ASSESSING STUDENT PERFORMANCE ON EC2000 CRITERION 3.a-k

Linda C. Strauss, Patrick T. Terenzini
The Pennsylvania State University

Abstract

The instruments currently available to measure the 11 student learning skills specified in *EC2000*'s Critrion 3 are vulnerable to challenges to their validity and reliability. This paper describes the development and evaluation of a 36-item measure specifically designed to assess student performance on those outcomes. Development followed standard test-construction procedures, including a comprehensive review of the literature and available instruments, item and content reviews by engineering faculty and student focus groups, a large pilot test, and a field administration involving 4,558 seniors at 39 randomly selected colleges of engineering. The resulting nine factor-scales align closely with the *EC2000* criteria and retain 72% of the original item variance. All but two scales have internal consistency reliabilities above .83. The instrument appears to be a conceptually faithful, psychometrically sound, and practical tool for assessing student learning on Criterion 3.

****

In 1996, the ABET Board of Directors adopted a transformative set of criteria for program accreditation that shifted the reaccreditation focus away from an emphasis on meeting curricular, resource, faculty, and facilities standards toward a new focus on student learning outcomes.[1,2] The new standards, called *Engineering Criteria 2000: Criteria for Accrediting Programs in Engineering in the United States,*[2] emphasized 11 specific learning outcomes and the assessment of program achievement on those outcomes. The new criteria maintained the previous standards' emphasis on the development of students' mathematical, scientific, and technical knowledge, but the new criteria also emphasized developing other professional skills, such as solving unstructured problems, communication, and teamwork skills. Programs are now required to present evidence of student achievement in each of 11 learning outcome areas specified in Criterion 3.a-k.[2]

The Problem

Although the EC2000 criteria specified a broad spectrum of both technical and professional skills, they did so in a way that still allows engineering programs some flexibility in their interpretation and increases the likelihood of closer alignment of the learning criteria with program goals. That interpretational flexibility, however, also created ambiguities in defining and measuring the skills that students must demonstrate if a program is to meet the intent of the criteria.

The generality of the specifications and the associated flexibility in operationalizing EC2000's Criterion 3 learning outcomes (hereafter, the "a-k criteria") have led to the emergence of a wide array of items, scales, and instruments for assessing student performance on one or more of the criteria. Few, if any, of the existing measures appear to have been developed according to the instrument/test-development standards generally recommended. Although the language and steps vary to some degree, instrument development experts typically recommend at least the following basic process:[3,4]

1. Identify content domains, the important concepts that need to be included in the instrument, and determine the weight or relative emphasis each should receive.
2. Write initial items/questions and appropriate response options for each domain.
3. Edit resulting items and response options, for example, multiple choice or open-ended options.
4. Pilot-test items.
5. Analyze the psychometric properties of items and scales to evaluate both the accuracy (reliability) of the measurements and whether the questions in the instrument actually measure the intended concepts (validity).
6. Make final item selections.

Most instruments now available for assessing engineering student learning outcomes, however, simply ask students to rate themselves on a collection of single-item measures, each representing one of the a-k criteria. Often, the Criterion 3.a-k statements are themselves the items, either verbatim or with slight modifications. Single-item measures are known to be unreliable and unlikely to withstand any rigorous examination of their validity for measuring the intended construct.

A few researchers have developed psychometrically complex measures of individual a-k criteria[5] or of some (but not all) of the 11 criteria.[6,7,8] Instruments in the first category of such measures, psychometrically sound as they may be, are typically highly focused in their content, necessarily lengthy, and, consequently, time-intensive to complete. Instruments in the second cluster of measures provide information on only some of the Criterion 3 learning outcomes and are, thus, incomplete measures. Both groups of measures are ill-suited for broader self-study or program evaluation purposes that require assessment of all a-k criteria or are used to guide important overall programmatic decisions affecting curriculum, effective pedagogies, and resource allocations.

Practicing what it preaches, ABET undertook its own self-assessment to determine whether EC2000 is promoting the student learning outcomes specified in Criterion 3.a-k. In 2002, ABET commissioned the Center for the Study of Higher Education (CSHE) at Penn State University to undertake this assessment. The study, entitled “Engineering Change: A Study of the Impact of EC2000,” is a national study of the impact of the new outcomes criteria and associated curricular and programmatic changes on learning among undergraduate engineering students.

As part of that effort, the CSHE project group developed a measure of student performance in each of EC2000’s 11 student learning criteria. This paper describes the development process and the psychometric characteristics of that instrument. Additionally, the current work is intended to provide the engineering education community with a psychometrically sound instrument for assessing the extent to which students in any given engineering program are achieving the learning outcomes specified in Criterion 3.

Conceptual Framework

The study rests on a conceptual framework hypothesizing that post-EC2000 changes in student learning will occur to the extent that engineering programs have modified their curricula, instructional practices, institutional policies, and faculty cultures and attitudes in response to EC2000. The framework suggests that linkages between EC2000 and learning outcomes are largely indirect: preparation for an EC2000 re-accreditation is hypothesized to promote curricular and other changes that, in turn, affect student learning. Figure 1 portrays these hypothesized relationships.

Design, Population, and Sample

The main project’s design entails data collection from several sources, including graduating seniors, alumni, faculty members, program chairs, deans, and employers. The instrument described here was designed for use with alumni and senior students. Data collection followed a modified time-series design,[10] taking cross-sectional samples both pre- and post-implementation of the EC2000 criteria. This design includes data collected from two separate groups of students at different points in time to allow evaluation of any changes over time.

The population for the study was defined to include all undergraduate engineering programs holding ABET accreditation since 1990 or earlier in the following seven fields: aerospace, chemical, civil, computer, electrical, industrial, and mechanical engineering. This disciplinary array permitted study of the disciplines that produce the vast majority of undergraduate engineering degrees (chemical, civil, electrical, and mechanical), as well as disciplines with strong ties to industry sectors (aerospace, computer, and industrial). Of the 1,241 currently ABET-accredited engineering programs in the targeted disciplines, 1,024 were accredited in 1990 or earlier.

The project team selected specific programs for participation in the study based on a two-stage, disproportionate, stratified random sample. In the first stage, the project group identified all institutions offering ABET-accredited undergraduate degrees since 1991 or earlier in at least two of the seven programs listed above. The second stage entailed a 7x3x2 sampling design that included the seven targeted disciplines, three EC2000 "adoption status" groups (based on when an institution first elected to undergo reaccreditation based on the EC2000 criteria: earlier than required, when required, or deferred when optional), and whether an institution participated in a National Science Foundation Engineering Education Coalition (yes or no). Aerospace and industrial engineering programs were over-sampled to ensure an adequate number of responses for robust statistical analysis. The sample was augmented further by the addition of four EC2000 pilot institutions (first reviewed in 1996 and 1997) and three Historically Black Colleges and Universities. Efforts were made to include Hispanic-Serving Institutions in replacing original institutions. The final sample included 203 programs at 40 institutions.

Item Development

Because a preliminary review of the existing engineering literature and surveys identified no psychometrically sound, comprehensive, and practical measure of all the learning outcomes specified in Criterion 3.a-k, CSHE project staff set out to create such a measure following the standard test/instrument development procedures listed above. The work of the engineering scholars and practitioners who produced the initial EC2000 criteria was the equivalent of the initial step in this process through their specification of the relevant content domains (the Criterion 3.a-k learning outcomes). Because that group offered no guidance on the *relative* importance of the 11 criteria, the test to be developed would treat all as equally important. Whether such a specification of weights is possible, appropriate, or desirable remains a set of questions for the engineering education community to consider.

Following the guidelines listed above, the CSHE project group developed the process summarized in Figure 2. Given the a-k outcomes to be measured, instrument development began with a thorough and comprehensive search for potentially relevant items reported in the engineering education literature or used in currently available instruments. Literature sources included leading scholarly journals in engineering education and higher education. The search also included surveys and literature produced by the NSF coalitions.

The search yielded 286 items that researchers and program evaluators have used to measure one or more of the skills described in Criterion 3.a-k. Project staff followed an iterative process to reduce the number of items to those with high relevance to one of the a-k criteria. First, the item bank was compiled and sorted according to each item's appropriateness for an a-k criterion. Project group members then discussed the merits of each item and eliminated those thought to be irrelevant, ambiguously worded, or relevant to multiple criteria. The project group also wrote new items when a criterion area had fewer than three relevant items.

In keeping with Step 3 in the list of instrument-development tasks, project staff next engaged five engineering faculty members at Pennsylvania State University in a review and critique of the items for each criterion. These faculty members were selected for their knowledge of ABET and the EC2000 learning criteria, for their extensive experience teaching undergraduate students, and for their active involvement in a recent self-study and ABET-reaccreditation review at Penn State. The faculty members represented aerospace, civil, and mechanical engineering; one member of the group was the associate dean for undergraduate engineering education. Faculty members and project researchers met to review the list of items and to discuss suggestions for item revisions, deletions, and additions. Based on faculty members' comments, project staff further refined and added items.

A second meeting with these same five faculty engineers to review changes made based on recommendations from the earlier meeting led to more item polishing and a final set of items. This overall process yielded 38 items, with three to five items believed *a priori* to represent each of the 11 a-k criteria.

Pilot Testing

With the instrument's content validity (domain coverage) confirmed both conceptually (by the ABET-sponsored group that developed the a-k statements) and operationally (through the faculty review of the item sets for each criterion), CSHE researchers undertook an empirical assessment of the construct validity of the instrument and the internal consistency reliability[11] of its (to this point only *a priori*) scales. In the summer of 2003, project staff using the Penn State data warehouse, created a list of all undergraduate students majoring in aerospace, chemical, civil, computer, electrical, industrial, or mechanical engineering and who had earned more than 72 academic credits (placing them in "senior" status). Working with the Survey Research Center of the Penn State Social Science Research Institute, project staff contacted 1,704 students by e-mail, explaining the purposes of the survey and asking them to complete the 38 item survey. The contact e-mail went out in the name of the dean of the College of Engineering. A total of 368 (21.6%) of the students completed the on-line survey; no follow-up contact was attempted.

Project staff then undertook a series of analyses of the responses of the 368 seniors to the 38 a-k criteria. These analyses included a principal components factor analysis, with varimax rotation. Factor analysis identifies patterns in the item responses and yields clusters of items that appear to be tapping the same construct. Those items can then be combined into scales, allowing more efficient item analyses and reporting. The analysis of the 38 items in this study indicated eight underlying factors, or scales. The analyses also included assessment of the internal consistency reliability (Cronbach's Alpha) of the scales suggested by the principal components analysis. Items loading below .40 on a factor (indicating the contribution to that factor is small), or above .40 on two or more factors (suggesting that the item was ambiguous) were deleted. Similarly, items were deleted when tests indicated their removal from a scale would appreciably raise the scale's internal consistency reliability.

In addition to these quantitative analyses, project staff arranged a qualitative focus group with graduating engineering seniors. Administrators in the College of Engineering suggested seven students whom they believed could provide useful insight into how students might interpret the survey items. Five students agreed to participate in the session. These students provided significant feedback regarding, for example, organization of the 38 items, lack of specificity in the question stem, changes in the response option scales, and whether both in- and out-of-class experiences should be considered as influences on their abilities in the 3.a-k outcome areas.

Results of the quantitative and qualitative studies led to deletion of two of the original 38 items. Comments of the focus group led to rewording of several other items.

Finally, the project's National Advisory Board reviewed the resulting 36 items and the scales identified in the factor analyses of these items. The Advisory Board includes college of engineering deans, representatives of industry and one of the major engineering societies, and engineering education assessment experts. The board reviewed and approved the 36 items and the resulting factor structure from the item development process and the pilot study, providing additional confirmation of the instrument's content and construct validity.

Field Administration

In spring, 2004, the population of 12,621 seniors nearing graduation in any of seven engineering fields (aerospace, chemical, civil, computer, electrical, industrial, and mechanical) on 39 nationally representative campuses selected according to the sampling design described above were sent the final survey instrument (available at http://www.ed.psu.edu/cshe/abet/instruments.html) (one of the original 40 institutions did not provide student responses). That measure included the 36 final a-k items, as well as other items and scales tapping seniors' experiences and perceptions of their engineering programs. Students evaluated their abilities on each of the 36 items in response to the following request: "Thinking about your in-class and out-of-class experiences, please rate your ability to do the following:" Students rated their abilities on a five-point scale, were 1 = "no ability," 2 = "some ability," 3 = "adequate ability," 4 = "more than adequate ability," and 5 = "high ability."

Hard copies of the instrument (and the URL for an electronic version) were sent to students at 20 of the institutions, and electronic versions only were sent to the students at the other 20 institutions. The cover letter in both administrations was written by the dean of the institution's college of engineering. Follow-ups included a postcard sent two weeks after the initial mailing and a complete follow-up (similar to the initial mailing) sent two weeks after the postcard.

Data Analysis

A total of 4,558 (36.1%) of the seniors completed and returned the survey. After resolving missing and incomplete data issues, the usable number of responses was 4,461, a usable response rate of 35.3 percent. To correct for response bias, respondents were weighted so as to produce distributions that were representative of their institution's senior population according to gender and engineering major. Respondents were also weighted to correct for varying response rates across institutions. Thus, institutions with relatively high or low response rates would not be over- or under-represented in the dataset. The resulting weighted sample is representative of the national population of seniors receiving engineering degrees in the targeted fields in the 2003-04 academic year.

To determine the optimal number of meaningful factors, project staff again examined the revised (final) 36 learning outcome items in a series of principal components analyses with varimax rotations. Table 1 summarizes the results of these analyses and subsequent reliability analyses. These analyses produced a nine-factor solution, only slightly different in structure and item content from those produced in the pilot test. The final, nine-factor structure retains 72.2 percent of the overall item variance among the 36 items, fit the correlation matrix best and was the most consistent conceptually with the a-k criteria. All 36 tems loaded above .40 on a single factor and were retained to develop factor scales. Table 1 shows the items on each of the nine factors, as well as the internal consistency reliability coefficients (Cronbach's Alpha) for each of the factor scales. The alphas were above .83 on all but two scales. On those scales, the alphas were .74 and .78.

Figure 3 maps the relationships between the nine-factor solution and the original a-k criterion statements. As can be seen in the figure, and with two exceptions, each factor aligns clearly in its content with one of the 11 a-k criteria. In the two exceptions, two sets of items expected to load on discrete factors reflecting two EC2000 criteria (e and j) loaded on another factor (nos. 1 and 2, respectively). Although the generally clear alignment of factor scales with distinct EC2000 outcomes criteria is psychometrically noteworthy, the failure of two sets of items to load on discrete criteria probably reflects to some degree the conceptual and practical overlap among the original criteria statements. Indeed, the overlap might be expected considering the definitional flexibility in the original statements.

Although the calculation of factor scores would provide uncorrelated measures of student performance on each dimension, such statistically independent measures would in all likelihood be unreasonable representations of what in practice are probably interrelated skills (it is expected that both the skills represented in a-k and the scales do overlap). Thus, factor scale scores were calculated: students' responses to each item comprising a given scale were summed and then divided by the number of items in the scale. This procedure, however, allows for intercorrelations among scales[12] and, in this study, probably provides a closer representation of the reality of the conceptual overlap among the a-k criteria. Table 2 provides information on the degree of overlap among the nine factor scales. As can be seen there, the intercorrelations are moderate to high. Such collinearity would be a cause for concern *if* the scales were to be used as independent, or predictor, variables. Because their use is most likely to be as outcome (dependent) measures for self-study and ABET re-accreditation purposes, however, the scales' intercorrelations are considerably less threatening to meaningful interpretation.

Limitations

The instrument and scales reported on here, like all other tests and measures, are limited in several ways. First, the engineering skills specified in EC2000's Criterion 3 are complex and in many cases multi-dimensional abilities. Although the scales operationalizing those learning outcomes appear to have sound content validity and generally high internal consistency reliabilities, those measures probably reflect only partially the complexity and subtlety of the talents that underlie the specified skills. Although more precise representation of the skills implied in the a-k learning criteria might be a desirable condition, such increased precision would require specification of the component skills implied in each learning outcome criterion. In addition, more items and scales would be needed to operationalize those subdimensions, with a corresponding increase in the length of the measure. The dilemma is a clear illustration of the classic "depth vs. breadth" tradeoffs frequently required in instrument development.

Second, the *EC2000* authors intended the a-k criteria to apply to student learning in all colleges or schools of engineering. It remains an empirical question, however, whether the nine factor-scale structure of the measure described here will perform similarly across all kinds of engineering schools. Although the nine scales appear to have strong content validity, these items and scales were reviewed and validated by engineering faculty and students at only one large, research university and by members of the project's National Advisory Board, most of whom had research university backgrounds. Thus, although it seems unlikely that the skills tapped by the scales described here will vary with institutional type, size, or curricular structure, that possibility remains.

Third, the characteristics reported are based on the responses of students in only seven engineering disciplines. Although these disciplines produce the vast majority of undergraduate engineering degrees, the possibility remains that the measure may perform differently with students in disciplines not represented in this study. Indeed, the instrument's performance may vary somewhat across the disciplines used in this study. Similarly, the instrument may perform differently with students who are not as advanced in their programs as were the graduating seniors who participated in this study.

Fourth, the measure described here contains only nine scales in contrast to the 11 skill sets articulated in EC2000's Criterion 3. It is possible that the difference indicates the present items' failure to differentiate clearly between Criterion 3.c ("Design[ing] a system, component, or process to meet desired needs") and 3.e ("Identify[ing], formulat[ing], and solv[ing] engineering problems"), and between 3.h ("Understand[ing] the impact of engineering solutions in a global and societal context") and 3.j ("Know[ing] contemporary issues"). It is also possible, however, that the items' failure to load on four, rather than two, discrete factors reflects a moderate to high degree of conceptual and practical correspondence between 3.c and 3.d and between 3.g and 3.j. Only future study using different item sets is likely to resolve this question.

Fifth, the scales are based on seniors' reports of the progress they believe they have made in each skill area during their undergraduate engineering program. Thus, these measures do not require students to demonstrate each skill on more objective measures (e.g., standardized tests or demonstrations of skill levels). Nonetheless, the research literature indicates a moderate to strong correspondence between students' self-reports and more objective measures of learning outcomes. Pike,[13] for example, found that the correlation between self-reports and objective measures was a function of the extent to which the self-reported items and scales reflected the learning content under examination. Similarly, Anaya[14] concluded that self-reports of learning gains were valid proxies for the educational skills measured by the verbal and mathematics tests that comprise the Graduate Records Examination. As summarized by Hayek, Carini, O'Day, and Kuh,[15] five conditions should exist for self-reported information to be valid. First, the information requested is known to the participant. In this case, the survey asks students about their own experiences and abilities. Second, the questions are clearly phrased.[16] With two pilot tests and a review by a national board, this condition would appear to be met. Third, the questions relate to recent activities.[17] Since students in the study are still enrolled in their undergraduate engineering experiences these events are more current than recent. Fourth, the respondents think the questions merit a serious and thoughtful response.[18] The response rate and general completeness of responses to the overall survey form would suggest students took the study seriously. One can speculate that students who did NOT take the survey seriously probably did not complete the survey, although this is an empirical question. Finally, answering the questions does not threaten the privacy of the respondent or evoke socially desirable responses.[19] Students were advised that the survey procedures precluded their identification, thereby at least reducing, if not eliminating, these two threats. Moreover, although standardized measures or skill demonstrations have certain advantages over self-reports, they also have their own limitations when adopted for program assessment purposes. Those constraints include availability, length, cost, and relevance to engineering programs and the EC2000 learning outcomes. The self-report instrument described here was designed specifically to assess student achievement on the EC2000 learning criteria and to be easy and inexpensive to use with large groups. Nonetheless, one must acknowledge the trade-offs being made. (Other references dealing with the correspondence between student self-reports and objective measures can be found at: <http://www.ed.psu.edu/cshe/abet/resources.html>.)

Finally, this instrument is not intended to replace the complete self-study process currently required by ABET. Indeed, the measure is intended to facilitate program self-reflection. The self-study process requires engineering programs to assess, discuss, and correct program weaknesses, some of which may relate to student learning outcomes. As noted earlier, this instrument is only part of a larger effort to study the impacts on students in engineering. Engineering programs must continue to assess not only students, but curricula, faculty culture, and organizational characteristics that may be influencing student learning outcomes. No instrument can replace the important discussions happening within engineering programs. This measure is intended to be a tool to assist in the identification and monitoring of student strengths and weaknesses.

Conclusions

Engineering programs facing ABET reaccreditation must show evidence that their graduates possess certain skills. Virtually all of the instruments currently available to assess those abilities present measurement problems. Some instruments are psychometrically questionable, relying either on single-item measures of each of the 11 EC2000 learning outcomes or on scaled-measures that assess some of the criteria but not others. Still other instruments tap an EC2000 criterion in detail. Because of the length of these latter measures, however, they may not be feasible for use in a comprehensive program assessment that must cover student experiences and other outcomes besides the one central to the instrument. This paper describes the efforts taken to develop a psychometrically sound, practical, and useful instrument for assessing the learning outcomes specified in EC2000's Criterion 3. The resulting instrument appears to be an acceptable contribution to the assessment tools available for self-study and ABET reaccreditation review.

The instrument development process followed the guidelines and standards recommended by test and instrument development experts. The content and items were the culmination of a process that included a thorough literature review and an iterative review process that included faculty and students to ensure that the content of the items reflects the intention of a-k and item wording is meaningful to students.

The analyses reported above indicate that the measure is conceptually clean: its scales correspond closely to nine of the 11 EC2000 learning outcomes criteria. It also appears to be psychometrically sound. Development followed carefully the 11 EC2000 Criterion 3 statements, providing a sound foundation for the measure's content validity, and engineering faculty members attested to the measure's operational content validity through their multiple reviews of the item statements. Principal component factor analyses indicated an acceptable level of construct validity in the correspondence between the instrument's nine factors and EC2000's 11 criteria. In addition, the scales representing the nine learning outcomes have generally strong internal consistency reliabilities (all coefficient alphas being above .74, with 7 of the 9 alphas being above .83). In short, the nine-scale instrument appears to be an appropriate and acceptably accurate and parsimonious representation of the skills specified in Criterion 3.a-k.

The measure, moreover, appears to be practical, meeting the challenge of finding an appropriate intersection between psychometric rigor and practicality.[3] The instrument allows engineering programs to use these scales to meet ABET requirements with relative ease and low to moderate expense. The items can be completed by students easily and in a relatively short period of time. These scales permit sound but also parsimonious assessment the a-k outcomes, leaving room in survey instruments for the assessment of other areas of interest to engineering programs without overburdening students and, reducing survey response rates.

These scales are intended to help engineering programs across the country to assess with confidence their students' learning on learning outcomes criteria specified in EC2000. One hopes these items and scales will facilitate institutional and program efforts to respond to the EC2000 requirements and to maintain their commitment to the continuous quality improvement of their programs.

Bibliographic Information

1. Weiss, M., & Aldridge, M. D. (2002). Assessing the effects of the ABET/NSF/Industry regional faculty workshops. *International Journal of Engineering Education, 18* (2): 245-253.
2. Engineering Accreditation Commission. (1997). *Engineering Criteria 2000: Criteria for Accrediting Programs in Engineering in the United States (3rd ed.)*. Baltimore, MD: Accreditation Board for Engineering and Technology, Inc.
3. Muchinsky, P. M. (2004). When the Psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology*, *57* (1): 175-209.
4. Tinkelman, S. N. (1971). Planning the objective test. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 46-80). Washington, DC: American Council on Education.
5. Adams, S. G. (2001). *The effectiveness of the E-team approach to invention and innovation with multidisciplinary student teams*. 5th Annual National Collegiate Inventors and Innovators Alliance National Conference. March 7-10, Washington, D.C.
6. Colbeck, C. L., Cabrera, A. F., & Marine, R. J. (2001). Assessing overall competence of faculty: EC criterion
7. *ASEE Conference Proceedings, 2001.* CD Rom. Washington DC: American Society for Engineering Education.
8. Terenzini, P. T., Cabrera, A. F., & Colbeck, C. L. (1999). Assessing classroom activities and outcomes, In *Proceedings*, *1999*, *Frontiers in Education Conference*, IEEE, 19-24.
9. Cabrera, A. F., Colbeck, C. L., & Terenzini, P. T. (2001). Developing performance indicators for assessing classroom teaching practices and student learning: The case of engineering. *Research in Higher Education, 42* (3), 327-352.
10. Campbell, D. T. & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: R. McNally.
11. Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Applied Social Research Methods Series, Volume 17. Newbury Park, CA: Sage.
12. Armor, D. J. (1974). Theta reliability and factor scaling. In H. Costner (Ed.). *Sociological Methodology: 1973-1974.* San Francisco: Jossey-Bass, pp. 17-50.
13. Pike, G.R., “The relationship between self reports of college experiences and achievement test scores,” *Research in Higher Education*, vol. 36, 1995, pp. 1-22.
14. Anaya, G., “College impact on student learning: comparing the use of self-reported gains, standardized test scores, and college grades,” *Research in Higher Education*, vol. 40, 1999, pp. 499-527.
15. Hayek, J. C., Carini, R. M., O’Day, P., T., & Kuh, G. D. “Triumph or tragedy: Comparing student engagement levels of members of greek-letter organizations and other students,” *Journal of College Student Development*, vol. 43(5), 2002, pp. 643-663.
16. Laing, J., Sawyer, R., & Noble, J. “Accuracy of self-reported activities and accomplishments of college-bound seniors” *Journal of College Student Development*, vol 29(4), 1998, pp. 362-368.
17. Converse, J. M. & Presser, S. (1989*)* *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage.
18. Pace, C. R. (1985). *The credibility of student self-reports*. Los Angeles: University of California, The Center for the Study of Evaluation, Graduate School of Education.
19. Bradburn, N. M. & Sudman, S. (1988). *Polls and surveys: Understanding what they tell us*. San Francisco: Jossey-Bass.

Biographical Information

DR. LINDA C. STRAUSS. Dr. Strauss serves as a Senior Project Associate at the Center for the Study of Higher Education at Penn State University. Dr. Strauss publishes on organizational characteristics, underserved populations, and athletics in higher education. She previously served developed and operated programs for women and students of color in undergraduate engineering programs.

DR. PATRICK T. TERENZINI. A Distinguished Professor of Education, Terenzini has over 30 years of experience in higher education as a teacher, researcher, and administrator. He co-directed an evaluation of the Engineering Coalition of Schools for Excellence in Education and Leadership (ECSEL). He has published over 100 articles in refereed journals and made more than 150 invited presentations at national scholarly and professional conferences.

Figure 1. Conceptual framework for *Engineering Change: A Study of the Impact of EC2000*.

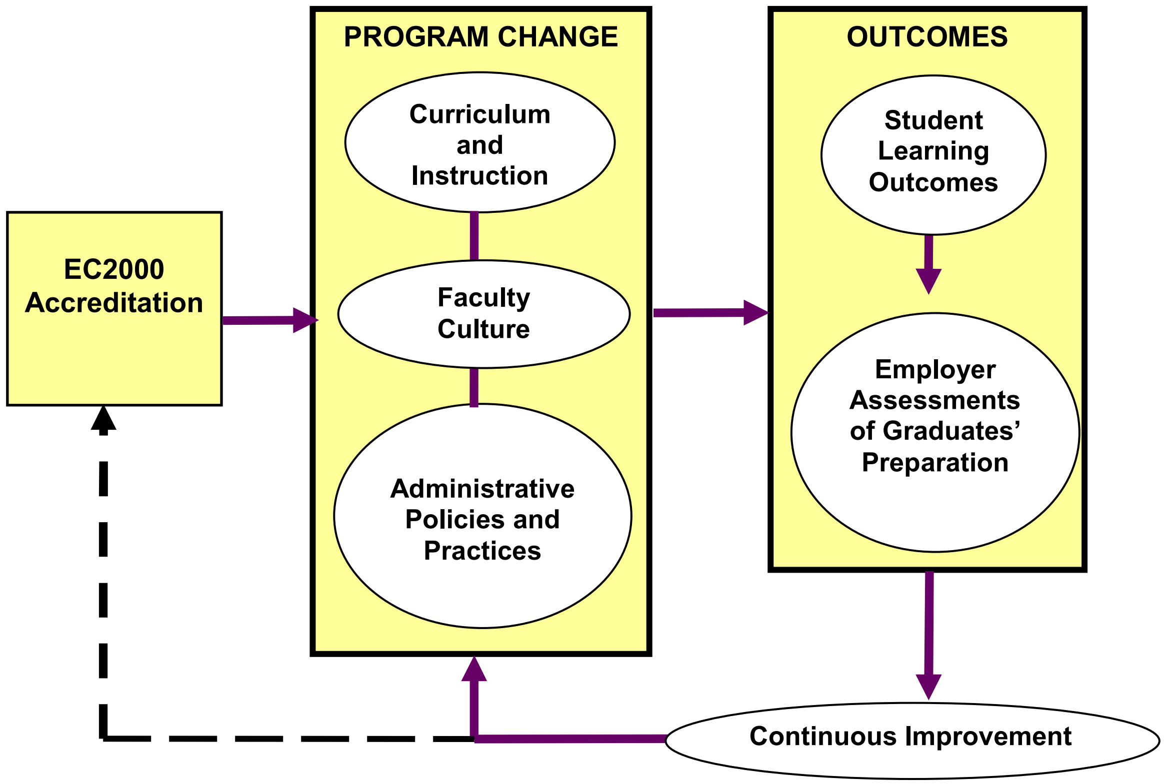

Figure 2. Process for development of items and scales operationalizing EC2000 learning criteria.

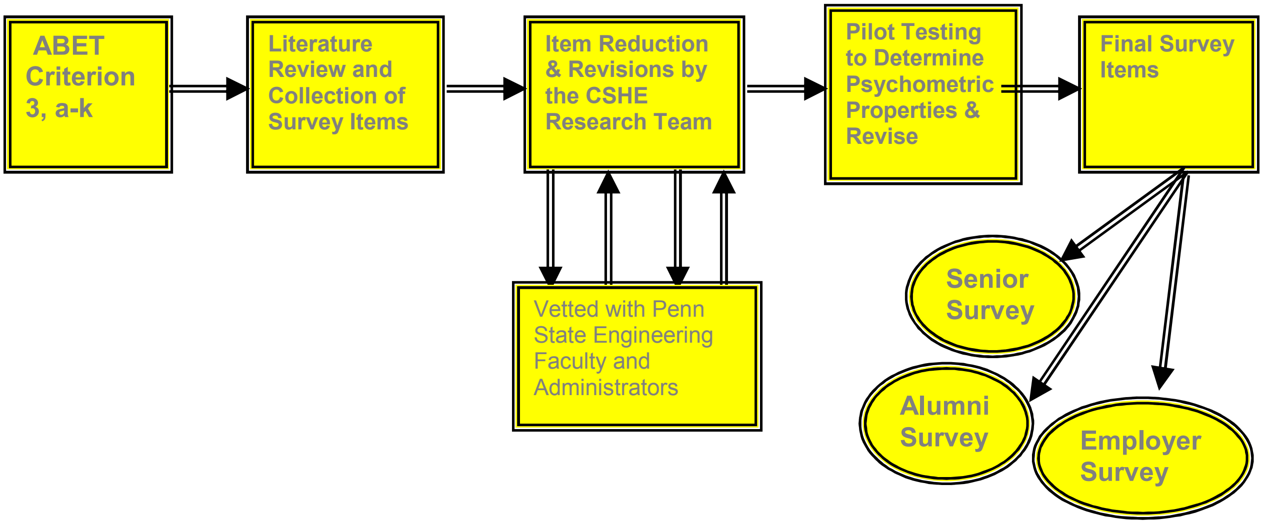

Figure 3. Correspondence of nine, factorially derived scales representing EC2000 learning outcomes.

EC2000 Criteria	**Factor Structure**
a. Apply knowledge of mathematics, Sciences, and engineering.	9. Applying Basic Skills
b. Design and conduct experiments, as well as to analyze and interpret data.	4. Experimental Skills
c. Design a system, component, or process to meet desired needs.	1. Design and Analytical Skills
d. Function on multi-disciplinary teams.	7. Group Skills
e. Identify, formulate, and solve engineering problems.	
f. Understand professional and ethical responsibilities.	3. Codes and Ethics
g. Communicate effectively.	5. Communications Skills
h. Understand the impact of engineering solutions in a global and societal context.	2. Societal and Global Issues
i. Recognize the need for, and engage in, life-long learning	8. Life-Long Learning
j. Know contemporary issues.	
k. Use the techniques, skills and modern engineering tools necessary for engineering practice.	6. Applying Engin. Skills

Table 1. Factor structure underlying items operationalizing EC2000 learning outcomes criteria.

Highest Loading Items	Number of Items	Factor Loadings	Scale Alpha	Variance Explained
1. Design and Analytical Skills	6		.92	12.0%
Design solutions to meet desired needs		.78		
Apply systematic design procedures to open-ended problems		.77		
Define key engineering problems		.76		
Formulate a range of solutions to an engineering problem		.75		
Understand essential aspects of the engineering design process		.68		
Apply discipline-specific engineering knowledge		.49		
2. Societal and Global Issues	5		.92	11.0
Understand contemporary issues (economic, environmental, political, etc.)		.80		
Understand that engineering decisions and contemporary issues		.79		
Understand the impact of engineering solutions in a societal context		.77		
Use knowledge of contemporary issues to make engineering decisions		.76		
Understand the impact of engineering solutions in a global context		.76		
3. Codes and Ethics	5		.87	8.4
Understand the engineering code of ethics		.79		
Consider ethical issues when working on engineering problems		.78		
Work through ethical issues in engineering		.78		
Understand technical codes and standards		.56		
Conduct yourself professionally		.48		
4. Experimental Skills	4		.89	7.8
Analyze evidence or data from an experiment		.79		
Interpret results of an experiment		.74		
Carry out an experiment		.74		
Design an experiment		.63		
5. Communication skills	4		.86	7.8
Convey ideas in writing		.78		
Convey ideas verbally		.76		
Convey ideas in formal presentations		.73		
Convey ideas in graphs, figures, etc.		.60		
6. Applying Engineering Skills	4		.94	7.5
Apply engineering tools in engineering practice		.75		
Apply engineering skills in engineering practice		.75		
Apply engineering techniques in engineering practice		.73		
Integrate engineering techniques, skills, and tools to solve real-world problems		.57		
7. Group Skills	3		.86	6.9
Work with others to accomplish team goals		.85		
Work in teams of people with a variety of skills and backgrounds		.83		
Work in teams where knowledge and ideas from multiple engineering disciplines must be applied		.65		
8. Life-long learning	3		.78	6.0
To what extent are you motivated to acquire and apply new technologies		.81		
To what extent are you willing to take advantage of new opportunities to learn		.80		
To what extent are you able to learn and apply new technologies and tools		.73		
9. Applying Basic Skills	2		.74	4.6
Apply knowledge of mathematics		.84		
Apply knowledge of physical sciences		.77		
				72.20%

Table 2. Intercorrelations among scores on nine scales reflecting EC2000 learning outcomes.

	Mean	S. D.	1	2	3	4	5	6	7	8	9
1. Design & Analysis Skills Scale	3.89	.71	1.00								
2. Social & Global Issues Scale	3.65	.82	.49	1.00							
3. Code & Ethics Scale	4.02	.74	.46	.65	1.00						
4. Experimental Skills Scale	3.91	.72	.73	.43	.42	1.00					
5. Communication skills Scale	4.00	.72	51	.55	.55	.51	1.00				
6. Applying Engineering Skills Scale	3.95	.77	.74	.58	.55	.61	.56	1.00			
7. Group Skills Scale	4.20	.72	.40	.47	.61	.36	.52	.47	1.00		
8. Applying Basic Skills Scale	4.05	.65	.69	.37	.33	.61	.40	.57	.29	1.00	
9. Life-Long Learning Scale	3.50	.53	.44	.33	.32	.39	.32	.45	.28	.38	1.00

All correlations are significant at the 0.01 level (2-tailed)