Defining Assessment: Foundation Knowledge Toward Exploring Engineering Faculty's Assessment Mental Models

Kai Jun Chew
Embry-Riddle Aeronautical University
Daytona Beach, Florida
chewk1@erau.edu

Amanda Ross
Virginia Tech
Blacksburg, Virginia
amandaross@vt.edu

Andrew Katz
Virginia Tech
Blacksburg, Virginia
akatz4@vt.edu

Holly M. Matusovich
Virginia Tech
Blacksburg, Virginia
matushm@vt.edu

***Abstract*—This full research paper documents assessment definitions from engineering faculty members, mainly from Research 1 universities. Assessments are essential components of the engineering learning environment, and how engineering faculty make decisions about assessments in their classroom is a relatively understudied topic in engineering education research. Exploring how engineering faculty think and implement assessments through the mental model framework can help address this research gap. The research documented in this paper focuses on analyzing data from an informational questionnaire that is part of a larger study to understand how the participants define assessments through methods inspired by mixed method strategies. These strategies include descriptive statistics on demographic findings and Natural Language Processing (NLP) and coding on the open-ended response question asking the participants to define assessments, which yielded cluster themes that characterize the definitions. Findings show that while many participants defined assessments in relation to measuring student learning, other substantial aspects include benchmarking, assessing student ability and competence, and formal evaluation for quality. These findings serve as foundational knowledge toward deeper exploration and understanding of assessment mental models of engineering faculty that can begin to address the aforementioned research gap on faculty assessment decisions in classrooms.**

Keywords—engineering, assessment, definition, decision, faculty

I. Introduction

Assessments are an important component of the engineering education learning environment [1]–[3]. The importance includes understanding student growth and proficiency in learning [4], assisting students with learning [5], [6], and for accountability purposes [7]–[9]. These underlie the need to understand how engineering educators make decisions with assessments, especially on how they design and implement assessments in their courses. This topic is particularly understudied in engineering education, and addressing this research gap is especially important considering the need to improve engineering education for various economic and social reasons [10]–[13].

This research begins addressing this research gap by presenting findings from an informational questionnaire implemented in the first phase of a larger study that seeks to explore and map out engineering faculty mental models on assessments. Herein, we answer the research question: "*How do engineering faculty who responded to a survey about mental models related to assessment define assessment*?" These findings, though not generalizable at this stage, will provide foundational knowledge that paves the way toward a more robust understanding of faculty thoughts and views about assessments and how they make decisions about assessments in their courses.

II. Literature Review

A. The Need for Deeper Exploration into the "Why" behind Assessment Decisions

Assessments are key components in an engineering learning environment. Some argue that there are differences between the terms assessment and evaluation [14], [15], but in this paper, the terms were not predefined as the study focused on understanding how engineering faculty members defined assessments, and some may see assessment and evaluation as similar. Pellegrino and colleagues called for a rethink of education assessments considering the advancement of learning sciences, supporting the notion that assessments are highly intertwined with other elements of teaching, such as learning outcomes and pedagogical approaches [2], [3]. In addition, assessments are crucial in the learning process, helping with various processes. One is to track students' growth and proficiency with the knowledge being taught [4]. Another important element of assessment is to assist students with learning the materials in a course through various processes such as compelling and motivating them to study for an assessment [5], [6], [16], [17]. Assessments are also essential in the focus on accountability, such as program assessments with ABET [7], [8], [18]–[20]. All in all, engineering education can benefit from more scholarship on improving assessments, considering assessments can come with different forms of definitions as described (tracking student learning, helping with learning process, working as tools for accountability). This paper strives to contribute to this goal.

Current literature on assessments, however, focuses a lot more on the "how" without understanding the "why" behind assessment usage and decisions. For instance, there have been publications focusing on guidelines to design and implement assessments, such as different ways of creating and administering tests [21]–[25]. Other guidelines include how to assess for ABET [26], [27] and how to design and implement alternative assessments in engineering courses [28]–[32]. Many of these works provided brief explanations of the decisions behind the assessment implementation. However, explicit faculty personal assessment philosophies, including how different contexts like the course and departmental contexts, influence their philosophies, should be further explored to understand why the faculty decided on these assessments. This uncovers a research gap in engineering education, which is to understand the inherent "why" faculty decided on certain types of assessments. For instance, tests have been shown to be deeply embedded in engineering education assessment culture [33]–[35], and there has been a lack of research on why faculty ultimately decide to continue using tests heavily in spite of other forms of assessments documented, though sprinkles of mentions of large class size and inertia were mentioned without a deeper dive into these explanations [36]. Understanding the inherent "why" should provide the engineering education community with important information in advancing assessment research and practice, specifically on the need for more diverse sets of assessments in light of knowledge improvement in terms of learning sciences and pedagogical approaches.

B. *Faculty Perspective on Assessments is Crucial*

To understand deeper the "why" behind assessment decisions, focusing on the engineering faculty perspective is essential. Research has shown the importance and strength of leveraging the knowledge of faculty perspectives to make change in educational practice. For example, research on teacher's beliefs and how they influence practice, including assessment decisions [37]–[40] has supported that beliefs, to varying degrees, influence practice in the classrooms [41], [42]. For instance, there has been research showing teacher's self-efficacy beliefs that influence their practice [41]. In engineering education, [43] used a multi-case study design to show the change of four faculty instructional beliefs from different departments while implementing model-eliciting activities (MEA) in a variety of engineering courses. Moore and colleagues found that these faculty shifted their instructional beliefs toward a more student-centered approach with the adoption of MEA, consistent with belief research that also shows practice can shift teaching beliefs [41]. This demonstrates that engineering instructors' teaching beliefs can change, and understanding what their beliefs are currently on assessment is an important step toward pursuing change in assessment beliefs among engineering faculty.

III. Research Purpose and Question

The arguments to understand engineering faculty mental models of assessment revolve around the need to further explore the fundamental "why" behind assessment decisions from the faculty perspective. The larger, NSF-funded research strives to address the aforementioned research gap by exploring and mapping out engineering faculty mental models (another faculty-perspective research approach) on assessments. Herein, the focus is on providing a foundational piece of knowledge for the exploration of assessment mental model by answering: "*How do engineering faculty who responded to a survey about mental models related to assessment define assessment*?"

This study is funded by the NSF Grants 2113631 and 2107008.

IV. Guiding Conceptual Framework

Mental models, the guiding conceptual framework, are internal representations people possess to describe, explain, and predict the various components of a system, namely a system's state, form, function, and purpose [44] [45]. Mental models allow one to plan out their future actions [46], [47]. The mental model approach is commonly used in fields such as risk communication [48] and system dynamics [49]. Within the field of engineering education, there is a nascent tradition of using the approach to study students and faculty. For example, McMahon used the mental model framework as a way to understand two communities of teacher's mental models of the engineering design process, and provided curricular recommendations based on the findings [50]. That study showed the utility of the mental model framework in course decisions and practice. Likewise, building on that tradition in the present study, the mental model approach is a useful way to understand assessment decisions among engineering faculty, especially exploring how faculty can describe, explain, and predict assessment decisions. For this paper, however, the focus of the mental model is on one of the components: Purpose. In mental model research, Purpose is associated with why a system exists, and the assessment definitions given by the participants focus substantially on why assessments are used in their courses [45]. Relevant to the present work, definitions of assessments shape part of the assessment form in engineering education. Thus, these definitions will serve as the foundational knowledge toward understanding the larger form of assessments, and other faculty mental model components of assessments in engineering education.

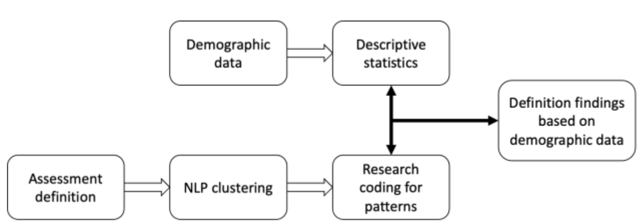
V. Method

This research leveraged mixed-method approaches to analyze an informational questionnaire implemented in the first phase of a larger research study. The informational questionnaire serves two purposes: 1) To recruit interview participants for the larger study and 2) to understand the lay of the land on how engineering faculty think about fundamental elements of assessments. The informational questionnaire contained demographic questions, the participant's personal perception of importance on the typical faculty evaluation domains (research, teaching, and service), participant's perception of how their department, college, and institution value the faculty evaluation domains, and assessment-related questions. These assessment-related questions were open-ended and included "How do you define assessment," "what types of courses do you typically teach," "what types of assessments do you use in your courses," and "to what degree do you feel responsible for making decisions about assessments in your courses?" This analysis specifically focuses on providing key demographic information (self-identified gender, and race/ethnicity, faculty position, tenure status, engineering disciplines, and years of teaching experience) on who has responded to the questionnaire and how these participants defined assessments. Essentially, these questions were defined and crafted specifically to understand who these participants were, where they worked and how they viewed assessments at a high level. As part of the research quality measures, the questionnaire was piloted with peers before the administration occurred.

In terms of recruitment, the research team enacted several strategies. One, the team reached out to peers in engineering with personal connections for participation and help with disseminating the questionnaire. Two, the team also compiled a list of email addresses from institution public websites. To guide the institution selection, the team used the ASEE Engineering by the Number [51] for guidance. For each engineering discipline, email addresses of faculty from the top five in terms of total bachelor's degrees awarded were collected from the public website. In addition, the team also compiled email addresses from faculty who work in the top ten institutions that awarded engineering degrees to underrepresented minorities by percentages to ensure a diverse set of participants were invited to provide perspectives on assessments. To date, about 3000 invitations were sent out, with 142 responded to the questionnaire. This research has obtained approval from the Institutional Review Board (IRB) of the home university. The participants were asked to provide consent to participate in the study before they entered the questionnaire.

The analysis process is illustrated in Figure 1. To analyze the demographic data, the authors employed descriptive statistics to tabulate and visualize the participants' information, which includes self-identified gender, race/ethnicity, faculty position, tenure status, engineering disciplines, and years of teaching experience [52]. For the open-ended responses on assessment definition, the authors leveraged the use of natural language processing (NLP) to provide the first cycle of analysis. This method was described in [53]. The basic approach is built upon sentence transformers [54], a relatively novel form of neural network architecture used in NLP applications for capturing semantic meaning in raw text. The method notably does not rely purely on simple word counts, which makes it more flexible for handling open-response items as in this study.

Fig. 1. Analysis strategies for the demographic data and assessment definition.



After using the sentence transformer to create a high dimensional numerical representation of each participant's written response, dimension reduction is used to create a lower dimensional representation. In this lower dimensional space, responses are clustered together. Theoretically, these clusters represent coherent themes. For example, responses about assessment being about comparing performance against objective or outcomes were all clustered together in this process, as described in the Results section.

The first cycle analysis yielded clusters based on the responses. The authors then coded the responses to label the clusters and identify potential patterns among the clusters and responses, leading to several categories of responses that emerged from the analysis. This process was inspired by thematic analysis [55].

After obtaining the descriptive statistics and the patterns from the assessment definitions, the researchers merged the data sets to seek out findings on the definition based on the demographic data, and this analysis stage was inspired by mixed method strategies of mixing or integrating quantitative and qualitative data at the analysis stage to seek out robust inferences from the data set [56], [57].

VI. Study Limitation

As with all research, this study has limitations. The majority of the respondents in the questionnaire worked in a Research 1 universities where research is prioritized for faculty [58], with several from teaching-focused institutions. In addition, the questionnaire was not designed as a cross-sectional instrument for inference generalization. This means the findings, though useful in beginning conversations about assessment definition in engineering education, cannot be immediately generalized to the general engineering faculty population. The low response rate of the questionnaire further the limitation and also shows the difficulty in conducting similar research in engineering education. Lastly, this paper focused on the specific question about how these participants defined assessments on the questionnaire, meaning many of the definitions provided were short and succinct. The definitions provided useful insights into understanding how engineering faculty defined assessments, but are limited to the words provided, sometimes without elaborate context. These are limitations that should be kept in mind while interpreting the findings.

VII. Results

A. *Overall demographic findings*

The data set analyzed for this paper contains 110 respondents. The following tables present the demographic findings from these respondents. The N=110 included participants that responded with an assessment definition. Those who left the question blank were not included in this analysis process (32 respondents out of the 142).

Tables I and II show that the data set has a majority of male and White respondents, though there are representations from female engineering faculty and respondents who identify with Asian, Black or African American, Hispanic, Latino, or Spanish Origin, and Middle Eastern or North African. Some identified with more than one race/ethnicity and chose not to identify.

TABLE I. SELF-IDENTIFIED GENDER OF RESPONDENTS

Gender	Frequency
Male	77
Female	28
Prefer not to say	5

TABLE II. SELF-IDENTIFIED RACE AND ETHNICITY

Race/Ethnicity	Frequency
White	74
Asian	13
Hispanic, Latino, or Spanish origin	6
Identified with more than one race/ethnicity	7
Black/African American	4
Middle Eastern or North African	2
Another Race/Ethnicity not Listed	2

In terms of academic positions, Tables III and IV show the makeup of the data set based on the faculty position and tenure status, showing that the majority of the respondents were tenured or on the tenure-track. There are more assistant professors, but the proportion of assistant, associate, and full professors is almost equal. Non-tenure track forms a small part of the data set, with positions like lecturer/instructor and professor of practice making up the non-tenure-track groups. A single research professor participated in the questionnaire.

TABLE III. FACULTY POSITIONS

Faculty position	Frequency
Assistant Professor	37
Full Professor	31
Associate Professor	23
Lecturer/Instructor	13
Administration	8
Professor of Practice	6
Distinguished Professor	3
Research Professor	1

TABLE IV. TENURE STATUS

Tenure status	Frequency
Tenured	53
Tenure Track	33
Non Tenure Track	22
Other	1

TABLE V. ENGINEERING DISCIPLINES

Discipline	Frequency
Electrical engineering	21
Mechanical/Manufacturing engineering	15
Aerospace/Ocean/Astro engineering	10
Chemical engineering	10
Civil engineering (non-structural)	9
Software Engineering/Computer Science	9
Bioengineering/Biomedical Engineering	7
Industrial/Systems Engineering	7
Computer Engineering	6
General Engineering	6
Environmental/Ecological Engineering	3
Structural/Architectural Engineering	3
Construction Engineering/Management	1

Engineering disciplines wise (Table V), there are more than 10 respondents for four engineering disciplines (electrical, mechanical or manufacturing, aerospace, ocean, or astro, and chemical). There are also representations from other disciplines like civil, software engineering or computer science, bioengineering or biomedical, industrial or system, computer engineering, and general engineering, and a small number who work in other engineering disciplines. Overall, there is a diverse representation in terms of engineering disciplines.

The data set has a large representation of those who have long years of teaching experience, with the plurality of the group having more than 15 years of teaching experience. The other groups have a fairly equal representation roughly in terms of proportion, with those with less than a year of experience being the least in the data.

TABLE VI. YEARS OF TEACHING EXPERIENCES

Years of Teaching Experience	Frequency
More than 15 years	48
11-15 years	12
7-10 years	15
4-6 years	16
1-3 years	16
Less than one year	3

B. Assessment Definition Findings

TABLE VII. CLUSTER THEMES

Cluster Theme	Frequency
Assess student learning with respect to learning outcomes	23
Benchmarking	20
Assess student learning	19
Assess for student ability and competence	17
“Formal evaluation,” evaluation, or evaluation for quality	16
External or program evaluation	8
Decision making	3
No attempt to define	4

On the assessment definition, the NLP and coding led to several clusters of the open-ended responses on how these respondents defined assessments, as tabulated in Table VII.

Table VII shows the cluster themes of the open-ended responses and their respective frequencies. As shown, most of the respondents defined assessments as ways to assess student learning with respect to learning outcomes, with example quotes presented below.

“*Determining learning by students with respect to learning objectives*.”

“*How well did students achieve the outcomes for the course as defined in the syllabus*.”

These quotes, and other similar definitions, all discussed measuring student learning and achievement of the learning outcomes listed in the course objectives. These are different than the other cluster theme of “Assess student learning” as definitions provided clustered under “assess student learning” did not explicitly mention learning outcomes as the level learning is assessed to, as exemplified by quotes below.

“*Seeking to quantify the amount of knowledge successfully transferred to students in the class*.”

“*Quantitative and qualitative methods for understanding a person's knowledge*.”

In some of the quotes, there were some definitions that specified the methods used, such as the example quote that mentioned quantitative and qualitative methods to measure a student’s knowledge. Overall, definitions categorized under “assess student learning” are broader and did not mention specific constraints like the learning outcomes.

There were about 20 respondents whose definitions fit the idea of “benchmarking”. In this case, these definitions specifically discuss the need to compare student performances to pre-existing standards or another group of their peers.

“*Evaluation of whether defined standards are met*.”

“*Determining in a quantitative way how someone is doing relative to their peers or to some standard*.”

This is different than the definitions categorized in “assess student learning with respect to learning outcomes” as many of these definitions focus a lot more on the language of “standard” or “baseline,” which imply a more general form of comparison as compared to learning outcomes in courses.

About 17 respondents defined assessments as ways to measure their students’ ability and competence, specifically mentioned words like “skill,” “ability,” and “experience” in these definitions.

“*Evaluating a student’s knowledge, skills, and experience*.”

“*How to judge understanding and ability to apply skills to engineering problems*.”

Although some of the definitions did mention learning outcomes, these definitions were more specific about the students’ skills, abilities, and competence when it comes to assessment.

As for the category of “Formal evaluation, evaluation, or evaluation for quality,” these definitions specifically focus on evaluations of students for various reasons, such as the need for “formal” forms of assessment, the need to evaluate to improve the quality of education, and general mention of evaluation.

“*Assessment is the formal evaluation of your work (progress or deliverables)*”

“*Measuring the state of something, measuring a level of quality.*”

“*Feedback, evaluation, information.*”

It must be noted that some definitions did mention feedback as part of the evaluation.

On other cluster themes, some mentioned external (tenure) or program (ABET) evaluation, though it must be noted that these definitions had no relation to student assessment.

“*P&T evaluation, ABET self-evaluation, survey feedback from alumni and employers*.”

Some other definitions mentioned the need for assessment for decision-making for evidence-based practices, for example.

“*Gathering for data-driven decision-making and evidence-based practices*.”

Lastly, some respondents did not attempt to define assessments. Some left short comments such as follows, providing no information on how they defined assessments.

“*In one line*?”

"*No idea*."

C. Assessment Definitions based on Demographics

TABLE VIII. "ASSESSING STUDENT LEARNING" AND "ASSESSING STUDENT LEARNING (LEARNING OUTCOMES)" ASSESSMENT DEFINITIONS BY FACULTY POSITIONS

Faculty position	Frequency
Assistant Professor	23
Associate Professor	8
Full Professor	6

Analyzing the cluster themes of assessment definitions with the demographic data, an interesting observation emerged. Many respondents who are assistant professors had definitions categorized as part of "assess student learning" and "assess student learning with respect to learning outcomes." As a comparison, associate, full professors, and lecturers/instructors had definitions that spread across the different cluster themes with no emerging pattern with these positions with respect to these two cluster themes. Specifically, 23 out of 37 of the assistant professors in the data set had definitions categorized in these two themes (Table VIII).

In addition, there is no emerging pattern with other cluster themes based on the positions aside from both themes on assessing student learning. This pattern is also consistent with the tenure status as 21 respondents who were on the tenure track had "assessing student learning" as their assessment definitions, while 15 tenured professors were in the same group.

It must be noted that there are no observable patterns when cross analyzing the assessment definition cluster based on other demographic findings

VIII. DISCUSSION AND IMPLICATIONS

Our results reveal several key findings. First, the assessment definitions provided by the respondents are largely aligned with existing literature. The idea of assessment being used for assessing student learning is typical in many assessment research literature, especially on assessing achievement based on certain outcomes [1], [2]. For Pellegrino and colleagues, this type of assessment is known as assessment of individual achievement, and a large number of definitions provided by the respondents revolve around this idea as many definitions focused on the need to assess student learning [2]. This aligns with Suskie's definition of assessments too, as Suskie argued that assessment is "deciding what we want our students to learn and making sure they learn it," implying the specific focus on learning outcomes [1, p. 8]. In engineering education, literature on assessment have also defined "assessment" around student learning.

Second, it is not surprising to observe that many of the engineering faculty in the data set defined assessments based on the idea of "baseline," "standards," or "external evaluation," and "evaluation for quality." This is consistent with existing literature as engineering education assessment has intimate relationships with accreditation with ABET for the need for quality control of engineering education [7]. There have been existing literature that provide guidance on how assessments can be designed and implemented toward ABET program evaluations, and having engineering faculty that define assessments based on ABET and accreditation goals is consistent with the overall accreditation culture of engineering education [18]–[20], [26]. It must also be noted that definitions that fell under "benchmarking" can also be traced to similar literature on accreditation for quality as many of the "benchmarking" definitions mentioned the need to assess against standards, meaning the language used to standardize engineering programs toward continuous improvement of program quality to produce future engineering students [7], [59].

Third, in terms of defining assessment for measuring student ability and competency, [60], when discussing the need for the different methodologies for assessment, some participants used the term "assessment" generally for measuring individual student's competencies like exam and homework assignment scores. This shows that definitions of measuring student ability and competency exist in engineering education literature.

Interestingly, the data lacked definitions that mentioned assisting students to learn in the course. Classroom assessments have been a substantial scholarship and practice domain in education, and classroom assessments are conceived as assessments that are specifically leveraged for learning among students [2], [6], [16], [61] instead of focusing on the need to measure students. Although there were mentions of feedback from assessments in some of the definitions, overall, there was a clear pattern that assisting students to learn with assessments is scarce among the definitions.

Another thing to note is the pattern when respondents who were assistant professors had defined assessments more around assessing student learning. Although there is no existing literature that engage with this topic as far as the authors knew, the authors surmised that this could be an important topic to focus on in the larger study in understanding the disparity observed here between assistant professors (more junior faculty) and associate and full professors (more senior faculty). The large study may explore this pattern in the next stage of the research.

These assessment definitions can serve as foundational knowledge toward understanding engineering faculty's mental models on assessment. As previously described, assessment definition is considered as part of the assessment mental model component of Purpose. With these findings, the engineering education community can begin researching how assessment mental models work, and the larger study this research stemmed from will leverage and expand upon these findings to explore and map out engineering faculty mental models, which can help understand how engineering assessment decisions are being made in classrooms. Ultimately, these findings can contribute to efforts in improving assessments in engineering classrooms, such as identifying engineering faculty who may be more open to changing assessment approaches in their classrooms and using innovative assessments. In addition, the engineering education community can leverage these findings to pursue further scholarship on assessments in engineering education to contribute to efforts in making changes on how engineering education conducts assessments.

References

[1] L. Suskie, *Assessing student learning: A common sense guide*. Jossey-Bass, 2018.

[2] J. W. Pellegrino, N. Chudowsky, and R. Glaser, *Knowing what students know: The science and design of educational assessment*. Washington, DC, 2001.

[3] G. P. Wiggins and J. McTighe, "What is backward design?," in *Understanding by design*, 2011, pp. 7–19.

[4] L. Lachlan-Haché and M. Castro, "Proficiency or growth? An Exploration of two approaches for writing student learning targets acknowledgments," 2015. [Online]. Available: https://www.air.org/sites/default/files/Exploration-of-Two-Approaches-Student-Learning-Targets-April-2015.pdf.

[5] G. Gibbs, "Using assessment strategically to change the way students learn," in *Assessment matters in higher education: Choosing and using diverse approaches*, S. Brown and A. Glasner, Eds. 1999, pp. 41–53.

[6] L. A. Shepard, "The role of assessment in a learning culture," *Educ. Res.*, vol. 29, no. 7, pp. 4–14, 2000.

[7] J. W. Prados, G. D. Peterson, and L. R. Lattuca, "Quality assurance of engineering education through accreditation: The impact of engineering criteria 2000 and its global influence," *J. Eng. Educ.*, vol. 94, no. 1, pp. 165–184, 2005, doi: 10.1002/j.2168-9830.2005.tb00836.x.

[8] J. T. Brown, "The seven silos of accountability in higher education: Systematizing multiple logics and fields," *Res. Pract. Assess.*, vol. 11, no. 2017, pp. 41–58, 2014.

[9] P. Nagy, "The Three Roles of Assessment : Gatekeeping, Accountability , and Instructional Diagnosis," *Can. J. Educ.*, vol. 25, no. 4, pp. 262–279, 2000.

[10] National Academy of Engineering, "The engineer of 2020: Visions of engineering in the new century," Washington, DC, 2004. doi: http://www.nap.edu/catalog/10999.html.

[11] L. Jamieson and J. Lohmann, *Creating a culture for scholarly and systematic innovation in engineering education*. Washington, DC: American Society of Engineering Education (ASEE), 2009.

[12] L. L. Long III, "Toward an antiracist engineering classroom for 2020 and beyond: A starter kit," *J. Eng. Educ.*, vol. 109, no. 4, pp. 636–639, 2020, doi: 10.1002/jee.20363.

[13] K. A. Douglas, A. Rynearson, Ş. Purzer, and J. Strobel, "Reliability, validity, and fairness: A content analysis of assessment development publications in major engineering education journals," *Int. J. Eng. Educ.*, vol. 32, no. 5, pp. 1960–1971, 2016.

[14] T. W. Banta and C. A. Palomba, *Assessment essentials: Planning, implementing, and improving assessment in higher education*, 2nd ed. San Francisco, CA: Jossey-Bass, 2015.

[15] D. M. Mertens and A. T. Wilson, *Program evaluation theory and practice: A comprehensive guide*, 2nd ed. New York, NY: The Guilford Press, 2019.

[16] J. H. McMillan, Ed., *Sage handbook of research on classroom assessment*. Thousand Oaks, California: Sage Publications, 2013.

[17] S. M. Brookhart and J. G. DeVoge, "Testing a theory about the role of classroom assessment in student motivation and achievement," *Appl. Meas. Educ.*, vol. 12, no. 4, pp. 409–425, 2000, [Online]. Available: www.tandfonline.com/doi/abs/10.1207/S15324818AME1204_5.

[18] L. J. Shuman, M. Besterfield-Sacre, and J. McGourty, "The ABET 'Professional Skills' - Can They Be Taught? Can They Be Assessed?," *J. Eng. Educ.*, vol. 94, no. 1, p. 41, 2005, [Online]. Available: http://proquest.umi.com/pqdlink?did=792141431&Fmt=7&clientId=57708&RQT=309&VName=PQD.

[19] N. Soundarajan, "Preparing for accreditation under EC 2000: An experience report," *J. Eng. Educ.*, vol. 91, no. 1, pp. 117–123, 2002, doi: 10.1002/j.2168-9830.2002.tb00680.x.

[20] J. R. Lohmann, "EC 2000: The Georgia Tech experience," *J. Eng. Educ.*, vol. 88, no. 3, pp. 305–310, 1999, doi: 10.1002/j.2168-9830.1999.tb00451.x.

[21] R. M. Felder, "Designing tests to maximize learning," *J. Prof. Issues Eng. Educ. Pract.*, vol. 128, no. 1, pp. 1–3, 2002, doi: 10.33250/14.04.264.

[22] D. Timmermann and C. H. Kautz, "Multiple choice questions that test conceptual understanding: A proposal for qualitative two-tier exam questions," 2015, doi: 10.18260/p.24516.

[23] D. M. Qualters, T. C. Sheahan, E. J. Mason, D. S. Navick, and M. Dixon, "Improving learning in first-year engineering courses through interdisciplinary collaborative assessment," *J. Eng. Educ.*, vol. 97, no. 1, pp. 37–45, 2008, doi: 10.1002/j.2168-9830.2008.tb00952.x.

[24] W. Fellin and G. Medicus, "Multiple Choice Tests: More than a Time Saver for Teachers," *Int. J. Eng. Pedagog.*, vol. 5, no. 3, p. 15, 2015, doi: 10.3991/ijep.v5i3.4376.

[25] D. Chadha, M. Maraj, and A. Kogelbauer, "Opening Up Assessment in the Age of COVID-19: Exploring the Utility of Online Open-Book Exams," *Adv. Eng. Educ.*, vol. 8, no. 4, pp. 1–5, 2020.

[26] R. M. Felder and R. Brent, "Designing and teaching courses to satisfy the ABET engineering criteria," *J. Eng. Educ.*, vol. 92, no. 1, pp. 7–25, 2003, doi: 10.1002/j.2168-9830.2003.tb00734.x.

[27] E. Koehn, "ABET program criteria: Review and assessment for a civil engineering program.," *J. Eng. Educ.*, vol. 90, no. 3, pp. 445–455, 2001, [Online]. Available: http://0-search.ebscohost.com.library.uark.edu/login.aspx?direct=true&db=a9h&AN=48056580&site=ehost-live&scope=site.

[28] P. S. Steif and J. A. Dantzler, "A statics concept inventory: Development and psychometric analysis," *J. Eng. Educ.*, vol. 94, no. 4, pp. 363–371, 2005, doi: 10.1002/j.2168-9830.2005.tb00864.x.

[29] R. A. Streveler, R. L. Miller, A. I. Santiago-roma, M. A. Nelson, M. R. Geist, and B. M. Olds, "Rigorous methodology for concept inventory development: Using the 'assessment triangle' to develop and test the Thermal and Transport Science Concept Inventory (TTCI)," *Int. J. Eng. Educ.*, vol. 27, no. 5, pp. 968–984, 2011.

[30] K. C. Midkiff, T. A. Litzinger, and D. L. Evans, "Development of engineering thermodynamics concept inventory instruments," in *Proceedings - Frontiers in Education Conference*, 2001, vol. 2, p. F2A/3, doi: 10.1109/FIE.2001.963691.

[31] J. M. Williams, "The engineering portfolio: Communication, reflection, and student learning outcomes assessment," *Int. J. Eng. Educ.*, vol. 18, no. 2, pp. 199–207, 2002.

[32] C. A. Cherrstrom, C. Raisor, and D. Fowler, "Student chemical engineering reflective ePortfolios," *Chem. Eng. Educ.*, vol. 49, no. 3, pp. 157–166, 2015.

[33] O. Rompelman, "Assessment of student learning: Evolution of objectives in engineering education and the consequences for assessment," *Eur. J. Eng. Educ.*, vol. 25, no. 4, pp. 339–350, 2000, doi: 10.1080/03043790050200386.

[34] S. Sheppard, K. Macatangay, A. Colby, and W. M. Sullivan, *Educating engineers: Designing for the future of the field*. San Francisco, CA: Jossey-Bass, 2009.

[35] S. M. Lord and J. C. Chen, "Curriculum design in the middle years," in *Cambridge handbook of engineering educatinon research*, A. Johri and B. M. Moskal, Eds. New York, NY: Cambridge University Press, 2014, pp. 181–199.

[36] J. Cuseo, "The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students," *J. Fac. Dev.*, vol. 17, pp. 5–21, 2007.

[37] J. Skott, "The promises, problems, and prospects of research on teachers' beliefs," in *International handbook of research on teachers' beliefs*, H. Fives and M. G. Gill, Eds. New York, NY: Routledge, 2015, pp. 13–30.

[38] N. Barnes, H. Fives, and C. M. Dacey, "Teachers' beliefs about assessment," in *International handbook of research on teachers' beliefs*, H. Fives and M. G. Gill, Eds. Routledge, 2015, pp. 284–300.

[39] M. F. Pajares, "Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct," *Rev. Educ. Res.*, vol. 62, no. 3, pp. 307–332, 1992, doi: 10.3102/00346543062003307.

[40] H. Fives and M. M. Buehl, "Spring cleaning for the 'messy' construct of teachers' beliefs: What are they? Which have been examined? What can they tell us?," in *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.*, vol. 2, 2011, pp. 471–499.

[41] M. M. Buehl and J. S. Beck, "The relationship between teacher's beliefs and teacher's practices," in *International handbook of research on teachers' beliefs*, H. Fives and M. G. Gill, Eds. Routledge, 2015, pp. 66–84.

[42] L. A. Bryan, "Nestedness of beliefs: Examining a prospective

elementary teacher's belief system about science teaching and learning," *J. Res. Sci. Teach.*, vol. 40, no. 9, pp. 835–868, 2003, doi: 10.1002/tea.10113.

[43] T. J. Moore *et al.*, "Changes in Faculty Members' Instructional Beliefs while Implementing Model-Eliciting Activities," *J. Eng. Educ.*, vol. 104, no. 3, pp. 279–302, 2015, doi: 10.1002/jee.20081.

[44] P. N. Johnson-Laird, "Mental models and human reasoning," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 43, pp. 18243–18250, 2010, doi: 10.1073/pnas.1012933107.

[45] W. B. Rouse and N. M. Morris, "On looking into the black box: Prospects and limits in the search for mental models," *Psychol. Bull.*, vol. 100, no. 3, pp. 349–363, 1985, [Online]. Available: http://scholar.google.com/scholar?q=related:QM4p5zGC8jMJ:scholar.google.com/&hl=en&num=30&as_sdt=0,5.

[46] K. Carley and M. Palmquist, "Extracting, representing, and analyzing mental models," *Soc. Forces*, vol. 70, no. 3, pp. 601–636, 2016.

[47] C. D. Wickens and A. Kramer, "Engineering psychology.," *Annu. Rev. Psychol.*, vol. 36, no. 1, pp. 307–348, 1985, doi: 10.1146/annurev.ps.14.020163.001441.

[48] M. G. Morgan, B. Fischhoff, A. Bostrom, and C. J. Atman, Eds., *Risk communication: A mental models approach*. New York, NY: Cambridge University Press, 2002.

[49] J. K. Doyle and D. N. Ford, "Mental models concepts for system dynamics research," *Syst. Dyn. Rev.*, vol. 14, no. 1, pp. 3–29, 1998, doi: 10.1002/(SICI)1099-1727(199821)14:1<3::AID-SDR140>3.0.CO;2-K.

[50] A. P. McMahon, "Mental models elementary teachers hold of engineering design processes: A comparison of two communities of practice," 2012, doi: 10.18260/1-2--21686.

[51] American Society for Engineering Education, "Engineering & engineering technology by the numbers 2019," Washington, DC, 2020. doi: 10.1016/b978-0-12-809372-6.00007-4.

[52] A. Field, J. Miles, and Z. Field, *Discovering statistics using R*. Thousand Oaks, California: Sage Publications, 2012.

[53] A. Katz, M. Norris, A. M. Alsharif, M. D. Klopfer, D. B. Knight, and J. R. Grohs, "Using Natural Language Processing to Facilitate Student Feedback Analysis," 2021.

[54] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," in *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 2019, pp. 3982–3992, doi: 10.18653/v1/d19-1410.

[55] V. Braun and V. Clarke, "Thematic analysis," in *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.*, H. Cooper, Ed. 2012, pp. 57–71.

[56] J. W. Creswell, "Chapter ten: Mixed-methods procedures," in *Research design: Qualitative, quantitative, and mixed methods approaches*, Thousand Oaks, California: Sage Publications, 2009, pp. 203–224.

[57] E. Creamer, "A primer about mixed methods for research in an educational context," *Int. J. Learn. Teaching, Educ. Res.*, vol. 15, no. 8, pp. 1–22, 2016.

[58] The Carnegie Classification of Institutions of Higher Education, "Carnegie classification: Listing." http://carnegieclassifications.iu.edu/ (accessed Oct. 19, 2018).

[59] ABET, "Accreditation policy and procedure manual: Effective for reviews during the 2022-2023 accreditation cycle." Baltimore, MD, 2022, doi: 10.1136/oem.2003.008847.

[60] B. M. Olds, B. M. Moskal, and R. L. Miller, "Assessment in engineering education: Evolution, approaches and future collaborations," *J. Eng. Educ.*, vol. 94, no. 1, pp. 13–25, 2005, doi: 10.1002/j.2168-9830.2005.tb00826.x.

[61] S. M. Brookhart, "Developing measurement theory for classroom assessment purposes and uses," *Educ. Meas. Issues Pract.*, vol. 22, no. 4, pp. 5–12, 2003.