Strategies for Assessing Course-Specific Outcomes

David G. Meyer <meyer@purdue.edu>
Purdue University
School of Electrical & Computer Engineering
West Lafayette, Indiana 47907

Abstract

A proven method for satisfying the Accreditation Board for Engineering and Technology (ABET) "Criterion 3" requirements is the formulation of outcomes specific to "core" courses in a curriculum, which are tied to the program outcomes. The challenges of assessing such course-specific outcomes are described in this paper, with a focus on practical realities and lessons learned through seven trials in two different computer engineering courses spanning nearly five years. Issues addressed include formulation of outcomes, choice of evaluation instruments, static vs. dynamic assessment thresholds, instructor overhead, maintaining consistency with prior grading practices, and remediation strategies. Outcome demonstration success rate data are presented for representative trials.

Introduction

ABET, Inc., the recognized accreditor for college and university programs in applied science, computing, engineering, and technology, is a federation of 28 professional and technical societies representing these fields. ABET has provided leadership and quality assurance in higher education for over 70 years. The criteria for accrediting engineering programs[1] published by ABET are intended to assure quality and to foster the systematic pursuit of improvement in the quality of engineering education that satisfies the needs of constituencies in a dynamic and competitive environment. Although institutions may use different terminology, for purposes of Criterion 3, program outcomes are statements that describe what students are expected to know and be able to do by the time of graduation. These relate to the skills, knowledge, and behaviors that student acquire in their matriculation through the program. For the purpose of accreditation, engineering programs must demonstrate that their students attain:

(a) an ability to apply knowledge of mathematics, science, and engineering;
(b) an ability to design and conduct experiments, as well as to analyze and interpret data;
(c) an ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability;
(d) an ability to function on multi-disciplinary teams;
(e) an ability to identify, formulate, and solve engineering problems;
(f) an understanding of professional and ethical responsibility;
(g) an ability to communicate effectively;
(h) the broad education necessary to understand the impact of engineering solutions in a global, economic, environmental, and societal context;
(i) a recognition of the need for, and an ability to engage in life-long learning;
(j) a knowledge of contemporary issues; and
(k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

According to supporting documentation provided by ABET [2], "…engineering programs must have in place an appropriate assessment process that produces documented results that demonstrate that students have achieved each and every item listed in (a) through (k). It is expected that all students will demonstrate achievement of every item listed in (a) through (k). Programs must show, even by appropriate sampling, that there is convincing evidence to assume that all students by the time they have graduated have demonstrated achievement, to a level acceptable to the program, of every item listed in (a) through (k). However, it is not necessary for evidence to be provided for each and every student." Further, "student self-assessment, opinion surveys, and course grades are not, by themselves or collectively, acceptable methods for documenting achievement of outcomes."

Atman *et al.*[3] provide an overview of the methodologies that are available for use in assessing undergraduate engineering programs, along with research questions associated with these methodologies that are currently outstanding. To measure how well engineering students can apply classroom knowledge and skills to realistic design problems, *authentic assessment* and *performance-based assessment* methods can be used. The key to authentic assessment, according to the authors, is to "create a context in which the student can individually or collaboratively demonstrate an ability to apply a well-developed problem-solving strategy" which might involve "problem definition, gathering relevant information, generating solution alternatives, choosing the optimum solution given implicit and explicit constraints, assessing and improving the proposed solution, and effectively reporting results…"

Felder and Brent[4] provide guidance on the formulation of course learning objectives and assessment methods that address outcomes (a) through (k) of Criterion 3. They also describe a series of possible "program-level" and "course-level" assessment tools. Jenkins and Kramlich[5] provide an interesting discussion of assessment methods that have worked and have not worked in their experiences at the University of Washington. Numerous other papers in the literature address the issue of outcome assessment in light of current accreditation requirements.

The initial basis for the outcome assessment strategy adopted by the Purdue University College of Engineering for meeting ABET "Criterion 3" (or, "EC2000") requirements was formulation of a series of course-specific learning outcomes for the core and breadth courses required in the Electrical and Computer Engineering (ECE) curriculum, and subsequently mapping those learning outcomes to the program objectives. No attempt was made at the outset, however, to specify how these course-specific outcomes were to be assessed (and, over five years later, there remains a wide variability in outcome assessment strategies employed by various professors). This prompted the author to embark on an empirical study of course-specific outcome assessment strategies, the results of which are summarized in this paper.

The basic questions addressed in this study include:

1. formulation of content-specific learning outcomes that can be consistently and quantitatively assessed;
2. formulation of effective outcome assessment instruments along with mechanisms to determine outcome demonstration thresholds;
3. formulation of grading strategies that incorporate outcome demonstration thresholds yet produce results consistent with prior (accepted) grading practices; and
4. formulation of outcome remediation strategies that are both fair and efficient.

The hope is that through sharing what has worked and, in particular, what has *not* worked, others working toward the same goal of improving applicable outcome assessment strategies might benefit. A related study on outcome assessment strategies for capstone design[6] delves further into issues related specifically to assessment of design-intensive courses.

Outcome Formulation

Two courses for which the author is responsible serve as the basis for the development of the outcome assessment strategies described here:

- *Introduction to Digital System Design* (ECE 270), a sophomore level lecture/lab course on digital logic and system design with an annual enrollment of approximately 500 students.
- *Microprocessor System Design and Interfacing* (ECE 362), a junior-level lecture/lab course on embedded microcontroller systems with an annual enrollment of approximately 300 students.

For the sophomore-level course in this study, a set of ten outcomes was initially formulated (the mapping of these course-specific outcomes to the "Criterion 3" outcomes specified by ABET, listed previously, are in parentheses):

1. an ability to analyze static and dynamic behavior of digital circuits (a,b);
2. an ability to map and minimize Boolean functions as well as represent them in various standard forms (a,e);
3. an ability to design and implement combinational logic circuits (a,c,e,k);
4. an ability to use a hardware description language to specify a digital circuit (a,c,e,k);
5. an understanding of various combinational "building blocks" such as decoders, multiplexers, and encoders (a,c,e,k);
6. an ability to design and implement arithmetic logic circuits (a,c,e,k);
7. an understanding of the behavior exhibited by latches and flip-flops (a);
8. an ability to design and implement sequential circuits (a,c,e,k);
9. an understanding of various sequential "building blocks" such as counters and shift registers (a,c,e,k); and
10. an ability to design and implement a simple computer (a,c,e,k).

At the moment (and for next few semesters), this set of ten outcomes seemed reasonable: it represented a relatively equal division of the course content into key outcomes that could be readily tested on traditional hourly examinations. The goal was to utilize a definitive "exam module" to assess students' success in demonstrating a given outcome. The reality of creating/grading that many examination modules (plus providing opportunities for remediation on each of these) prompted a "compaction" of the list to the set of six outcomes currently employed:

1. an ability to analyze static and dynamic behavior of digital circuits (a,b);
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement them as combinational logic circuits (a,c,e,k);
3. an ability to use a hardware description language to specify combinational logic circuits, including various "building blocks" such as decoders, multiplexers, encoders, and tri-state buffers (a,c,e,k);
4. an ability to design and implement arithmetic logic circuits (a,c,e,k);
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify them (a,b,c,e,k); and

6. an ability to design and implement a simple computer based on combinational and sequential building blocks (a,c,e,k).

The current set allows the outcomes to be tested in pairs using three two-hour evening exam sessions (which is fairly standard for design-oriented courses at Purdue University). Alternately, each outcome could be tested separately using a series of hourly (in class) exams. Both strategies have been utilized during the course of this study. Sample exams used to assess these outcomes are posted on the ECE 270 course web site[7].

For the junior-level course in the series, the major categories of topics were fewer in number than in the sophomore-level course, yielding a relatively compact set of four learning outcomes (that were tested using a series of four hourly exams):

1. an ability to design and implement a simple computer (a,c,e,k);
2. an ability to write programs for a computer in assembly language (e,k);
3. an ability to interface a microprocessor to various devices (a,c,e,k); and
4. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller (j,k).

Subsequent curriculum changes (specifically, increased emphasis on embedded system design) prompted revision of the course learning outcomes as follows:

1. an ability to write programs for a computer in assembly language (e,k);
2. an ability to interface a microprocessor to various devices (a,c,e,k);
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller (j,k); and
4. an ability to design and implement a microcontroller-based system (a,c,e,j,k).

Currently, in-lab "practical exams" are used to assess outcomes 1-3, while an embedded system design "mini-project" (implementation of a turn-key microcontroller-based device) is used to assess outcome 4. Sample exams used to assess outcomes 1-3 along with a description of the design mini-projected are posted on the ECE 362 course web site[8].

First Accreditation Visit Under EC2000

The initial set of learning outcomes and strategies for assessment were barely in place when the first accreditation visit occurred, Fall 2001. One thing became abundantly clear in the discussions with the Visitors that ensued: outcome demonstration success could not simply be based on whether a student received a *passing grade* for a course. Effectively, the Visitors tacitly informed us: "We want to see evidence of *failing grades* assigned for cases in which students [who would otherwise be passing but] failed to successfully demonstrate one or more course outcomes." In other words, the business of assigning course grades could no longer be conducted as usual.

Several major issues emerged in the formulation of a tractable outcome assessment strategy. First, and perhaps foremost, is the issue of balance. On one side of the scale, fairness to students must be considered, both in terms of providing students with a sufficient number of opportunities to demonstrate each course outcome as well as ensuring that students' *own* work is used as the basis for outcome demonstration success. The only reliable way to ensure the latter (in courses that are primarily "content" oriented) is to test outcome demonstration success using proctored quizzes or exams. Another aspect of the balance issue is overhead for the instructor, both in terms of keeping the incremental workload associated with outcome assessment and tracking to a minimum as well as keeping the outcome assessment process "contained" within a given term/semester.

The next issue that needs to be addressed in the formulation of a tractable outcome assessment strategy is choice of evaluation instruments. Possibilities include exams (whole or question subsets), quizzes (written/oral), homework assignments, labs, and papers/presentations. For lower-division "content" courses, proctored exams/quizzes have proven to be the very effective; while for senior design projects, papers and presentations are generally the preferred evaluation instruments.

One of the most challenging issues encountered was determination of outcome demonstration "passing thresholds". Choices include *static* thresholds (plus what absolute value to choose) or *dynamic* thresholds (plus the algorithm used to "adjust" them). A further complication is making the outcome demonstration thresholds chosen consistent with traditional grading cutoff (A-B-C-D-F) thresholds: assigning course grades consistent with proven prior practice, yet reflecting meaningful application of outcome assessment thresholds. Yet another challenge is making a concerted effort to achieve consistency semester-to-semester, professor-to-professor, course-to-course, etc. Suddenly, what initially appeared to be simple had become complex.

The strategies attempted will be described using the sophomore-level course (*Introduction to Digital System Design*) as an illustrative example. Parallel studies conducted in the post-requisite junior-level course (*Microprocessor System Design and Interfacing*) will also be cited.

Initial Strategies

Armed with an understanding of the basic issues involved, several "refinements" of outcome assessment strategies employed in the sophomore/junior courses are chronicled[9]. What "went wrong" with each of the prior attempts will be described, providing the cumulative rationale for the strategy currently employed. The preliminary strategies tried, in chronological order, include the following:

1. fixed passing threshold (60%) on weighted sum of (selected) lab, homework, and exam scores.
2. fixed passing threshold (60%) on primary assessment, remediation homework, and final assessment.
3. fixed passing threshold (60%) on primary assessment, final assessment, and remediation homework.
4. fixed passing threshold (60%) on primary and final assessments; use of "conditional failure" grade for those who would otherwise be passing, with opportunity to take remediation assessment the following semester.

Trial 1

The initial strategy employed for outcome assessment was to calculate a (weighted) sum of scores on selected, pertinent graded items (homework, lab experiments, exam scores) and apply a fixed threshold (here, 60%) to determine whether or not the outcome was successfully demonstrated. It only took one semester to determine the ineffectiveness of this strategy: basically, *everyone* "passed", primarily because it was impossible to ensure that students' *own* work was being evaluated on labs and homework. Further, the remediation strategy was ill-defined: graded items below the prescribed passing threshold (60%) had to be "made up" on an *ad hoc*, one-on-one basis.

Trial 2

The first refinement to the initial strategy was to define three distinct opportunities for demonstrating a given outcome:

1. successful completion of a primary outcome assessment exam
2. successful completion of a remediation homework problem set
3. successful completion of a final outcome assessment exam

A fixed threshold (60%) was applied to each of the evaluation instruments.

Here, students who failed the initial ("primary") outcome assessment exam were given the opportunity to complete a remediation homework set. While the remediation strategy was now clearly defined (compared with Trial 1), the remediation homework provided very little "filtering" – the only students who did not successfully demonstrate an outcome were the ones who "forgot" to turn in the homework. Further, there was significant (excessive) overhead associated with processing the remediation homework. Finally, it was impossible to ensure that students' own work was being evaluated on the remediation homework sets.

Trial 3

In an attempt to reduce the overhead associated with the remediation homework, the next refinement attempted was to "swap" the order of the remediation homework and the final assessment exam, based on the theory that there should be less remediation homework since it would be preceded by an additional "exam" attempt. The same fixed threshold (60%) was applied to determine whether or not an outcome had been successfully demonstrated. Even the casual reader could probably guess the major limitation here: that this scheme *only* works if the final assessment exam is (or, in the case of author's institution, *happens* to be) scheduled *early* during "finals week" (and also that it assumes students are willing to stay in town after the final exam in order to complete the remediation homework when and where necessary). Obviously, this attempted refinement ("mutation" might be a better word) resulted in excessive finals week overhead – grading final exams *plus* remediation homework. Further, it was still impossible to ensure that students' own work was being evaluated on the remediation homework.

Trial 4

The third refinement was based on an arcane, rarely used grading option available at the author's institution: the so-called "E" grade, which someone many years ago decided could be used to designate "conditional failure" (meaning that the student was "otherwise passing" but deficient in a specific area which, once satisfied, would permit the grade to be improved to passing). At first glance, this mechanism appeared to be the perfect vehicle to use for distinguishing "normal" failures from "outcome deficiency" failures. The big plus was the ability to eliminate the remediation homework as one of the outcome assessment instruments, thus reducing the outcome demonstration attempts to (two) proctored exam situations. To provide a "third attempt", students were allowed to take a remediation exam over each unsatisfied outcome the following semester to improve their "E" grade.

What sounded good at the time once again became wrought with excessive overhead: the lesson learned here was to keep the chosen assessment strategy *self-contained* within a given semester/term. Another issue that had not emerged previously was that the fixed threshold for passing (60%) became a significant factor. With the easy-to-pass remediation homework eliminated from the picture, the difficulty of writing exams that produced a predictable mean/distribution became a major challenge. Basically, there are too many factors "beyond an instructor's control" – particularly timing of exams relative to others students might be taking.

It is instructive to examine the data relative to this trial. Table 1 lists the primary assessment exam statistics. Upon completion of the primary assessments, only 53 out of 274 students had successfully demonstrated all ten course outcomes (less than 20%). About 26% (72 out of 274) had passed less than half the outcomes at this point. A serious drawback, then, was additional pressure placed on students to successfully demonstrate a significant number of outcomes on the final assessment, given that only two opportunities were provided to demonstrate a given outcome.

Table 2 illustrates the cumulative success rate achieved upon completion of the final assessment exam. Nearly 82% of the students 224 out of 274) were able to successfully demonstrate all ten course outcomes on either the primary or the final assessment. Of those who were not successful, half (25 out 50) had only failed to demonstrate a single outcome; the maximum number of outcomes failed was six (accomplished by 1% of the total number of students).

An interesting adjunct of Trial 4 was providing students the opportunity to not only use the final assessment for remediation of any number of (up to *all ten*) course outcomes, but also use it improve their score (relative to the primary assessment) on any (up to *all ten*) outcomes. Perhaps obviously, many students took advantage of this rather generous offer (intended as an incentive to "put it all together at the end" and master the course outcomes) and improved their course grades dramatically. Too many others, however, did not, and despite the offer still managed somehow to fail at least one outcome – in total, 42 out of 274 students (15%) received the fateful grade of "E" (conditional failure). If Bilbo (of *Lord of the Rings* fame) had been the instructor, he might have expressed the following sentiment about this "Wheel of Fortune" anomaly: *"I helped the students in the top half of the class more than I anticipated, and helped the students in the bottom half of the class less than I had hoped."* While this sounds bad for the students, it was even worse for the instructor – recall the opportunity of a remediation exam (third attempt) offered the *following semester*. The overhead associated with handling this (scheduling remediation exams and getting students to show up for them) was, in a word, excruciating.

Before leaving this trial, there is one additional feature of the data depicted in Tables 1 and 2 worth considering. Note that, based on the primary assessments, a number of students failed to successfully demonstrate outcomes 1-4. It would be natural to fear that, because students had to wait until the final assessment to demonstrate these outcomes (when the material was no longer "fresh" in their minds), a significant percentage would *again* fail to demonstrate them. This is clearly not what happened. As evidenced by the data in Table 2, only 1-2% of the students failed to demonstrate Outcomes 1-4 on the final assessment. This is perhaps due to the "soak time" of the material, as well as its cumulative nature. The data debunk the argument that remediation opportunities, to be successful, must be "immediate."

Table 1. Trial 4 Primary Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Avg. Score	Passed	Failed
1. an ability to analyze static and dynamic behavior of digital circuits	71.5%	75.2%	24.8%
2. an ability to map and minimize Boolean functions as well as represent them in various standard forms	63.4%	57.7%	42.3%
3. an ability to design and implement combinational logic circuits	75.8%	84.3%	15.7%
4. an ability to use a hardware description language to specify a digital circuit	65.0%	68.6%	31.4%
5. an understanding of various combinational “building blocks” (e.g., decoders, multiplexers, encoders)	61.7%	55.8%	44.2%
6. an ability to design and implement arithmetic logic circuits	62.0%	55.8%	44.2%
7. an understanding of the behavior exhibited by latches and flip-flops	80.5%	89.4%	10.6%
8. an ability to design and implement sequential circuits	65.1%	65.0%	35.0%
9. an ability to analyze, design, and implement sequential circuits and use a hardware description language	54.7%	46.4%	53.6%
10. an ability to design and implement a simple computer	54.5%	41.6%	58.4%

Table 2. Trial 4 Final Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Successful
1. an ability to analyze static and dynamic behavior of digital circuits	98.5%
2. an ability to map and minimize Boolean functions as well as represent them in various standard forms	99.3%
3. an ability to design and implement combinational logic circuits	99.6%
4. an ability to use a hardware description language to specify a digital circuit	99.6%
5. an understanding of various combinational “building blocks” (e.g., decoders, multiplexers, encoders)	97.8%
6. an ability to design and implement arithmetic logic circuits	94.5%
7. an understanding of the behavior exhibited by latches and flip-flops	98.2%
8. an ability to design and implement sequential circuits	93.8%
9. an ability to analyze, design, and implement sequential circuits and use a hardware description language	85.4%
10. an ability to design and implement a simple computer	91.2%

Seasoned Strategies

For the next set of refinements, a dynamic threshold was utilized to determine successful outcome demonstration. To determine an appropriate threshold choice, the author analyzed outcome demonstration data spanning several semesters. A relatively simple scheme that appeared to produce consistent results was to use EXAM MEAN – STANDARD DEVIATION as the threshold (limited to the range of 40% to 60%). Another refinement was to require students to successfully demonstrate *at least half* of the course outcomes on the primary assessment exams in order to qualify for a passing grade (i.e., taking the final assessment). Thus, students would be limited to re-taking (at most) half of the course outcomes on the final assessment for the purpose of remediation and/or improving their grade. In theory, limiting students to retaking at most half of the outcomes on the final assessment exam was viewed as the answer to several pressing issues: (a) reducing the overhead associated with the final assessment exam, (b) making the final assessment of each outcome more extensive than was possible previously (where students were allowed to retake all the outcomes, thus necessitating very short final exam modules for each outcome), and (c) helping curb the grade inflation associated with outcome score improvement (replacing the primary outcome assessment score with the score received for that outcome on the final).

Trial 5

The first trial utilizing the dynamic threshold described above (Spring 2004) produced much higher "initial pass rates" relative to previous trials (note that the number of outcomes had been compacted to six as of this trial, even though the body of material covered in the course was basically the same). Based on the primary assessments, 149 out of 277 students (53.8%) passed all six course outcomes. Only 26 students (9.4%) failed the course outright based on an inability to successfully demonstrate at least half of the outcomes on the primary assessments. After the final assessment, 242 out of 277 (87.4%) had passed all six outcomes. These results seemed to be more "reasonable" than the Trial 4 data detailed previously. A nagging problem that continued in Trial 5, despite a reduction in the number of outcomes that could be re-taken on the final by 50%, was an inordinate amount of "grade inflation" associated with the "outcome improvement" provision (the maximum of the primary and final assessments was still being used as the effective score for a given outcome). No significant reduction in post-final "grade inflation" was realized due to the change in strategy. Also, there was still a fair amount of overhead associated with the final assessment, since *all* students were given the opportunity to re-take as many as 50% of the total number of outcomes.

Trial 6

In an attempt to mitigate the post-final grade inflation associated with the "outcome improvement" provision of previous trials, the final assessment was "re-purposed" in Trial 6 as an instrument for outcome remediation *only*. Here, the primary assessment exam became the sole contributor to the course grade determination; the final assessment exam for a given outcome was only taken by those who did not pass the primary assessment exam for that outcome, and all that mattered was whether or not the dynamic threshold had been achieved. The post-final grade inflation associated with previous trials was thereby eliminated, and the overhead associated with the final assessment exam was significantly reduced. As documented in Table 5, the overall outcome demonstration success rates were not impacted by this change of strategy. This trial took place over two consecutive semesters (Fall 2004 and Spring 2005), for which the overall outcome demonstration success rates were 85.2% and 82.6%, respectively.

Table 3. Trial 5 Primary Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Average Score	Passing Threshold	Passed	Failed
1. an ability to analyze static and dynamic behavior of digital circuits	68.4%	54.1%	86.3%	13.7%
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement …	81.0%	60.0%	92.4%	7.6%
3. an ability to use a hardware description language to specify combinational logic circuits, including various …	72.2%	56.4%	86.0%	14.0%
4. an ability to design and implement arithmetic logic circuits	60.6%	40.0%	79.0%	21.0%
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify	81.8%	60.0%	91.9%	8.1%
6. an ability to design and implement a simple computer based on combinational and sequential building blocks	57.8%	40.0%	75.5%	24.5%

Table 4. Trial 5 Final Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Successful
1. an ability to analyze static and dynamic behavior of digital circuits	95.7%
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement …	94.9%
3. an ability to use a hardware description language to specify combinational logic circuits, including various …	92.4%
4. an ability to design and implement arithmetic logic circuits	90.3%
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify	90.6%
6. an ability to design and implement a simple computer based on combinational and sequential building blocks	88.8%

Table 5. Trial 6 Final Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Successful (Fall 2004)	Successful (Spring 2005)
1. an ability to analyze static and dynamic behavior of digital circuits	94.4%	92.9%
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement …	93.0%	94.0%
3. an ability to use a hardware description language to specify combinational logic circuits, including various …	93.0%	92.4%
4. an ability to design and implement arithmetic logic circuits	90.8%	91.3%
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify	90.8%	92.4%
6. an ability to design and implement a simple computer based on combinational and sequential building blocks	87.3%	87.0%

Trial 7

An acknowledged issue with Trial 6 was perhaps going "too far" in eliminating the effect of the final assessment exam on the course grade calculation. Effective Fall 2005 (Trial 7), the opportunity for students to retake an outcome for grade improvement was reintroduced. Instead of using the maximum of the primary and final assessment exams as the effective score for a given outcome (as was done previously), a *weighted average* was used instead (here, a weight of 60% was applied to the primary assessment score and a weight of 40% was applied to the final assessment score). If a student elected not to re-take an outcome on the final, a weight of 100% was applied to the primary assessment score for that outcome. As was the case for Trial 5, students were limited to re-taking at most 50% of the outcomes on the final assessment exam (here, up to three of the six outcomes). The post-final grade inflation that occurred was nearly an *order of magnitude smaller* (only about 0.1on a 4.0 scale, compared with as much as 1.0 on a 4.0 scale for Trial 5). Note that, in theory, the overall outcome demonstration success rate should not have been impacted by the changes made in how the final assessment exam contributed to the course grade calculation. Inexplicitly, however, the success rates were somewhat lower than expected for this trial, as documented in Table 6. Overall, only 78.2% of the 129 students enrolled passed all the outcomes based on the final assessment, which was about 5% lower than expected. While there was some additional overhead involved with scheduling and processing the final assessment exam (compared with Trial 6), it was not excessive. Students certainly seemed pleased, however, to have the opportunity to improve their course grade on the final. Trial 7 is currently in use as of this writing (Spring 2006).

Table 6. Trial 7 Final Assessment Statistics for Sophomore Digital Systems Course.

Outcome	Successful (Fall 2005)	Successful (Spring 2006)
1. an ability to analyze static and dynamic behavior of digital circuits	87.2%	TBD
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement …	88.7%	TBD
3. an ability to use a hardware description language to specify combinational logic circuits, including various …	85.7%	TBD
4. an ability to design and implement arithmetic logic circuits	85.7%	TBD
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify	85.7%	TBD
6. an ability to design and implement a simple computer based on combinational and sequential building blocks	81.2%	TBD

Parallel Trials

In parallel with the trials in the sophomore-level digital systems course, the author ran similar trials in the junior-level microcontroller course. Not surprisingly, very similar results were obtained. For reference, comparable Trial 4 data is provided in Tables 7 and 8 (results are shown for the three learning outcomes common to all seven trials). Here, 72 out of 94 (76.6%) of the students demonstrated all outcomes based on the final assessment (this after a fairly high failure rate on the primary assessment exams). A total of 10 students (11%) who would have otherwise passed the course failed to demonstrate at least one outcome, and therefore received a grade of "E" (conditional failure).

Table 7. Trial 4 Primary Assessment Statistics for Junior Microcontroller Course.

Outcome	Avg. Score	Passed	Failed
1. an ability to write programs for a computer in assembly language	66.1%	70.2%	29.8%
2. an ability to interface a microprocessor to various devices	58.8%	55.3%	44.7%
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	59.6%	60.6%	39.4%

Table 8. Trial 4 Final Assessment Statistics for Junior Microcontroller Course.

Outcome	Successful
1. an ability to write programs for a computer in assembly language	84.0%
2. an ability to interface a microprocessor to various devices	81.9%
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	83.0%

Comparable Trial 5 results for the junior microcontroller course are reported in Tables 9 and 10. Comparing these to the results depicted in Tables 6 and 7, note the much more reasonable "initial pass rates" afforded by the dynamic threshold (EXAM MEAN – STANDARD DEVIATION). Of the 152 students enrolled, 92.8% (141) were able to successfully demonstrate all outcomes based on the final assessment. The number of students who failed the course outright (based on failure to demonstrate at least half of the outcomes on the primary assessment exams) was 8 (about 5%). As was the case for the sophomore course, the results for the junior course seem more reasonable than the corresponding Trial 4 results obtained.

Table 9. Trial 5 Primary Assessment Statistics for Junior Microcontroller Course.

Outcome	Average Score	Passing Threshold	Passed	Failed
1. an ability to write programs for a computer in assembly language	62.1%	45.0%	88.7%	11.3%
2. an ability to interface a microprocessor to various devices	65.5%	48.5%	85.0%	15.0%
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	69.9%	54.1%	86.0%	14.0%

Table 10. Trial 5 Final Assessment for Junior Microcontroller Course.

Outcome	Successful
1. an ability to write programs for a computer in assembly language	94.1%
2. an ability to interface a microprocessor to various devices	94.1%
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	92.8%

Comparable Trial 6 results for the junior microcontroller course are reported in Table 11. Like the sophomore pre-requisite course, this trial was in place for two consecutive semesters (Fall 2004 and Spring 2005), for which the overall outcome demonstration success rates were 94.7% and 94.1%, respectively.

Table 11. Trial 6 Final Assessment for Junior Microcontroller Course.

Outcome	Successful (Fall 2004)	Successful (Spring 2005)
1. an ability to write programs for a computer in assembly language	97.7%	94.1%
2. an ability to interface a microprocessor to various devices	95.9%	94.1%
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	95.3%	95.6%

Finally, comparable Trial 7 results for the junior microcontroller course are reported in Table12. The overall outcome demonstration success rate for this trial was 94.0%, which is comparable to the two Trial 6 results detailed above. This is in contrast to the apparently anomalous results obtained for the sophomore pre-requisite course, where a 5% reduction in overall outcome demonstration success rate was observed for Trial 7 relative to Trial 6.

Table 12. Trial 7 Final Assessment for Junior Microcontroller Course.

Outcome	Successful (Fall 2005)	Successful (Spring 2006)
1. an ability to write programs for a computer in assembly language	98.8%	TBD
2. an ability to interface a microprocessor to various devices	95.2%	TBD
3. an ability to effectively utilize the wide variety of peripherals integrated into a contemporary microcontroller	94.0%	TBD

Lessons Learned

The primary lesson learned throughout the various trials is the rather delicate balance between: (a) assigning course grades consistent with proven prior practice; (b) providing incentives for students to successfully demonstrate outcomes; (c) establishing reasonable, meaningful thresholds for outcome demonstration success that are decoupled to the extent possible from the "exam difficulty" (and other factors beyond the instructor's control, such as which other courses have exams during the same time period); and (d) determining a fair level of "filtering" based on outcome demonstration success (relative to awarding passing grades). There is also tension between providing a reasonable number of attempts for outcome demonstration (ideally limited to proctored exam/quiz situations) while keeping the incremental workload associated with tracking outcome compliance to a reasonable level.

The results obtained in Trials 6 and 7 appear to successfully balance most of these constraints in a reasonable manner. The threshold for successful outcome demonstration is indeed providing some "filtering" (i.e. causing students who would otherwise have passed the course to fail based on an outcome demonstration deficiency), as mandated by our accreditation visitors. Trial 6 minimizes the overhead associated with outcome remediation, while Trial 7 perhaps strikes a better balance between overhead and student incentive (yet keeping grade inflation associated with the incentive at a nominal level). Perhaps most importantly, though, the entire process (application of outcome assessment) is *helping students learn the course material* as well as helping them to focus their learning on the aspects of the course material considered key.

Summary and Conclusions

A wide range of outcome assessment strategies have been devised, deployed, and already discarded as a consequence of the new accreditation standards spelled out in ABET 2000. Converging toward "best practices" has proven to be a non-trivial exercise, here spanning seven trials over nearly five years. Quantitative comparison of different outcome assessment and tracking strategies has proven to be a valuable, instructive exercise that has provided effective, working solutions to the questions originally posed. Specifically, a set of content-specific learning outcomes has been formulated that can be consistently and quantitatively assessed; outcome assessment instruments have been formulated along with a mechanism to determine outcome demonstration thresholds; grading strategies have been formulated that incorporate outcome demonstration thresholds, yet produce results consistent with prior (accepted) grading practices; and, finally, outcome remediation strategies have been formulated that are both fair and efficient. The most important result, however, is that effective application of outcome assessment (using appropriate evaluation instruments, outcome demonstration success thresholds, incentives, and grading strategies) *promotes and helps learning.*

Bibliography

1. http://www.abet.org
2. http://www.abet.org/Linked%20Documents-UPDATE/Program%20Docs/EAC%20Guidelines%20for%20Criterion3.pdf
3. Atman, C. J., *et al.*, "Matching Assessment Methods to Outcomes: Definitions and Research Questions," 2000 *American Society for Engineering Education Conference Proceedings*.
4. Felder, R. M. and Brent, R., "Designing and Teaching Courses to Satisfy the ABET Engineering Criteria." *J. Engr. Education,* 92(1), 7–25 (2003).
5. Jenkins, M. G., and Kramlich, J. C., "Assessment Methods under ABET EC2000 at the University of Washington – Lessons Learned: What Works and What Doesn't," 2002 *American Society for Engineering Education Conference Proceedings.*
6. Meyer, D. G., "Capstone Design Outcome Assessment: Instruments for Quantitative Evaluation ", 2005 *Frontiers in Education Conference Proceedings.*
7. http://shay.ecn.purdue.edu/~dsml/ece270
8. http://shay.ecn.purdue.edu/~dsml/ece362
9. Meyer, D. G., "Outcome Assessment: Practical Realities and Lessons Learned", 2004 *Frontiers in Education Conference Proceedings.*