AC 2008-1851: APPLES AND ORANGES? A PROPOSED RESEARCH DESIGN TO EXAMINE THE CORRESPONDENCE BETWEEN TWO MEASURES OF ENGINEERING LEARNING

Patrick Terenzini, The Pennsylvania State University
Distinguished Professor of Education and Senior Scientist in the Center for the Study of Higher Education.

Lisa Lattuca, Pennsylvania State University
Associate Professor of Education and Senior Research Associate in the Center for the Study of Higher Education

Matthew Ohland, Purdue Engineering Education
Associate Professor and Director of First-Year Engineering in the Department of Engineering Education

Russell Long, Purdue University
Director of Project Assessment

Apples and Oranges? A Proposed Research Design to Examine the Correspondence Between Two Measures of Engineering Learning

Abstract

In 2004, ABET commissioned *Engineering Change*, a study of the impact of Engineering Criteria 2000 (EC2000) on the preparation of undergraduates for careers in engineering. One legacy of that study is a database of EC2000-specific self-reported student learning outcomes at 40 institutions, including precollege characteristics and engineering program outcomes for more than 4,300 graduates of the class of 2004. A second dataset, the Multiple-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD), compiles institutional data, including demographic and academic transcript records and Fundamentals of Engineering (FE) scores, from nine universities from 1987-2005. In this paper, we propose a design to combine data from the two databases to assess the correspondence between the self-reported student learning outcome measures in the *Engineering Change* study and the MIDFIELD dataset's information on program-level performance on the FE examination, the only objective test of students' engineering knowledge.

Introduction

Throughout its history, U.S. higher education has been mindful of questions about educational quality and institutional accountability. Formal accreditation mechanisms emerged in the early 20th century. Although the public has periodically engaged in these discussions, those who fund higher education – state and federal government, business and industry, and philanthropic foundations – have wielded the greatest influence.[1] Financial accountability is a dimension of these concerns, but the evaluation and assessment of educational effectiveness has emerged over the past two decades as an important corollary.

The current period of emphasis on accountability in the U.S. began in the 1980s and is roughly contemporaneous with expressions of heightened concern about the quality of engineering education programs and practices. The pressure for greater accountability, and the national conversations about the appropriate metrics for judging and ensuring educational quality that ensued, influenced the policy context for these discussions and the deliberations of accreditors. The Council for Higher Education Accreditation (CHEA), which recognizes individual accrediting agencies, now endorses assessment of student learning outcomes as one dimension of accreditation. Its endorsement, however, followed changes in the accreditation criteria in many regional and professional agencies that had already reduced their emphasis on quantitative measures of available resources and mandated that judgments of educational effectiveness be based on measurable outcomes.[2] Today, the higher education community generally accepts the need for assessment data to inform decision-making and acknowledges the need for rigorous methods that can provide this information to programs, colleges and universities, accreditation agencies, and state and federal governments.

This paper proposes a research design for a study of the correspondence between two publicly available assessment tools: the Fundamentals of Engineering (FE) examination and the student learning outcome scales developed for a national study of the impact of EC2000 on the preparation of undergraduate engineering students.

Measuring Student Learning

Although the importance of assessment in the reaccreditation process and in engineering schools' quality improvement efforts now appears to be well established, development of measures and instruments to assess student learning has lagged. Indeed, one might reasonably argue that the absence of rigorous designs and instruments may well be a major obstacle to engineering's efforts to improve and be responsibly accountable.

Currently, only one standardized, nationally normed instrument exists to assess learning in engineering. The Fundamentals of Engineering (FE) examination is used in licensing engineers throughout the U.S. The National Council of Examiners for Engineering and Surveying (NCEES), a national non-profit organization representing engineering and surveying licensing boards in all U.S. states and territories, develops, scores, and administers the FE examination (see http://www.ncees.org/). A central element in the first stage in the engineering licensing process, the FE covers material typically taught in undergraduate ABET-accredited programs in an array of engineering fields. The examination consists of 180 multiple-choice items administered during two, four-hour periods. The morning portion of the exam covers material common to all disciplines. The afternoon portion consists of a general examination or one specific to any of six fields (chemical, civil, electrical, environmental, industrial, and mechanical engineering).

The FE, however, is not a wholly satisfactory set of criterion measures for assessing desired engineering learning outcomes, whether those specified by EC2000,[3] the National Academy of Engineering's *The Engineer of 2020*,[4] or other national reports. The FE, for example, tends to concentrate on engineering subject-area and knowledge acquisition. Less attention is devoted to the engineering skills students may or may not have developed. Some have argued that FE scores are appropriate for assessing certain of ABET's EC2000 Criterion 3.a-k outcomes, specifically "Criterion 3: (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; (e) an ability to identify, formulate, and solve engineering problems; (f) an understanding of professional and ethical responsibility, and (k) an ability to use the techniques, skills, and modern engineering tools necessary for engineering practice".[5] The FE, however, does not address other skills specified by EC2000 or other reports (for example, the abilities to work in groups, communicate effectively, or recognize the interconnections between engineering solutions and economic, social, environmental, or cultural implications). Although there is potential that the FE outcomes are equally valid across a broad set of engineering disciplines, given the widely varying rates at which students in various disciplines take the exam, any set of test-taker data is likely to be more valid for some disciplines (for example, civil engineering) than for others (e.g., computer or biomedical engineering). This possibility is particularly likely given that validity is not a property of the instrument, but is instead related to the scores, which must be interpreted in context.[6]

ABET resisted rigid specification of what institutions must to in assessing their students' learning and discouraged reliance on any single measure. The consequence of the generality of ABET's specifications and the associated flexibility in operationalizing EC2000's Criterion 3 learning outcomes 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 these measures, however, appear to have been developed according to the instrument/test-development standards generally recommended. The absence of a widely used and broadly applicable set of measures of engineering learning outcomes has forced administrators and faculty members to wrestle with the challenges and compromises inherent in educational assessment. Because the institutional costs of accreditation and assessment can be substantial, institutions often turned to locally developed measures that make minimal demands on financial and staff resources. These efforts typically take the form of survey questionnaires in which students are asked to report how much progress they believe they have made in one content or skill area or another. These reports usually consist of one or more items intended to reflect the focal outcome. The variability of these local items and scales is considerable. One study's review of instruments used in studies published in archival journals or conference proceedings identified 286 outcome items that mapped to one of the 11 EC2000 outcomes. Between 20 to 40 survey items were associated with each of 11 a-k outcomes.[7] The item bank containing these items was subsequently reduced through editing, re-writing, and writing original items and became the foundation for developing nine factorially derived scales that (with two exceptions) map unambiguously to the EC2000 learning outcome criteria.

In addition to the measurement uncertainties evident in the wide variety of available items and instruments scattered throughout the research literature, engineering faculty and staff members also question the validity of student self-reports for measuring learning, challenges based on skepticism of students' abilities to evaluate themselves objectively and the absence of any demonstration of what has been learned on some standardized measure or in some hands-on assignment.

Considerable research over the past 30 years has examined the correspondence between self-report based measures of learning and skill development and objective measures of the same traits or skills. Although results vary depending on the traits examined and the measures used, these studies report correlations of .50 to .70, on average, between self-reports and such objective criterion measures as the ACT Comprehensive Test, the College Basic Academic Subjects Examination, and the Graduate Record Examination.[8,9,10,11,12] For example, Pike[11] 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[8] 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. In a meta-analysis of 44 studies comparing self-reported versus actual grades and test scores, Kuncel, Crede, and Thomas report correlations between the two of .90 for college GPA and .82 for total SAT score.[13] The study suggests that only the very lowest performing students significantly misrepresented their grades and scores.

Hayek, Carini, O'Day, and Kuh[14] and Kuh[15] concluded that, all things considered, self-reports are likely to be valid and appropriate for use in quality assurance and performance improvement systems provided five general conditions are met: 1) the information requested is known to the respondents; 2) the questions are phrased clearly and unambiguously; 3) the questions refer to recent activities; 4) the respondents think the questions merit a serious and thoughtful response; and 5) answering the questions does not threaten, embarrass, or violate the privacy of the respondent or encourage the respondent to answer in socially desirable, rather than in truthful ways. Under these conditions, they suggest, student self-reports are both valid and reliable, especially for measuring the outcomes for groups of students.

Study Purpose

The centrality of the a-k learning outcomes in accreditation and in quality assurance and accountability self-studies continues to drive engineering assessment efforts to identify or develop cost-effective measures of student learning outcomes. Despite engineering education assessment's widespread reliance on students' self-reports of learning gains, studies of the correspondence between such self-reports and more objective measures of engineering student learning are missing from the literature. This study will seek to rectify that situation by examining the correspondence between the self-reported learning outcome measures developed for the Engineering Change (EC) study, which examined the impact of the implementation of EC2000 outcomes-based accreditation criteria, and FE scores for five institutions. At the time the EC2000 study was conducted, FE scores were unavailable to test the criterion validity of the measures using an objective test of learning in engineering. The opportunity to do so, however, is now at hand. In exploring that correspondence, the study seeks to expand understanding of the measurement characteristics and correspondence between measures now in wide use or rapidly gaining prominence.

Data Sources

The methodology to be used in this study must be carefully designed if it is to be credible and provide useful psychometric insights on widely used measures of engineering learning, as well as on the validity of self-reported learning outcome measures in educational research. In this paper, we thus present an analytical plan for evaluating the criterion-related validity of the scales developed for the EC2000 study using data available through the Multi-Institution Database for Investigating Engineering Longitudinal Development (MIDFIELD) project, including student FE examination scores and grade-point averages (from transcript records). Presentation and discussion of the study design and procedures at ASEE will permit the engineering education research community to vet the study before it is conducted and to recommend improvements for the final study design. Findings from the study, thus, will be presented in subsequent publications.

The Engineering Change Database. In 2002, ABET commissioned a study of the effects of the introduction of new accreditation criteria, *Engineering Criteria 2000*.[3] The new standards shifted the basis for reaccreditation from a resources- to outcomes-based model. EC2000 required programs to demonstrate their graduates' achievement in 11 engineering skill areas. The large national assessment that ensued defined the population for the study to include all undergraduate engineering programs holding ABET accreditation since 1990 or earlier in seven engineering fields. This disciplinary array permitted study of the disciplines that produce the vast majority of undergraduate engineering degrees (chemical, civil, electrical, industrial, and mechanical), as well as disciplines with strong ties to industry sectors (aerospace and computer). Together, these programs award 82 percent of all undergraduate engineering degrees conferred. Of the 1,241 currently ABET-accredited engineering programs in the targeted disciplines, 1,024 were accredited in 1990 or earlier. The sampling design produced a dataset based on a two-stage, 7x3x2, disproportional, stratified random sample. In the first stage, 40 institutions offering at least two of the seven focal programs were randomly selected from within three strata: seven engineering disciplines, three levels of EC2000 adoption status (early, on-time, deferred), and two levels (yes or no) of participation in an NSF-funded engineering education coalition. To ensure an adequate number of responses for analysis within each discipline, institutions with programs in the smaller disciplines were over-sampled, and Historically Black and Hispanic-Serving institutions were purposively included.

All graduating seniors in the 40 institutions in the targeted disciplines received an invitation to participate in the study. The Graduating Senior Survey solicited information on a wide array of topics, including basic demographic information, level of participation in out-of-class activities related to engineering education, self-reported student-learning outcomes associated with each of the 11 outcomes criteria, classroom experiences, and plans for the future. Usable responses were received from 4,330 graduating seniors (a 36% response rate) on 39 campuses. Weighting procedures subsequently adjusted for unrepresentativeness introduced by disproportional sampling and for minor response bias related to students' sex, race/ethnicity, and discipline.

Instrument Development. A lengthy instrument development process (including a review of the literature and relevant instruments in use, item writing and editing, discussions with engineering faculty members, pilot testing, factor analysis, and item analyses) led to the development of 50 items to operationalize the 11 EC2000 learning outcomes criteria. Responses to 36 items in the national survey produced a set of factorially derived (principal components) scales corresponding to nine of the eleven EC2000 outcome criteria. Respondents indicated their level of achievement on each of the 36 items using a 5-point scale, where 1 = “No Ability” and 5 = “High Ability.” The final, nine-factor solution retained 75.3 percent of the overall item variance among the 36 survey items. Scales were formed using only those items loading above .40 on a factor, and no item loaded above that standard on more than one factor.[16]

The nine outcomes scales measure students’ knowledge and abilities in 1) math and science (a two-item scale with Cronbach’s alpha = . 74), 2) experimental skills (four items assessing the ability to design and carry out an experiment; alpha =. 89), 3) engineering skills (four items assessing abilities to use engineering tools and skills in practice; alpha = .94), 4) design and problem-solving (a six-item scale assessing students ability to solve open-ended problems and design solutions; alpha = .92), 5) communications skills (four items measuring abilities to convey ideas in writing, verbally, and in graphs; alpha = .86), 6) group skills (three items assessing the ability to work with others; alpha = . 86), 7) knowledge of societal and global issues (a five-item scale measuring awareness and understanding of societal contexts and contemporary issues; alpha = .92), 8) ethics and professionalism (five items assessing understanding of standards and codes; alpha = .87), and 9) life-long learning (three items measuring motivation to continue to learn, alpha = . 78). More complete descriptions of the instrument and scale development processes are given in Strauss and Terenzini[17] and Lattuca, Terenzini, and Volkwein.[7]

Five of the EC2000 study institutions are also part of the MIDFIELD study (Clemson, Georgia Tech, North Carolina A&T, University of Florida, and Virginia Tech). All are doctoral-level research universities. Across those five institutions, 216 (44%) of the 491 seniors graduating in 2004 who participated in the EC2000 study reported they had also taken the FE examination. Of the resulting 23 programs offered by the five institutions, 14 programs have five or more respondents in the EC2000 database. Because (as will be seen below) the unit of analysis for the study must be programs, 14 programs will not provide sufficient statistical power. Consequently, study team members are seeking additional institutions in the EC2000 study to add to the MIDFIELD database. Such an augmentation will perhaps double the number of institutions and programs used for analyses.

The MIDFIELD Database. An outgrowth of the Southeastern University and College Coalition for Engineering Education (SUCCEED), MIDFIELD currently includes unit-record data from all undergraduate, engineering degree-seeking students at nine public universities in the southeastern United States. An earlier version of the database was compiled in 1996, and the current version was compiled starting in 2004, adding data on courses taken and grades earned for all students, expanding the database to include full transcript records. The database includes records from 1987-2005, although some partner institutions are updated only to Spring 2004. All participating institutions are doctoral granting research institutions. MIDFIELD has 69,776 first-time-in-college students.[18] Engineering students are overrepresented at the MIDFIELD institutions. In 2005, 330 institutions nationwide enrolled 310,022 undergraduate engineering students,[19] and engineering students represented approximately 10% of the 3,320,249 students enrolled in all majors at those institutions.[20] However, within the MIDFIELD data, over 20% of students matriculate in engineering. This is in part because MIDFIELD includes six of the 50 largest U.S. engineering programs (measured by undergraduate enrollment), resulting in a population that includes approximately one-twelfth of all engineering graduates of U.S. engineering programs annually. The percentage of women and of Latinos (regardless of gender) among these engineering graduates is representative of other U.S. programs. African-American students, however, are significantly overrepresented in the MIDFIELD dataset, as the MIDFIELD participants include four of the top five producers of African-American engineering graduates, including two HBCU; together, all partner schools graduate one-fifth of all U.S. African-American engineering B.S. degree recipients each year. These ratios are taken from the most recent data available; the exact percentages vary from year to year.[19]

Some examinations have KR_{20} statistics with a range higher or lower than other exams primarily because of the homogeneity or heterogeneity of the knowledge domain being tested and the length of the examination. Kuder-Richardson (KR_{20}) coefficients reflect the internal consistency reliabilities of scales derived from a set of dichotomously coded items.) KR_{20} values above .80 are considered entirely adequate (the statistic can vary from .00 to 1.0). The minimum acceptable KR_{20} value is .70. If a KR_{20} value approaches .70, the items of the exam are reviewed for irregularities, and the length of the examination is evaluated. If the KR_{20} drops more that .05 below its historical average for an examination, sources of unreliability are considered. Possible explanations include flawed items, trick questions, misleading item wording, unsuitable illustrations, guessing taking place, inadequate directions, and irregularities in administration.[21]

MIDFIELD studies usually involve large sample sizes,[22] but the associated data on FE test-takers is limited by the number of years and disciplines studied, by the test-taking rate in various disciplines, and by aggregation of the unit of analysis to the program level.

Proposed Analytical Plans

The EC2000 study dataset has students as the unit of analysis; MIDFIELD's FE data are at the program level. Consequently, the first step in the analytical process will be to aggregate EC2000 student records to the program level within each institution. Covariate-adjusted means (and standard deviations) for each of the nine EC2000 scales will be calculated and added to the EC2000 study database. Covariates will include students' sex, high school grade-point average, SAT/ACT scores (using a common metric), engineering discipline, response rates within program, and the percentage of students in a program who took the FE examination.

Following the merger of program-level data from the two studies, zero-order correlations will be used to assess the correspondence between EC2000 and FE scores. Several measures will be correlated. Variables from the EC2000 dataset will include all nine outcome scale scores. FE variables in the analyses will include average scaled scores (by program) for items that align with a particular EC2000 criterion and the percentage of FE takers who pass the examination (both by program and institution-wide). Finally, the standard deviations of EC2000 measures and FE scaled measures will be examined to evaluate the dispersion of specific sets of scores.

The Proposed Study's Implications

This paper has described the design and methods proposed for a study of the program-level correspondence between engineering students' self-reported learning gains and their performance on the Fundamentals of Engineering examination, standardized and nationally normed measure of engineering learning. The study must be carefully designed if it is to be credible and provide useful psychometric insights on widely used measures of engineering learning, as well as on the validity of self-reported learning outcome measures in educational research.

The study's findings will be of great interest to those who conduct research on the outcomes of undergraduate engineering education, as well as engineering administrators and faculty who must conduct assessment as a condition of accreditation. Both communities require evidence of the validity of publicly-available measures such as the EC2000 student outcomes scales. Each community also has a stake in ensuring that the research procedures used to establish the validity of widely-available self-report measures are rigorous and complete.

References

[1] Stark, J. S., & Lattuca, L. R. *Shaping the college curriculum: Academic plans in action*. Boston: Allyn & Bacon.

[2] Volkwein, J. F., Lattuca, L. R., Caffrey, H. S., & Reindl, T. *What Works to Ensure Quality in Higher Education Institutions.* AASCU/CSHE Policy Seminar on Student Success, Accreditation, and Quality Assurance. Washington, DC: AASC&U and Penn State University, 2003.
[3] ABET. *Engineering criteria 2000.* Baltimore: Author, 1997.
[4] National Academy of Engineering. *The engineer of 2020: Visions of engineering in the new century*. Washington, DC: National Academies Press, 2004.
[5] LeFevre, W., Steadman, J. W., Tietjen, J. S., White, K. R., & Whitman, D. L. (2005). *Using the fundamentals of engineering (FE) examination to assess academic programs*. Clemson, SC: National Council of Examiners for Engineering and Surveying, n.d., p. 2.
[6] Messick, S., "Standards of validity and the validity of and standards in performance assessment," *Educational Measurement: Issues and Practice*, *14* (4), 1995, pp. 5–8.
[7] Lattuca, L. R., Terenzini, P. T., & Volkwein, J. F. Engineering Change: Findings from a Study of the Impact of EC2000, Final Report. Baltimore, MD: ABET, Inc., 2006.
[8] Anaya, G. College impact on student learning: Comparing the use of self-reported gains, standardized test scores, and college grades. *Research in Higher Education, 40,* 1999, pp. 499-526.
[9] Dumont, R. G., Troelstrup, R. L. Exploring relationships between objective and subjective measures of instructional outcomes. *Research in Higher Education, 12*, 1980, pp. 37-51.
[10] Hansford, B.C., & Hattie, J.A. The relationship between self and achievement/performance measures. *Review of Educational Research*, *52*, 1982, pp. 123-142.
[11] Pike, G. R. The relationships between self-reports of college experiences and achievement test scores. *Research in Higher Education, 36*, 1995, pp. 1-22.
[12] Pike, G. R. Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education*, 37, 1996, pp. 89-114.
[13] Kuncel, N. R., Crede, M., & Thomas, L. L. The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*(1), 2005, pp. 63-82.
[14] 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, 43*(5), 2002, pp. 643-63.
[15] Kuh, G. D. Imagine asking the client: Using student and alumni surveys for accountability in higher education. In J. C. Burke & Associates (Eds.), *Achieving accountability in higher education: Balancing public, academic, and market demands* (pp. 148-172). San Francisco: Jossey-Bass, 2005.
[16] Armor, D. J. Theta reliability and factor scaling. In H. Costner (Ed.). *Sociological methodology: 1973-1974 (pp. 17-50).* San Francisco: Jossey-Bass, 1974.
[17] Strauss, L. C., & Terenzini, P. T. Assessing student performance on EC2000 Criterion 3.a-k. *Proceedings of the 2005 American Society for Engineering Education Annual Conference and Exposition.* Washington, DC: American Society for Engineering Education, 2005.
[18] U.S. Department of Education. "The integrated postsecondary education data system (IPEDS) glossary." Retrieved December 12, 2007, from http://www.nces.ed.gov/ipeds/glossary/.
[19] American Society for Engineering Education. "Engineering data management system." Retrieved December 12, 2007, from http://www.asee.org/datamining/reports/.
[20] U.S. Department of Education. "IPEDS educational peer tool and peer analysis system." Retrieved December 12, 2007, from http://nces.ed.gov/ipedspas/.
[21] National Council of Examiners for Engineering and Surveying. *NCEES standards development procedures manual*. Clemson, SC: Author, 2007.
[22] Ohland, M.W., S.D. Sheppard, G. Lichtenstein, Ö. Eris, D. Chachra, and R.A. Layton, "How Can We Teach Them If They Aren't There? Matriculating, Engaging, Persisting, and Migrating in Engineering," *in press*, invited paper in "How People Learn Engineering," special issue, *J. Eng. Ed. 97*(2), April 2008.