Outcome Assessment: Practical Realities and Lessons Learned

David G. Meyer
Purdue University, School of Electrical and Computer Engineering
West Lafayette, IN 47907 meyer@purdue.edu

***Abstract* - The challenges of implementing ABET 2000 outcome assessment strategies for a sequence of three computer engineering courses (spanning sophomore-level to senior design) are described in this paper, with a focus on practical realities and lessons learned. 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.**

Index Terms - ABET 2000, dynamic assessment thresholds, evaluation instruments, outcome assessment, remediation strategies, static assessment thresholds.

INTRODUCTION

A veritable plethora of outcome assessment strategies have been devised, deployed, and ultimately discarded as a result of DYOTAD: the "do your own thing (and we'll tell you whether or not we like it) accreditation dilemma." This paper documents the author's journey toward formulating a coherent and practical outcome assessment strategy for a sophomore-to-senior sequence of computer engineering courses. Short of providing entertainment value, the hope is that through sharing what has worked and, in particular, what has *not* worked, others working toward the same goal might benefit.

While initially somewhat skeptical of the ABET 2000 paradigm, the author has become fully committed to the idea of outcome assessment in engineering education. The major benefit is that outcome assessment provides an opportunity to focus students' attention on the key topics being taught (and that instructors hope students will retain), which leads to better learning as well as improved performance in post-requisite courses.

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

- ECE 270 *Introduction to Digital System Design*, a sophomore level lecture/lab course on digital logic and system design with an annual enrollment of approximately 500 students [1].
- ECE 362 *Microprocessor System Design and Interfacing*, a junior-level lecture/lab course on embedded microcontroller systems with an annual enrollment of approximately 300 students [2].
- ECE 477 *Digital Systems Design Project*, a senior-level course that involves a semester-long, team-oriented, multidisciplinary design project [3].

OUTCOME FORMULATION

When confronted with the necessity of formulating a set of learning outcomes for the entire undergraduate curriculum, the ECE School Administration utilized a relatively uninspired approach: lock all the faculty in a room (euphemistically referred to as a "retreat") and not allow anyone to leave until the task was complete. Upon trying to actually deploy an outcome assessment strategy based on these somewhat hastily devised lists, however, some significant revisions had to be made.

For the sophomore-level course (ECE 270), a set of ten outcomes was finally formulated:

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

At the moment (and for next few semesters), this set of ten outcomes seemed reasonable: it represented a fairly 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.
2. an ability to represent Boolean functions in standard forms, to map and minimize them, and to implement them as combinational logic circuits.
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.
4. an ability to design and implement arithmetic logic circuits.
5. an ability to analyze, design, and implement sequential circuits and use a hardware description language to specify them.
6. an ability to design and implement a simple computer based on combinational and sequential building blocks.

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).

For the junior-level course in the series (ECE 362 *Microprocessor System Design and Interfacing*), the major categories of topics were fewer in number than in the sophomore-level course, yielding a relatively compact set of four learning outcomes:

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

The current set allows the outcomes to be tested using a series of four hourly exams.

Finally, for the senior design course (ECE 477), the set of learning outcomes common to all Purdue ECE senior design options was adopted:

1. an ability to apply knowledge obtained in earlier coursework and to obtain new knowledge necessary to design and test a microcontroller-based digital system.
2. an understanding of the engineering design process.
3. an ability to function on a multidisciplinary team.
4. an awareness of professional and ethical responsibility.
5. an ability to communicate effectively, in both oral and written form.

MAJOR ISSUES

The initial set of learning outcomes and strategies for assessments 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 what algorithm should be 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 (ECE 270 *Introduction to Digital System Design*) as an illustrative example.

STRATEGIES ATTEMPTED

Armed with an understanding of the basic issues involved, several "refinements" of outcome assessment strategies employed in the sophomore/junior courses are chronicled. 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:

- fixed passing threshold (60%) on weighted sum of (selected) lab, homework, and exam scores.
- fixed passing threshold (60%) on primary assessment, remediation homework, and final assessment.
- fixed passing threshold (60%) on primary assessment, final assessment, and remediation homework.
- fixed passing threshold (60%) on primary and final assessments; use of "E" 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 Purdue, *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 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 sets.

TRIAL 4

The third refinement was based on an arcane, rarely used grading option available at Purdue: the "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 for "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 whatever assessment strategy chosen *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 (even after over twenty years of experience). Basically, there are too many factors "beyond an instructor's control" – particularly timing of exams relative to other courses students might be taking.

It is instructive to examine the data relative to this trial. Table I provides the primary assessment exam statistics.

TABLE I
TRIAL 4 PRIMARY ASSESSMENT STATS (ECE 270)

Outcome	Avg. Score	Passed	Failed
1	71.5%	75.2%	24.8%
2	63.4%	57.7%	42.3%
3	75.8%	84.3%	15.7%
4	65.0%	68.6%	31.4%
5	61.7%	55.8%	44.2%
6	62.0%	55.8%	44.2%
7	80.5%	89.4%	10.6%
8	65.1%	65.0%	35.0%
9	54.7%	46.4%	53.6%
10	54.5%	41.6%	58.4%

Upon completion of the primary assessments, only 53 out of 274 students had successfully demonstrated all ten course outcomes (less than 20%). Over 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 II 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).

TABLE II
TRIAL 4 FINAL ASSESSMENT STATS (ECE 270)

Outcome	Successful
1	98.5%
2	99.3%
3	99.6%
4	99.6%
5	97.8%
6	94.5%
7	98.2%
8	93.8%
9	85.4%
10	91.2%

An interesting adjunct of this trial 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 I and II worth noting. 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 II, 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.”

TRIAL 5 (CURRENT STATUS)

The strategy currently employed (the “fourth refinement”) finally recognizes the need to utilize a dynamic threshold for successful outcome demonstration, rather than a static one. To determine an appropriate choice, the author analyzed outcome demonstration data spanning several semesters for both ECE 270 and ECE 362. A fairly 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.

Preliminary results indicate much higher “initial pass rates” based on use of the dynamic threshold described earlier (note that the number of outcomes has been compacted to six for this trial). 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 seem more "reasonable" than the Trial 4 data reported earlier.

TABLE III
TRIAL 5 PRIMARY ASSESSMENT STATS (ECE 270)

Outcome	Avg. Score	Thresh	Passed	Failed
1	68.4%	54.1%	86.3%	13.7%
2	81.0%	60.0%	92.4%	7.6%
3	72.2%	56.4%	86.0%	14.0%
4	60.6%	40.0%	79.0%	21.0%
5	81.8%	60.0%	91.9%	8.1%
6	57.8%	40.0%	75.5%	24.5%

TABLE IV
TRIAL 5 FINAL ASSESSMENT STATS (ECE 270)

Outcome	Successful
1	95.7%
2	94.9%
3	92.4%
4	90.3%
5	90.6%
6	88.8%

PARALLEL TRIALS

In parallel with the trials in ECE 270, the author ran similar trials in the junior-level course ECE 362. Not surprisingly, very similar results were obtained. Comparable "Trial 4" data is provided in Tables V and VI. Here, 72 out of 94 (76.6%) of the students demonstrated all four 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).

The equivalent of Trial 5 has already been completed in ECE 362, and reported in Tables VII and VIII. Comparing the results depicted in Tables V and VII, 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 four course 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%). Again, this seems more reasonable than the Trial 4 results obtained.

TABLE V
TRIAL 4 PRIMARY ASSESSMENT STATS (ECE 362)

Outcome	Avg. Score	Passed	Failed
1	57.2%	50.0%	50.0%
2	66.1%	70.2%	29.8%
3	58.8%	55.3%	44.7%
4	59.6%	60.6%	39.4%

TABLE VI
TRIAL 4 FINAL ASSESSMENT STATS (ECE 362)

Outcome	Successful
1	80.9%
2	84.0%
3	81.9%
4	83.0%

TABLE VII
TRIAL 5 PRIMARY ASSESSMENT STATS (ECE 362)

Outcome	Avg. Score	Thresh	Passed	Failed
1	76.7%	58.1%	88.8%	11.2%
2	62.1%	45.0%	88.7%	11.3%
3	65.5%	48.5%	85.0%	15.0%
4	69.9%	54.1%	86.0%	14.0%

TABLE VIII
TRIAL 5 FINAL ASSESSMENT STATS (ECE 362)

Outcome	Successful
1	96.7%
2	94.1%
3	94.1%
4	92.8%

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 Trial 5 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. 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).

DESIGN COURSE VARIANCES

Space does not permit a detailed look at the outcome assessment strategies optimal for "design" courses (in contrast to the "content" courses detailed previously), but one issue is clear: the kinds of outcomes (related to teamwork, communication, professional skills) are totally different. The major challenge here has been to quantify the assessment of these inherently qualitative outcomes and apply appropriate thresholds. Another challenge for a team-oriented project course is ensuring equitable distribution of workload and grade determination based on both individual as well as corporate contributions.

Evaluation instruments chosen to quantitatively evaluate the five course outcomes (listed previously) include:

1. a design component homework (e.g. schematic design, printed circuit board layout design, etc.).
2. the individual lab notebook.
3. the team-defined project success criteria.
4. a professional component homework (e.g. reliability and safety analysis, patent liability analysis, etc.).
5. the formal design review, final video presentation, and final written report.

Remediation needs to be handled differently as well: here, students who initially fail to demonstrate an outcome (e.g. receive a score on a design or professional homework below the prescribed passing threshold) must be given an opportunity for prompt remediation (e.g. rewriting the deficient paper, correcting the printed circuit board layout, etc.). Because the evaluation instruments are much more "standardized" than in a content course (where one is forced to constantly come up with "new" exam questions), fixed passing thresholds (60%) are applied to each of the five outcomes. Good results have been obtained using this strategy the past few semesters. Details about the course, including archived web sites of completed design projects, can be found at [3].

SUMMARY AND CONCLUSIONS

A wide range of outcome assessment strategies have been devised, deployed, and already discarded as a consequence of DYOTAD – the "do your own thing (and we'll tell you whether or not we like it) accreditation dilemma." Different kinds of courses (in particular, "content" vs. "design") require different outcome assessment strategies, and finding the "best practices" for each case is non-trivial. Quantitative comparison of different outcome assessment and tracking strategies has proven to be a valuable, instructive exercise. 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.* "Outcome assessment is a good thing."

REFERENCES

[1] http://shay.ecn.purdue.edu/~dsml/ece270

[2] http://shay.ecn.purdue.edu/~dsml/ece362

[3] http://shay.ecn.purdue.edu/~dsml/ece477