the faculty are already assessing each assignment
within their course for the course assessment process (something each faculty should be
doing for any course they teach), the time spent to assess assignments or a portion of the
assignment as an embedded indicator is a minimal addition. Only those assignments
being used as an embedded indicator require a collection of student work (the best, the
average, the worst). The constant assessment of the course content and program outcomes by the faculty results in assessment being part of the fabric of the program, not something that is done every six years when ABET comes to town. The additional “course focused” requirements to be completed by the faculty for other assessment methods, capstone and internal exams, are processes that they should be already doing as a part of the senior design process and the exam questions should not take too much additional time to generate based on recently assessing the course they have just taught that year. The assessment of the embedded indicators collected each year improves the assessment process through the skills the faculty develop. The chair does the major lifting on some of the assessment processes and the faculty participates through an annual program assessment presentation during department meetings where they provide input as to recommended program changes based on assessment of the consolidated data from all of the assessment methods. The program assessment process that has emerged is “Course Focused” and has limited the number of embedded indicators collected in each course and for each outcome and ensures the time spent by faculty on assessment surrounds the normal activities they are already performing as part of their normal teaching activities.

The assessment of collected embedded indicators each year by the faculty have greatly improved their critical eye and improved the overall quality and focus of the embedded indicators collected. More importantly, the process clearly displays the faculty closing the loop through not only the improvement of student skills with changes to courses and the curriculum, but also changes and improvement to the assessment process as well. Normally many programs show a drop in assessment between ABET visits, but tying the process for the faculty to a “Course Focused” effort ensures that the faculty remain engaged each and every year. Additionally, the faculty have noted that the more critical assessment of student learning has improved their assessment of research work and improved the quality of research papers and reports.

Through annual assessment of the entire process, the adjustment as to where a direct measure is collected is simple to execute since the program does not naturally select the best course to collect from on the first try. The faculty understand that the team assessing a given outcome are focused only on how best to meet that outcome. Occasionally, a course can be completely overloaded in collecting embedded indicators. The annual assessment of the collected direct measures by the faculty brings not only the entire team together to perform assessment of the program through the embedded indicators, but their collective experiences for that year are used to determine whether the program is demonstrating the outcomes and where is the best location to collect each embedded indicator.

A much clearer picture of student demonstration of an outcome is possible through the use of all assessment methods, especially direct measures. If the direct measures that the faculty must provide data for are tied directly to their teaching roles, then the time required for collection and assessment against an outcome is a minimal increase over normal course operations and management.
5.0 Conclusions

Direct measures are only one of many assessment measures available to programs to collect the necessary data to allow for assessment of whether the program can demonstrate that their students are meeting program outcomes. Through a “Course Focused” process, the time spent by faculty collecting and assessing the data collected is minimized since each faculty member should already be assessing student performance on course assignments and exams. If the right assignments and exams can be located that directly demonstrate a given outcome, then the collection of that data and its assessment is only an small extension of the assignment assessment process that should already be taking place. The faculty at the University of Texas at Tyler have noted that the course assessment process is taking about one-third as much time for courses they have taught before and only half of the time for new courses in comparison with the first time they completed one. The requirement to assess each assignment was onerous at first, but clearly is something that each faculty should do, but many who do never write anything down. If nothing is written down, the same faculty member may make the same mistakes the next time the course is taught, especially if the course is not taught each year. Additionally, the written assessment process clearly improves the performance of a newly assigned faculty member to the course. The assignment assessment process also directly provides for review of previously recommended changes to the course as well as data for recommended changes for the next offering.

Direct measures, if crafted properly, combined with properly led and trained faculty can result in an efficient assessment process tied directly to what faculty should already be doing within the planning, operating, and assessing of the courses they teach. A consistent well-designed process leads to inherent time efficiencies that are sustainable, while completing program assessment needed for accreditation. The program completed its second annual assessment after the ABET visit and the faculty realize that it is taking less time to perform robust assessment because it is a necessary daily part of the fabric of the program and with improvement of their assessment skills, the entire process takes less time each cycle.

References


2 Texas Assessment of Knowledge and Skills (TAKS), (http://www.tea.state.tx.us/index3.aspx?id=3839&menu_id=793), Assessed 13 January 2011.


9Industrial Advisory Committees, 27,700 hits on Google, University web sites for their Industrial Advisory Committees, (http://www.abet.uprm.edu/Symposium/Abstracts/General_Assessment_Processes/Armstrong_ProgramAbstract_9.pdf), Assessed 13 January 2011.


12ABET Webinar, “How to Develop a Survey,” 9 November 2010.


14Nalbone, T., Referred to the UT Tyler program assessment process as “Course Focused” during development of an abstract for a proposal for an ABET presentation, Tyler, TX, 22 Nov 2010.


16Course Assessment Document for CENG 4341, Fall 2007 and 2008 and Spring 2009 and 2010, University of Texas at Tyler.


