2018 ASEE Annual Conference & Exposition
June 24–27, 2018 Salt Palace Convention Center Salt Lake City, UT

Paper ID #23336

Critical Analysis of the Validity of the Fundamentals of Engineering Mechanical Exam

Dr. Alex C. Szatmary, King's College

Alex Szatmary teaches mechanical engineering in a new program at King's College, Wilkes-Barre. In his technical research, he uses mathematical models to study how cells get to places in the body.

Critical analysis of the validity of the Fundamentals of Engineering Mechanical Exam

Abstract

The purpose of the Fundamentals of Engineering Exam (FE) is primarily to ensure that Engineers in Training possess at least minimal competency on a broad range of engineering subject matter. It is secondarily used in curricular assessment; for this purpose, it has been suggested that it has particular value as the only nationally-normed exam on engineering. Here, I argue that the FE is not suitable for curricular design or assessment. The FE has not been shown to be a valid instrument and there is evidence in the literature that it is not. Furthermore, it is unclear, from the Exam Specifications alone, what content the exam actually assesses. To clarify how competency in subjects is operationally defined on the FE Mechanical, I wrote learning objectives based on questions from the FE Mechanical Practice Exam on Computational Tools, Fluid Mechanics, and Mechanical Design and Analysis. I show that the practice exam is not content valid with respect to the exam specifications or the FE Reference Handbook. Furthermore, the questions are posed at a low cognitive level and are mostly an exercise in applying simple formulas so performance on the FE may have more to do with motivation, test-taking skills, and familiarity with the handbook than competency in engineering. Furthermore, the FE is poorly aligned with ABET student outcomes; in addition to the poor validity of the exam, evidence of student performance on the exam is too coarse to be useful for continuous improvement.

Introduction

The Fundamentals of Engineering (FE) Exam is a nationally-normed assessment of engineering competency across a broad range of subject matter [1]. It is administered by the National Council of Examiners for Engineering and Surveying (NCEES). The FE is typically taken around the time of graduation; passing the FE is a requirement for becoming an Engineer in Training in the United States. After several years of work experience as an EIT, an engineer is eligible to take the Principles and Practice of Engineering exam (PE), which is a requirement for licensure. FE exams are available for chemical, civil, electrical and computer, environmental, industrial, and mechanical engineering; there is also an “other disciplines” exam. Some subjects, such as Mechanical Design and Analysis, appear only on the FE Mechanical; other subjects include different topics and are assessed differently on each discipline’s exam.

Although the FE is designed to be a step in the licensure process, it is also frequently used in program assessment [2]. Programs can get reports from NCEES indicating how their students did, on average, in each subject area on the FE.

A program might find it desirable to align curriculum with the FE for several reasons. First, the FE Exam Specifications [3] list topics within subjects and thus may seem like a simple starting point when deciding on course content. Also, ensuring coverage of FE topics might help students to do well on the exam. Finally, the FE is only useful for assessment to the extent that the content that it assesses corresponds to the content of the curriculum. Anecdotally, some faculty consider the FE when deciding what to cover in a course. Courses have been added to curricula to improve FE scores [4, 5, 6]. The technical content of several programs (typically general engineering programs at state universities) was selected from the beginning to cover the FE [7, 8, 9].

This project arose in a new mechanical engineering program. My intention was to use the FE Exam Specifications as a single simple source on what one might expect in a mechanical engineering program. The FE also seemed like a straightforward, objective assessment. However, as I tried to develop learning objectives that comprehensively covered FE Mechanical content, I found it necessary to consult multiple sources beyond the Specifications. Furthermore, seeing how competency is assessed on the FE Practice Test raised questions for me regarding the FE as an assessment. This paper focuses on the FE Mechanical but these findings are broadly to be applicable to each version of the FE; throughout, I use "FE Mechanical" to refer to the particular exam and "FE" to refer to all versions of the exam.

For the FE to be useful in assessment, its psychometric reliability and validity must be demonstrated; furthermore, content validity is a useful frame for analyzing what the FE covers. Lawson [10] previously analyzed the FE in terms of psychometric reliability and validity. I could not find other analyses of the FE as a psychometric instrument. Others have described how they use the FE in assessment [4, 5, 6, 11]; however, their work assumes (rather than tests) the validity of the FE for this purpose. Lawson shows that although the FE may be valid for licensure purposes, it is not always valid for assessment of programs, and it is completely invalid for assessment of engineering schools; therefore, attaching performance funding to FE pass rates is inappropriate. Lawson also calls for further critical analysis of the FE.

In this paper, I review how the FE is formulated and scored and discuss how it is currently used in assessment. I show, from the literature, that we have reasons to doubt the validity of the FE. The content validity of the FE has not been discussed previously in the literature. Content validity must be determined subjectively. Therefore, I analyze how the FE Mechanical treats three specific subjects: Computational Tools, Fluid Mechanics, and Mechanical Design and Analysis. In an attempt to distill FE content into a single document, I wrote learning objectives corresponding to FE Mechanical Practice Exam questions for these topics. I show that the FE exam assesses topics that are not on the exam specification, that the exam specification is too vague to be useful, and that the FE content is probably more influenced by the handbook than by the exam specifications. Furthermore, some questions are badly written and FE questions assess engineering competency at a low cognitive level; these features undermine the construct and criterion validity of the instrument. Because assessment should be done in light of the objectives for a particular program, I also explain why the FE is not a suitable assessment for my program.

Finally, I argue that the FE is not useful for assessment for accreditation through ABET because, in addition to simply being a poor assessment, it does not assess most ABET student outcomes and it provides information that is too coarse to be useful for continuous improvement.

The FE is developed and evaluated through an opaque process

The FE Exam Specifications are developed by surveying engineering faculty members [1]. The structure and contents of this survey are not publicly known. It is not known how the content of the FE Reference Handbook was chosen. FE questions are written by volunteers who are "licensed practitioners and engineering educators" [1]. The NCEES Exam Development Procedures Manual describes how the exam is developed [12], but this document is not publicly available. Note that industry has no direct influence on the content of the exam.

The FE Reference Handbook is the only source of information that test-takers can consult during the exam [13]. Numerous items on the FE Exam Specifications, such as spreadsheets, sustainability, and hydraulic, pneumatic, and electromechanical components are not described at all in the Handbook. If something is not in the FE Reference Handbook, it is probably not on the exam [14]. Because the Handbook affects the content of the exam, how content is selected for the Handbook should be described.

Passing the FE is criterion-referenced in the sense that passing depends on how test-takers do relative to judgments by professional engineers of how well minimally-competent engineers should do on FE questions [10]. However, NCEES rightly discourages the use of raw pass rates in program assessment [1] and instead provides programs with subject matter reports which show a "performance index" for each subject; the performance index is not directly related to the raw number of questions that students answer correctly. The intention behind this reporting method is that it is a necessary way to address the facts that everyone answers a different set of questions, with different numbers of questions from each subject, and with different levels of difficulty. It is not possible to correlate performance indices with the criteria that are used to determine passing. Instead, NCEES recommends analyzing subject matter reports not by looking at just raw performance index values but by comparing the performance index with that of programs in a comparison group. In short, when used for program assessment, the FE is a norm-referenced measure. As described in the following section, this limits the face validity of the FE. Because subject matter reports are not given with respect to objective criteria, a program cannot use the FE to determine whether their students meet a desired level of performance.

Taken together, the lack of transparency from NCEES is troubling. For example, one might think that the FE Exam Specifications are consensus descriptions of what should be covered in undergraduate engineering programs. However, reported survey results depend on how questions are asked and how results are interpreted. Outcomes involving design of experiments, the Navier-Stokes equations, static equilibrium in 3D, and teamwork could be assessed to some extent on the FE, but they are not; is this a reflection of the consensus of faculty surveyed or a consequence of the survey process? How do we confirm that the content and emphasis of the FE reflects the opinions of the surveyed faculty members? Additionally, how are the specifications translated into questions? How is the handbook developed? Is it involved in determining exam content? What do the numbers on the subject matter reports even mean?

The FE might be reliable but it is not valid

Overview

Psychometric tests are evaluated based on their reliability and validity. Reliability means that the results are consistent over time and with different judges. Validity means that the test measures what it purports to measure. Below, I describe face, content, construct, and criterion validity and how they apply to the FE. The FE has two main uses: to confirm that EITs have at least minimal competency and to assess programs. For the first purpose, whether a person passes the FE is the key measure; for program assessment, subject matter reports indicating the performance of students on each FE subject are more useful. Therefore, this paper focuses on the reliability and validity of subject matter reports for a program's students.

Reliability

The FE exam is probably reliable but its reliability has not been thoroughly established. The FE is machine-graded so the scoring process is consistent. Reliability goes beyond scoring consistency, however. The reliability of the FE could be demonstrated more fully by showing whether a student's scores within a subject are correlated (internal reliability) as well as by having people re-take the FE and compare their scores between offerings (test-retest reliability); such studies have not been reported.

Face validity

Face validity means that a test appears to measure what it is intended to; because the FE is about engineering and it is intended to measure competency in engineering, the FE exam probably has some face validity. Lawson notes that FE scores are criterion-referenced, which contributes to the face validity of this measure [10]; however, subject matter reports are essentially norm-referenced. Therefore, the face validity of the FE (as it is used in assessment) is limited because subject matter scores do not show whether students are competent at the level set by the program.

Content validity

Whether a given subject matter report score is valid depends on how that subject is "operationalized" on the FE. What does it mean to be minimally competent in a given subject, say, statics? There are a variety of opinions: does someone have to be able to represent a real-life system as a set of free-body diagrams? To solve 3D problems involving rigid bodies? To be able to use computational tools to determine forces in structures? To account for friction or hydrostatic forces? The way that a subject is assessed on the FE represents an operational definition of what constitutes minimal competency in that subject. Understanding how engineering competency is operationalized on the FE requires analyzing what the FE measures (content validity) and how well it measures (construct validity).

Content validity means that a test assesses all aspects of what is being measured; in this paper, content validity is defined to also mean that a test does not measure content that it is not stated to measure. Below, I analyze FE materials and show that the FE does not have content validity. In short, FE assessment of subjects—that is, the FE's operational definitions of those subjects—is not comprehensive. Therefore, high scores in a subject might not indicate mastery of a subject if that subject is narrowly assessed on the FE.

The FE might assess things that are not, in fact, essential content for a given subject. If a program's students were to have a low score in a subject, this might be because the FE assesses things that are esoteric or not important to that particular program. Unfortunately, FE subjects are specified vaguely so it is unclear, from the Exam Specifications alone, what content is being assessed. These issues of content validity have to be analyzed subjectively based on available information about the FE.

Construct validity

Construct validity means that a test measures what it means to measure. Construct validity should be demonstrated using empirical, quantitative data to show that the measure is correlated with related measures and not unrelated ones. The only statistical evidence in the literature for the FE having construct validity compares FE performance across institutions for specific subjects [11]; differences in subject scores were attributable to whether students were required to take a corresponding course prior to taking the FE. FE scores in particular subjects should correlate with student grades in corresponding courses but Watson et al. found no such correlation [15]. There are mixed reports on whether individual FE scores correlate with student grade point average (GPA) [15, 16]. Because few validation studies have been done on the FE, one of the ways we can estimate its construct validity is to see how students do on mock FE exams. In one such study, mock FE scores correlated with GPA in mechanics courses but there was a weak correlation between scores on specific subjects and grades in corresponding courses [17]. This is the result one would expect if the FE were a measure of general intelligence more so than engineering competency.

Although construct validity must be established empirically, it can also be analyzed subjectively. The construct validity of the FE is further weakened because FE performance is apparently affected by several other factors. For example, student motivation to pass the FE exam is a major factor in exam performance [18]. Having taken an FE review course can raise the FE pass rate from about 80% to 95% [19, 20]. Miller [21] has suggested that familiarity with the FE Reference Handbook is probably a major factor in success on the FE.

Because the students who take the FE are typically a self-selected group, it is necessary to show that the FE results are representative of a program's entire student population; the only way to guarantee that FE results are representative is to require all students to take the FE as a graduation requirement [15]. On the other hand, if taking the FE is a graduation requirement, the validity of the results can be even worse unless a large majority of the students are motivated to put effort into preparing for and taking the exam [10].

Because the test uses multiple-choice questions, it probably assesses at low cognitive levels [10]. Faculty members are often surprised by how simple FE questions are [15]; the questions mostly involve finding formulas in the FE Reference Handbook and doing "plug and chug" [22]. With that said, the overall score needed to pass the FE was about 55% in 1998 [15] and is rumored to still be in the range of 50-55% [14]. If the FE is so easy, why are the scores so low? There are a few possible reasons. The exam must be taken quickly, allowing an average of 2.9 minutes per question. Questions could be worded in a way that is confusing to students. Many of the questions are based on the FE Reference Handbook which uses non-standard notation [21].

FE scores are not a pure measure of engineering competence because they also indicate motivation and probably general intelligence. To do well on the FE, it helps to be familiar with the FE Reference Handbook. Motivation to pass the FE would lead a student to take the time to become familiar with the many idiosyncratic features of the exam. The exam is fast-paced but passing only requires getting around half of the questions correct. Therefore, a prudent test-taker would make educated guesses on challenging questions and focus on solving easy questions correctly. General intelligence would probably help a test-taker make sense of the notation, figure out formulas and content that would not have been covered in coursework, quickly make guesses on multiple-choice questions with obvious distractors, and be strategic while taking the exam.

Criterion validity

Criterion validity means that a test has predictive power. It would be interesting if FE scores predicted performance of engineers in the real world but no study along these lines has been reported. Given that the FE mostly assesses ability to perform simple calculations, it seems unlikely that it is a good predictor of real-world engineering competency, which involves complexity and ambiguity, and requires broad professional skills [23, 24].

Summary

The FE is not a sound psychometric instrument. Its reliability has not been demonstrated and there are reasons to doubt its validity. Although criterion-referenced data is most useful for program assessment, FE scores on subject matter reports are norm-referenced, limiting the face validity of the FE. FE results are affected by motivation and possibly general intelligence, hindering its construct validity. It has not been shown to be predictive of engineering performance so it lacks criterion validity. The focus of the following analysis is to illustrate problems with the content validity of the FE and how engineering competency is operationally defined.

Content validity analysis

Content validity needs to be established subjectively, that is, by critically analyzing the content measured with a particular instrument. Content validity analysis of the FE has not been previously reported. Here, I present critical analysis of the FE topics Computational Tools, Fluid Mechanics, and Mechanical Design and Analysis. Because FE questions are confidential, an NCEES-developed practice exam [25] must suffice as a proxy for the actual exam. The FE Exam Specifications are somewhat broad and vague so one might turn to the FE Reference Handbook [13] as another way of guessing what might be on the exam. Thus, I address the questions of whether the FE Mechanical Practice Exam is content valid with respect to the exam specifications and the handbook, for these particular topics. Subsequent content validity analysis of other topics should be performed.

The exam specifications for Computational Tools, Fluid Mechanics, and Mechanical Design and Analysis are shown in Tables 1, 3, and 5, respectively. To develop an understanding of how each

Table 1: FE Mechanical Exam Specifications for Computational Tools

A. Spreadsheets
B. Flow charts

Table 2: Learning objectives based on FE Mechanical Practice Test questions on Computational Tools

Question	Learning objectives
10	Predict the contents of a spreadsheet cell when copied with automatic cell referencing.
11	Predict the result of algebraic calculations in spreadsheets.
12	Interpret pseudocode that employs if statements and loops.
13	Interpret flowcharts.

subject is operationally defined on the FE, I wrote learning objectives to describe the competencies assessed in each FE Mechanical Practice Exam question on these subjects; the learning objectives are shown in Tables 2, 4, and 6. Selected problems are quoted in Appendix A.

The FE Mechanical Exam Specifications are vague. For Computational Tools, "spreadsheets" and "flow charts" are broad categories. The Mechanical Design and Analysis specifications do not state which machine elements a test-taker should be able to do stress analysis on, what one should be able to do regarding power transmission, or what kinds of joining methods are covered. The Fluid Mechanics specifications are particularly unclear: what fluid properties must be known? What should a test-taker be able to do regarding power and efficiency? The categories "internal flow" and "external flow", or "incompressible flow" and "compressible flow" do not clarify things because all flows are internal or external, and incompressible or compressible. Which flows should be analyzed? With what methods?

Turning to the FE Reference Handbook for more detail is unsatisfactory because it does not describe several topics; others are addressed only briefly or in sections of the handbook for other disciplines. There is not a chapter on Computational Tools in the handbook, spreadsheets are

Table 3: FE Mechanical Exam Specifications for Fluid Mechanics

A. Fluid properties
B. Fluid statics
C. Energy, impulse, and momentum
D. Internal flow
E. External flow
F. Incompressible flow
G. Compressible flow
H. Power and efficiency
I. Performance curves
J. Scaling laws for fans, pumps, and compressors

Table 4: Learning objectives based on FE Mechanical Practice Test questions on Fluid Mechanics

Question	Learning objectives
55	Use formulas for specific gravity and kinematic viscosity.
56	Recall the definition of viscosity.
57	Find the definition of Archimedes' principle in the handbook.
58	Calculate the magnitude of a hydrostatic force on a rectangular face.
59	Calculate the velocity of a jet discharging from an orifice.
60	Calculate the resultant force acting on a control volume with uniform inlet and outlet velocities.
61	Calculate the Reynolds number. Use definitions of flow rates.
62	Calculate the Reynolds number.
63	Calculate volumetric flow rate.
64	Calculate the efficiency of a fan given pressure drop, flow rate, and power.

Table 5: FE Mechanical Exam Specifications for Mechanical Design and Analysis

A. Stress analysis of machine elements
B. Failure theories and analysis
C. Deformation and stiffness
D. Springs
E. Pressure vessels
F. Beams
G. Piping
H. Bearings
I. Power screws
J. Power transmission
K. Joining methods
L. Manufacturability
M. Quality and reliability
N. Hydraulic components
O. Pneumatic components
P. Electromechanical components

Table 6: Learning objectives based on FE Mechanical Practice Test questions on Mechanical Design and Analysis

Question	Learning objective	Handbook section
93	Calculate stress in a beam subject to bending.	Mechanics of Materials
94	Calculate the moment of inertia using a table. Calculate stress in a beam subject to bending. Calculate alternating and mean stress. Calculate factor of safety in fatigue using the modified Goodman theory.	Statics, Mechanics of Materials, Mechanical Design and Analysis
95	Calculate stress in a linear helical spring.	Mechanical Design and Analysis
96	Use Hooke's Law.	Mechanics of Materials
97	Calculate the tangential stress in a thin-walled pressure vessel.	Mechanics of Materials
98	Calculate strain under compound loading.	Mechanics of Materials
99	Calculate the tangential stress in a thin-walled pressure vessel.	Mechanics of Materials
100	Calculate strain due to thermal expansion.	Mechanics of Materials

absent from the handbook altogether, and flowcharts are in the chapter on Electrical and Computer Engineering. Mechanical Design and Analysis topics E and G are in the Mechanics of Materials chapter, M is in Industrial Engineering, and N, O, and P are not discussed at all.

None of these subjects is content valid with respect to either the exam specifications or the handbook. On the practice exam, one of the four Computational Tools questions (#12) purports to be about a flowchart; the so-called flowchart is actually pseudocode. For Fluid Mechanics, two questions involve calculating the Reynolds number; however, dimensionless parameters are not on the specifications. Several topics topics on the Fluid Mechanics specifications (E, G, I, J) are not assessed. Most topics in the fluid mechanics chapter of the handbook are not addressed.

The practice exam assesses Mechanical Design and Analysis in a strange way. There are 8 questions on mechanical design and analysis. Of the 16 topics listed on the specification, 6 (A-F) are assessed. Also, question 100 assesses thermal expansion, which is not listed anywhere in the exam specification and appears in the mechanics of materials section of the handbook. On the other hand, if we were to take the FE handbook's chapter on Mechanical Design and Analysis to operationalize this area, we would find that only 2 of the 8 questions assess mechanical design competency; instead, most of the questions use formulas that are in the Mechanics of Materials chapter. The exam specification, practice exam, and handbook are all in disagreement: the practice exam does not assess most of what is on the specifications, assesses content that is not on the specifications, and mostly assesses mechanics of materials rather than mechanical design and analysis. Furthermore, numerous topics present in the exam specification (M, N, O, and P) are absent from both the practice exam and the handbook so there is no way of determining how these topics might be assessed.

In addition to the problems with content validity, direct analysis of the practice exam indicates problems with construct validity. First, most questions just require applying simple formulas and definitions from the handbook, which suggests that exam performance depends more on familiarity with the handbook than with mastery of concepts. Furthermore, some of the questions are posed in a way that rewards test-taking skills or familiarity with the handbook rather than competency in the subject. For example, question 61 has bad distractors: the right answer is 1,200,000 while the wrong answers are all powers of 10. Question 57 asks for the definition of Archimedes' principle; the answer appears verbatim in the handbook. The handbook gives incorrect formulas for hoop and axial stress in a thin-walled pressure vessel (it uses the mean radius rather than the inner radius of the pressure vessel); someone could get penalized for knowing the right formula. These factors suggest that the FE measures familiarity with the handbook and test-taking skills, which undermines the construct validity of the FE.

Each of these subjects is assessed at a simplistic level. The Computational Tools questions do not even assess how to translate a mathematical expression into a spreadsheet formula or a line of code. The fluid mechanics questions mostly just require application of simple formulas and definitions. Engineers use dimensionless parameters such as the Reynolds number to relate results from models to predicted behavior; instead, the questions involving the Reynolds number just call for calculating it. Although questions similar to 59 and 60 could assess competency with Bernoulli's equation or Reynolds Transport Theorem (respectively), these questions are posed such that they can be solved by rote application of formulas without consideration of fundamental concepts. The mechanical design and analysis questions can, likewise, mostly be answered by simple evaluation of formulas. Some questions are less complex than they appear. For example, question 100 would seem to assess limits and fits (the only aspect of manufacturability addressed in the handbook); however, only the relevant dimensions are given so familiarity with the relevant terminology is not actually assessed.

The FE does not assess any of these subjects at a relevant level of sophistication for the objectives held by my program. In our programming course, the main assessments call for students writing code to solve straightforward engineering problems on take-home assignments and to create a project that addresses an open-ended problem. In fluid mechanics, students synthesize multiple models of fluid motion, calculus, scaling analysis, and experimental methods. In mechanical design, students design gearboxes, accounting for fatigue failure of gears, bearings, and rotating shafts. Each of these tasks is more challenging, rewarding, and relevant than answering FE questions.

Now, one could argue that it is unfair to expect a standardized exam to assess these things. In response to that, I would first say that one could design a standardized exam that did assess abilities at a higher level. In the case of mechanical design, for example, an exam could at least assess each of the calculations necessary to design a simple machine such as a gearbox (the FE omits stress analysis of gears, stress concentrations on rotating shafts, and reliability of bearings) as well as some questions involving refining a design (e.g., based on which parts have lower factors of safety). This might not be an open-ended design experience, but it is more likely than the FE to measure ability to design a machine.

Second, being able to write code, combine models of fluids, or design a machine constitutes minimal competency in these areas—an assessment that does not reach these levels cannot credibly claim to show even minimal competency. Therefore, the FE is unlikely to have criterion validity, i.e., it is probably not predictive of preparedness for professional practice.

The FE is not useful for assessment for ABET

Aligning FE results with ABET student outcomes is challenging. Other authors have noted misalignment between the FE and ABET Criteria 2000 [22, 26]; here, I give some examples of this misalignment and discuss whether data from the FE is helpful for continuous improvement. The ABET student outcomes which were in place until quite recently [27] are shown in Table 7; although ABET no longer uses these student outcomes, they are used here for consistency with the existing literature. Now, NCEES states that the FE is appropriate for assessing ABET outcomes (a), (b), (c), (e), (f), and (k) [2]; these statements are made without evidence or argument so I will not address all of them in detail. I concede that the exam seems to cover some aspects of outcomes (a) and (f). However, it is odd to assert that a multiple-choice exam is a good assessment of the "ability to design and conduct experiments..." (outcome (b)). The FE questions I analyzed in the case studies call for analysis but not detailed design (in the sense of using analysis to select parameters or components); they certainly do not call for design "...within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability." (outcome (c)).

The FE is only partially aligned with outcome (e), and then, only if "engineering problems" means "textbook engineering problems." Real-world engineering problems are typically ill-posed, have multiple goals, have multiple solution strategies, and are subject to non-engineering constraints [23]; FE questions do not bear these features. Furthermore, FE questions are already identified and formulated so performance on the FE is not content valid with respect to those attributes of the outcome.

One program has used the percentage of students taking the FE to assess outcome (i) [5], with the argument that taking the exam shows a recognition of the need for life-long learning. However, most engineering jobs do not require licensure and that there are numerous other ways in which young engineers could pursue life-long learning so this claim is not well-supported.

There are further factors to consider when applying FE results to continuous improvement. Because FE data is reported at the level of subjects, this data is most useful for subjects that correspond to individual courses. For example, the Mechanical Design and Analysis subject includes both traditional mechanical design items as well as piping, pressure vessels, and actuators; in a given program, it may be that most machine components are addressed in a mechanical design course, actuators are addressed in a course on mechatronics, and failure of pipes and pressure vessels is briefly covered in solid mechanics. If students score high on Mechanical Design and Analysis, that could indicate that all of these items are being taught well. However, if FE results on this subject are low, it is unclear where improvements should be made.

Even if the FE were valid, there would be two additional factors that prevent FE subject matter reports from being useful for continuous improvement. First, subject matter reports are interpreted using norm-referenced rather than criterion-referenced data. This means that if a program has a low score in a subject, it is unclear if action is needed because it is impossible to determine from FE data whether student performance meets criteria that are relevant to a particular program.

Table 7: ABET a-k student outcomes

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

Second, all performance on a given subject is aggregated; therefore, it is impossible to determine from subject matter reports what topics within a subject should be emphasized.

Most engineering institutions have at least one program that uses the FE for assessment [2]. However, to my knowledge, there are only three published accounts of the FE improving a program. NCEES [2] describes a program in which low FE scores for Hydraulics and Hydrologic Systems led to improvements for the corresponding course. NCEES is probably referring to the reports of Nirmalakhandan et al. [4, 28] regarding their program. In this case, there were already numerous sources of evidence that the hydraulics course needed improvement, including student evaluations, exit interviews, and observations by the instructor of a subsequent course. The improvements to the program included adding an additional course on the topic, acquiring new lab equipment, changing the pedagogical approach, and modifying the course to increase coverage of FE content. FE scores indeed rose [4].

Bull et al. [5] describe how FE scores were used by the Chemical Engineering Program at the United States Military Academy (USMA) to argue to the school's administration that new courses should be offered that would address FE topics. Additionally, a course was introduced in which FE topics were specifically reviewed. Naturally, FE scores improved on the subjects that were addressed. Likewise, Guarino et al. [6] describe how adding courses on computing and ethics led to rises in FE scores. Neither of these case studies is strong evidence that the FE is useful in assessment. First, although the FE is probably not a valid measure of student ability in a subject, it is not so invalid that it is totally unaffected by students having had a course in a given subject [11]. Second, because the FE specifications are based on faculty opinions rather than evidence on what working engineers need to be able to do, simply ensuring coverage of the FE cannot be taken as evidence that students are better prepared for professional practice. Therefore, these modifications are simply better alignment with the exam specifications, rather than improvements in quality.

Other reports on use of FE data focus on interventions with the goal of increasing the FE pass rate itself. Effective interventions focused on motivation [18, 29, 30], advising [7], and review courses [29, 30].

Assessment should be able to detect problems, indicate a course of action, and confirm that the action is effective. In the case of the hydraulics course, there was ample evidence beyond the FE that the course needed improvement. FE data certainly could not have called for the numerous improvements that were undertaken. For the example of the chemical engineering program, low FE scores on specific topics were used to make a case to the administration for creation and revision of courses [5]. This argument relies on the assumption that if something is on the FE, it must be important. Because the process of developing the FE specifications is opaque and does not involve industry, this assumption is not justified. Furthermore, even NCEES states that programs should not simply teach to the FE but rather to use the FE to measure competency in areas that a program identifies as relevant [1]. Feedback from alumni and employers would be a better source of information regarding what that particular program's graduates should have mastered; information on what is taught at peer institutions would also be enlightening. The authors note that their institution, the United States Military Academy, takes external data particularly seriously so FE data is influential in decision-making [5]. The USMA has a unique identity and distinct purpose so the benefit of the FE yielding "external data" is probably outweighed by it being a generic measure that does not support the particular mission of the USMA.

In contrast, Mazurek describes a pragmatic approach to the use of FE data at The Coast Guard Academy [18]. The institution noted a decrease in FE pass rates and concluded that the cause was low student motivation. Furthermore, the Academy had required all students to take the FE as a graduation requirement; it was hypothesized that students who had no interest in passing the FE were bringing down scores. Dropping this requirement and motivating students by informing them about the importance of licensure led to an increase in pass rates; presumably, the students who chose to take the FE were motivated to pass. Ultimately, the institution recognized that the FE is not the best assessment of their engineering programs.

Given that the FE has widespread use in assessment, if it were appropriate for this purpose, there should be abundant evidence of its efficacy. In the few examples that purport to show evidence that the FE can improve programs, it is only in cases where the problems are already obvious (numerous pieces of evidence in one case, courses being absent from curriculum in the others) and it can only indicate interventions as blunt as introducing new courses for the sake of teaching to the FE. Although the FE is probably not valid, it is used to track performance year-over-year.

In summary, even if the FE were valid, the best it could do is suggest new courses that cover content related to one or two ABET student outcomes and track performance over time. Furthermore, the FE is most relevant to outcome (a) which is, anecdotally, the one that is already the most addressed in engineering programs and the least in need of an additional measure.

Conclusion and recommendations

The FE should not be used for design of curriculum. First, although passing the FE is an essential step for licensure, there is not a nationwide need to better prepare engineering students for the FE. The FE has a high pass rate (about 70%) [31], while fewer than half of recipients of engineering bachelor's degrees take the FE, let alone pursue licensure. If there were a need to increase the fraction of engineers with licenses, helping young engineers understand the benefits of licensure would be more effective than taking special measures to make curriculum fit the FE.

Second, the FE Exam Specifications are vague, ill-posed, and developed by an opaque process that does not involve industry. Therefore, it is hard to argue that the FE Exam Specifications are good guidance regarding what engineers should know. It is unclear why this problem has not been previously reported but one possibility is that others have analyzed the FE at too high of a level. For example, Sullivan et al. [9] describe aligning curriculum with the FE but this alignment was done at the level of FE subjects and course objectives, not FE topics and learning objectives.

Learning objectives based on the FE Exam Specifications, the FE Reference Handbook, and the FE Practice Exam could, in principle, be used as guidance for developing a curriculum that comprehensively covers the FE. However, the Specifications are vague, the Practice Exam only samples a fraction of FE content, and learning objectives based on the Handbook amount to being able to use handbook formulas; therefore, it is futile to specify the FE in terms of learning objectives.

Third, and most importantly, teaching to the FE would be a distraction from the directions in which engineering education is being improved. To identify ways to improve engineering curriculum, the ASME Vision 2030 Project surveyed managers of engineers, recent mechanical engineering graduates, and leaders in education; the resulting report calls for increasing emphasis on professional skills [24]. Vision 2030 does not mention the FE or even licensure. Other organizations such as the National Academy of Engineering [32] and The Carnegie Foundation for the Advancement of Teaching [33] also call for greater development of professional skills. The FE, which is focused on technical details, is thus at odds with the agreed-upon directions for improvement of engineering curriculum.

The FE should not be used for program assessment. First, the FE is not a reliable or valid psychometric instrument: the content it assesses does not match the specifications, results are influenced by factors other than engineering competency, and it is unlikely that the FE can predict success in a professional environment. Second, even if the FE were valid, it would only be a credible tool for assessing one or two ABET outcomes and it cannot assess them in depth or comprehensively.

If the FE is not a valid measure, why is it used so commonly in program assessment? Apparently, there is a common assumption that the FE is valid; perhaps its use in licensure makes it appear credible. One could attempt to "teach to" the FE by consulting the exam specifications but not the handbook or practice exam questions; without comparing the specifications, handbook, and exam questions, the poor validity of the FE would not be apparent. The FE is a convenient source of data for ABET which may also explain its common use.

Alternatives to the FE exist. Programs can and should gather assessment data directly from courses and especially the culminating design experience; these tools can indicate student performance on a range of ABET outcomes. If it were necessary for a program to develop a comprehensive assessment of student technical proficiency, an exit exam could be an appropriate choice; the only advantages that the FE has over an in-house exit exam are the effort involved in developing the exam and the FE being nationally-normed. However, exit exams are probably superior to the FE because their content is known to the program, it can be customized to focus on student performance relative to criteria that are relevant to that program's goals, and it can guide instruction at a more detailed level. For example, one exit exam was able to point out specific topics for improvement, such as centroids and polar coordinates [34].

Others have suggested that the FE is important because it is nationally-normed [4]. Although it may seem desirable to study how one program's students compare nationwide, this kind of evidence is not necessary for program improvement and may even be a distraction. Indeed, ABET does not require the use of nationally-normed data [35]; when the EC2000 criteria were introduced, "nationally-normed" data was listed as one possible source of "evidence that may be used". Because each program serves a distinct constituency, assessment tools should focus on whether a program's students meet ABET student outcomes as operationalized by criteria set by that program.

Concept inventories are another appealing alternative to the FE for program assessment; they have been described as having "the potential to be one of the best "ABET EC 2000" assessment instruments for showing continuous improvement" [36]. These standardized tests are designed to be valid and reliable research instruments. Although they typically use multiple-choice questions, the distractors for the questions are convincing to students who do not have a strong conceptual knowledge; thus, despite being multiple-choice questionnaires that take 1-2 minutes per question, concept inventories can assess student working knowledge. Some analyses of concept inventories have found possible issues with validity [37]; with that said, analysis of this depth is not even possible for the FE without cooperation from NCEES.

If a program were to have problems with graduates having unsatisfactorily low FE pass rates, there are numerous effective interventions that can be tried. Faculty can increase student motivation to pass the FE by giving them examples of how the FE helped engineers develop their careers. FE review courses can increase pass rates [19, 20, 30]. Familiarity with the FE Reference Handbook is a major factor that affects performance on the FE. One strategic test-taker developed what he called an "ABC approach" [38]; he divided the exam into subjects he could get A) 85%, B) 60%, and C) 25% correctly and made sure that he studied until he had enough "A" and "B" subjects to be likely to get a passing score on the FE. Targeted interventions focusing on motivation and FE test-taking skills should be attempted prior to revising curriculum to conform to what one may imagine to be on the FE.

References

[1] NCEES, "Using the Fundamentals of Engineering (FE) examination as an outcomes assessment tool," NCEES,

Tech. Rep., 2014.

[2] S. F. Barrett, J. W. Steadman, and D. L. Whitman, “Using the Fundamentals of Engineering (FE) examination as an outcomes assessment tool,” NCEES, Tech. Rep., Feb. 2016.

[3] NCEES, “Fundamentals of Engineering (FE) Mechanical CBT exam specifications,” NCEES, Tech. Rep., 2013.

[4] N. Nirmalakhandan, D. Daniel, and K. White, “Use of subject-specific FE exam results in outcomes assessment,” *Journal of Engineering Education*, vol. 93, no. 1, pp. 73–77, 2004.

[5] G. Bull, M. Armstrong, and A. Biaglow, “Using the Fundamentals of Engineering exam to assess student performance in a chemical engineering curriculum,” in *Proceedings of the Fall 2016 ASEE Mid-Atlantic Regional Conference*, 2016.

[6] J. C. Guarino, J. R. Ferguson, and V. K. C. Pakala, “Quantitative assessment of program outcomes using longitudinal data from the FE exam,” *Proceedings of the 2013 ASEE Annual Conference & Exposition*, vol. 23, p. 1, 2013.

[7] R. Helgeson and E. Wheeler, “Passing the Fundamentals of Engineering examination as a graduation requirement in a general engineering program: Lessons learned,” *Proceedings of the 2005 ASEE Annual Conference & Exposition*, 2006.

[8] O. Pierrakos and H. Watson, “A comprehensive ABET-focused assessment plan designed to involve all program faculty,” in *Frontiers in Education Conference, 2013 IEEE*, 2013, pp. 1716–1722.

[9] S. Sullivan, R. Williams, W. Howard, J. Yao, and P. Kauffmann, “Identifying the content of a general engineering program using benchmarking and the Fundamentals of Engineering examination,” *Proceedings of the 2007 ASEE Annual Conference & Exposition*, 2007.

[10] W. D. Lawson, “Reliability and validity of FE exam scores for assessment of individual competence, program accreditation, and college performance,” *Journal of Professional Issues in Engineering Education and Practice*, vol. 133, no. 4, pp. 320–326, 2007.

[11] R. Wicker, R. Quintana, and A. Tarquin, “Evaluation model using Fundamentals of Engineering examination,” *Journal of Professional Issues in Engineering Education and Practice*, vol. 125, no. 2, pp. 47–55, 1999.

[12] B. E. Barry and M. W. Ohland, “Engineering ethics curriculum incorporation methods and results from a nationally administered standardized examination: background, literature, & research methods,” in *Proceedings of the 2009 ASEE Annual Conference & Symposium*, 2009.

[13] NCEES, *FE Reference Handbook*. NCEES, 2013.

[14] M. R. Lindeburg, *FE Mechanical Review Manual*. Professional Publications Inc, 2014.

[15] J. L. Watson, “An analysis of the value of the FE examination for the assessment of student learning in engineering and science topics,” *Journal of Engineering Education*, vol. 87, no. 3, pp. 305–311, 1998.

[16] E. W. Wheeler, “Determination of a predictive model for the Fundamentals of Engineering examination,” Ph.D. dissertation, The University of Memphis, 2012.

[17] N. Gnanapragasam, K. Kuder, and J. Dragovich, “Comparison between grade earned in a course and performance in a simulated FE exam for mechanics related courses,” in *Proceedings of the 2008 ASEE Annual Conference & Symposium*, 2008.

[18] D. F. Mazurek, “Consideration of FE exam for program assessment,” *Journal of Professional Issues in Engineering Education and Practice*, vol. 121, no. 4, pp. 247–249, 1995.

[19] E. Koehn, “Fundamentals of Engineering exam: motivation/review enhances pass rate,” *Journal of Professional Issues in Engineering*, vol. 115, no. 3, pp. 289–296, 1989.

[20] E. Koehn and R. D. Malani, “Review for and assessment of the Fundamentals of Engineering exam,” in *Proceedings of the 2005 ASEE Annual Conference & Exposition*, 2005.

[21] A. Miller, ““FE exam” the first “reality show” encounter for engineering and engineering technology undergraduates,” in *Proceedings of the 2006 ASEE Annual Conference & Exposition*, 2006.

[22] V. P. Drnevich and R. K. Tener, “Opposition to mandatory use of PE/FE exams as assessment tools,” in *Proceedings of the 1998 ASEE Annual Conference and Exposition*, 1998.

[23] D. Jonassen, J. Strobel, and C. B. Lee, “Everyday problem solving in engineering: Lessons for engineering educators,” *Journal of Engineering Education*, vol. 95, no. 2, pp. 139–151, 2006.

[24] ASME, *Vision 2030 Creating the Future of Mechanical Engineering Education*. ASME, 2012.

[25] NCEES, “FE Mechanical Practice Exam,” NCEES, Tech. Rep., 2017.

[26] K. J. Fridley, W. Back, and D. Williamson, “The ASCE BOK, ABET accreditation criteria, and NCEES FE Exam—are they appropriately aligned?” in *Proceedings of the 2016 ASEE Annual Conference & Exposition*, 2016.

[27] ABET, “Criteria for accrediting engineering programs, 2017–2018,” ABET, Tech. Rep., 2017.

[28] N. Nirmalakhandan and K. White, “Course refinement through outcome assessment: A case study,” *Journal of Professional Issues in Engineering Education and Practice*, vol. 126, no. 1, pp. 27–31, 2000.

[29] R. R. Foltz, T. W. Mays, and K. C. Bower, “Adding student led components to FE review courses to improve FE examination scores at The Citadel: A case study,” *Proceedings of the 2005 ASEE Southeastern Section Conference*, 2005.

[30] J. Bowen, “Efforts to better understand the relationship between civil engineering student preparation and success on the Fundamentals of Engineering exam,” in *Proceedings of the 2010 ASEE 2010 Annual Conference & Exposition*. American Society for Engineering Education, 2010.

[31] NCEES, “2016 squared: A year in numbers,” NCEES, Tech. Rep., 2017.

[32] National Academy of Engineering, US, *The Engineer of 2020: Visions of Engineering in the New Century*. Washington, DC: National Academies Press, 2004.

[33] S. D. Sheppard, K. Macatangay, A. Colby, and W. M. Sullivan, *Educating Engineers: Designing for the Future of the Field*. Jossey-Bass, 2008.

[34] P. Yearling, T. Anderson, and R. J. Durkin, “IUPUI Mechanical Engineering Technology Senior Assessment,” *Proceedings of the 2017 ASEE Annual Conference & Exposition*, 2017.

[35] ABET, “Criteria for accrediting engineering programs, 2018–2019,” ABET, Tech. Rep., 2018.

[36] D. Evans, G. L. Gray, S. Krause, J. Martin, C. Midkiff, B. M. Notaros, M. Pavelich, D. Rancour, T. Reed-Rhoads, P. Steif *et al.*, “Progress on concept inventory assessment tools,” in *Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference*. IEEE, 2003, pp. T4G1–T4G8.

[37] N. Jorion, B. D. Gane, K. James, L. Schroeder, L. V. DiBello, and J. W. Pellegrino, “An analytic framework for evaluating the validity of concept inventory claims,” *Journal of Engineering Education*, vol. 104, no. 4, pp. 454–496, 2015.

[38] A. Fasano, “The ugly truth behind the Fundamentals of Engineering exam (FE exam) and how you can finally crush it in less time than you think,” The Engineering Career Coach, Tech. Rep., 2014. [Online]. Available: ⟨https://engineeringcareercoach.com/ugly-truth-behind-fundamentals-engineering-exam-fe-exam-can-finally-crush-less-time-think/⟩

Appendix A: Selected FE Mechanical Practice Exam questions

The following questions are taken from the FE Mechanical Practice Exam [25]. Correct answers are indicated in bold.

Computational Tools

12. The flowchart for a computer program contains the following segment:

```
    VAR = 0
┌─► IF VAR < 5 THEN VAR = VAR + 2
│   OTHERWISE EXIT LOOP
└── LOOP
```

What is the value of VAR at the conclusion of this routine?
A. 0
B. 2
C. 4
D. 6

Fluid Mechanics

57. Archimedes' principle states that:

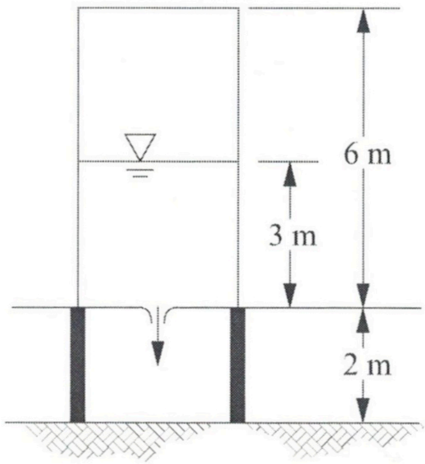
A. the sum of the pressure, velocity, and elevation heads is constant
B. flow passing two points in a stream is equal at each point
C. the buoyant force on a body is equal to the volume displaced by the body
D. a floating body displaces a weight of fluid equal to its own weight

[Note: the FE Reference Handbook states,

> **Archimedes Principle and Buoyancy**
> 1. The buoyant force exerted on a submerged or floating body is equal to the weight of the fluid displaced by the body.
> 2. **A floating body displaces a weight of fluid equal to its own weight**; i.e., a floating body is in equilibrium.

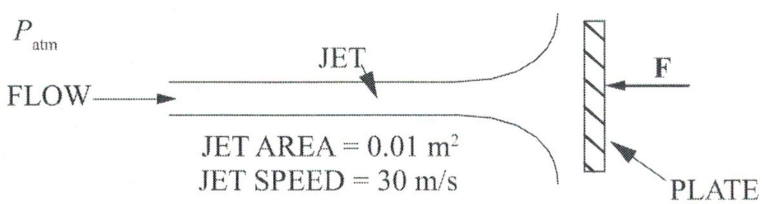
Emphasis added.]

59. Water is discharged to the atmosphere as a jet from a puncture in the bottom of a ventilated storage tank. The storage tank is a cylinder 6 m high mounted on a level platform 2 m off the ground. If losses are neglected, the jet velocity (m/s) when the tank is half full is most nearly:



A. 7.7
B. 9.9
C. 12.5
D. 50.8

60. A horizontal jet of water (density = 1,000 kg/m^3) is deflected perpendicularly to the original jet stream by a plate as shown below. The magnitude of force **F** (kN) required to hold the plate in place is most nearly:



A. 4.5
B. 9.0
C. 45.0
D. 90.0

61. The mass flow rate of sodium traveling through a pipe with an inside diameter of 0.1023 m is 22.7 kg/s. The mass density of the sodium is 823.3 $\mathrm{kg/m^3}$, and the dynamic viscosity is $2.32 \times 10^{-4}\ \mathrm{kg/(m \cdot s)}$. The Reynolds number for the sodium flow through the pipe is most nearly,

A. 10,000
B. 100,000
C. 1,000,000
D. 1,200,000

Mechanical Design and Analysis

100. A steel pulley with a minimum room-temperature bore diameter of 100.00 mm is to be shrunk onto a steel shaft with a maximum room-temperature diameter of 100.15 mm.

Assume the following:
Room temperature = 20°C
Coefficient of linear expansion of steel = $11 \times 10^{-6}/°C$
Required diametral clearance for assembly = 0.05 mm.

To shrink the pulley onto the room-temperature shaft with the desired diametral clearance, the pulley must be heated to a minimum temperature of most nearly,
A. 65°C
B. 136°C
C. 182°C
D. 202°C