Methods for Developing a Survey to Measure Engineering Judgement Across Undergraduate Coursework

Ms. Hayden Radel, University at Buffalo, The State University of New York

Hayden Radel is a junior Civil Engineering student at the University at Buffalo. Radel is currently working with Dr. Jessica Swenson, contributing to ongoing work on the development and measurement of engineering judgement through open-ended modeling problems.

Dr. Jessica Swenson, University at Buffalo, The State University of New York

Jessica Swenson is an Assistant Professor at the University at Buffalo. She was awarded her doctorate and masters from Tufts University in mechanical engineering and STEM education respectively, and completed postdoctoral work at the University of Michigan. Her research examines emotions within engineering problem solving and the student experience, engineering judgment, and elementary school teachers learning to teach engineering.

Methods for Developing a Survey to Measure Engineering Judgement Across Undergraduate Coursework

Abstract

This full methods paper describes the development and validation of a survey designed to measure students' engagement with engineering judgement while solving an open-ended modeling problem (OEMP). Engineering judgement is how engineers make decisions when faced with real-world, complex problems and is essential to professional practice, but is not captured through traditional textbook-based instruction. Prior work identified a taxonomy of engineering judgement types through retrospective student interviews. Building on this taxonomy, we developed a Likert-scale style survey to enable large-N, quantitative assessment. This survey has been administered at multiple universities across diverse undergraduate core technical courses.

This work provides instructors with a practical tool to understand and support students' development of engineering judgement, offers students a structured opportunity for self-reflection, and enables new large-scale research on how students apply theory in open-ended problem contexts. By providing an efficient, scalable assessment method, this survey broadens current approaches to analyzing engineering judgement and supports instructional design with open-ended problems.

Introduction

Engineers employ engineering judgement when approaching problems in all engineering disciplines. Undergraduate engineering students are expected to develop their ability to exercise engineering judgement across course work and extracurricular activities, which is confirmed by the inclusion of it in the ABET accreditation requirements [1]. One type of problem that was developed to encourage engineering judgement development in undergraduate students is ill-defined, open-ended problems [2], [3], [4], [5]. Ill-defined, open-ended problems allow course instructors to simulate real-world scenarios to allow their students to prompt students to make assumptions, create mathematical models, and navigate uncertainty; which all feed into their ability to exercise engineering judgement. These problems are often introduced to students in Senior Design Capstone Courses [2], but a growing number of researchers are interested in integrating the development of judgement across the curriculum [6], [7], [8].

Traditionally, engineering judgement has been understood to be a function of expertise, meaning an engineer's ability to exercise and apply it effectively depends largely on their experience in the field [9], [10], [11], [12]. However, recent work suggests that engineering judgement is not solely a product of experience; undergraduate students begin developing and exercising engineering judgement through coursework. Research on how engineering judgement is developed in students has used qualitative methods including faculty panels [8], case studies [13], and undergraduate and graduate focus groups [14]. With ongoing research on engineering judgement being done, a need for a quantitative tool to measure engineering judgement has arisen to help non-research instructors measure their students or students themselves to reflect on their judgement. As noted in a 2025 research brief, this must begin with a clear definition of what engineering judgement is [8]. The definition used for engineering judgement in this paper is making assumptions, assessing reasonableness, overriding a calculated answer, and using technology tools to create solutions to engineering problems [3].

This study is part of a multi-institution research project seeking to understand and scaffold the productive beginnings of engineering judgement in undergraduate students. Previous work developed the Engineering Modeling Judgment, [3], which used qualitative interviews with students to establish an engineering judgement taxonomy [3]. We are expanding on research questions developed for the NSF grant [15]. This paper continues work on project research question 1, "In what ways do undergraduate engineering students display the productive beginnings of engineering judgment?" [15].

This paper details the development of a survey to measure student's engagement with engineering judgement through open-ended modeling problems (OEMP). This is useful to assess and compare student experiences across different courses and institutions. The goal of this work is to provide instructors and researchers with a practical, theory-driven tool that can be used to assess and compare students' use of judgement. We used a top-down approach to survey question development and refinement [16], using the theorized model of Engineering Modeling Judgment [3]. The research questions we sought to answer are:

(RQ1) How do we design a method to collect large-N quantitative data on students' practice of engineering judgement?

(RQ2) How do we validate survey items to measure engineering judgement?

(RQ3) What are the implications of using retrospective student self-assessment to collect data?

Background

While designing a system, whether it be for spacecraft propulsion or a liquid-mass damper for a structure, engineers must apply engineering models to predict material behavior, loading conditions, and safety factors; for example Euler-Bernoulli Beam Theory for structural analysis, Coulomb Failure in soil mechanics, and the Navier-Stokes Equation for fluid mechanics. The engineering design process is often taught in introductory engineering courses as the systematic way engineers approach all design problems; we see engineering judgement applied in the decisions made *within* steps of the design process.

Engineering judgement does not have one consistent definition, but has been noted to include attitudes, behaviors, and capabilities [17]; application of knowledge to support problem-solving [8]; making assumptions, judging reasonableness, overriding numerical answers, and use of technology tools [3]. Engineering judgement has typically been researched in professional engineers [9], [10],[11],[12], as it is generally understood that the ability to apply it is based on experience and expertise. Gainsburg [9] analyzed how structural engineers collaborate to model structures where there are countless unknowns. The engineers created models of systems, forcing them to negotiate between the real system and their abstract understanding of how to solve the system [9]. This modeling is often expected of professional engineers and undergraduate engineering programs typically introduce it to students through close-ended problems in early undergraduate courses. However, in these courses, students are often provided with assumptions and simplified diagrams that remove the real-abstract negotiation that Gainsburg [9] noted. While these diagrams are helpful for teaching students how to model systems after these negotiations have occurred, they leave a gap in students' abilities to create and judge models of real systems. Previous research on engineering judgement has looked at how students navigate this gap [18] and how instructors can support students learning this [19].

One way that has been used to encourage students to engage with engineering judgement in coursework are open-ended modeling problems (OEMPs). OEMPs have previously been implemented at multiple universities with data collected from voluntary student participants through recorded group meeting conversations, retrospective interviews with researchers, collection of student work samples and interviews with course instructors [2], [3], [18], [19]. This has developed an understanding of how undergraduate students are engaging with engineering judgement through these problems. Engineering education researchers developed a taxonomy of engineering modeling judgement (EMJ) [3], shown in Table 1. The taxonomy looks at how students develop models and exercise engineering judgement through ill-defined, open-ended problems [3]. This taxonomy was built from retrospective interviews with students that were transcribed and qualitatively coded regarding their engagement with an open-ended ill-defined problem in a lower-level engineering course [3].

Table 1. EMJ Taxonomy [3] and survey question numbers based on subtype

Code	Emerging EMJ Type	Item Number (s)
EMJ 1: Making Assumptions		2_1
EMJ 1a	Making an assumption with no justification	
EMJ 1b	Assumption considers the user, client, or manufacturer	1_1, 1_2
EMJ 1c	Assumption makes the model more realistic, accurate, representative, or typical based on the student's research or experimentation for the class	1_3
EMJ 1d	Assumption makes the model more realistic, accurate, representative, or typical based on the student's personal lived experience outside the classroom	
EMJ 1e	Assumption makes the model solvable	
EMJ 1f	Assumption makes the model easier to solve and/or simplifies the model	1_4
EMJ 1g	Assumption does not affect the output of the model	
EMJ 1h	Assumption models what the student thinks is the worst-case scenario	1_6
EMJ 2: Assessing reasonableness		2_2
EMJ 2a	Assessing the reasonableness of assumptions that the student made	1_5, 1_7
EMJ 2b	Assessing the reasonableness of the output of the model	1_8
EMJ 2c	Assessing the reasonableness of the model as provided by the instructor in the problem	1_13
EMJ 3: Overriding a calculated answer		
EMJ 3a	Overriding a calculated answer with no justification	
EMJ 3b	Overriding a calculated answer considering the user, client, manufacturer, or safety	1_11,1_12, 1_13
EMJ 4: Using technology tools		2_6

EMJ 4a	Using a technology tool to help with analysis or computation	1_14
EMJ 4b	Assessing their use of the technology tool	1_15

The context of this particular study is to build a quantitative measurement of engineering judgement from the EMJ taxonomy described above. The previous work has focused on first and second-year engineering students in statics and dynamics courses, who are beginning their exposure with engineering problems. As the number of researchers and institutions interested in emerging engineering judgement in engineering students has grown, the need for a shared understanding and measure of its development has grown. Despite this growing interest, few tools have been developed for this task and current approaches tend to be time intensive. To address this gap, we designed a tool that aligns with the taxonomy that will be tested on students across all course levels in undergraduate engineering.

In the following sections of the paper, we describe a tool created to measure engineering judgement to collect a large-N data set across levels of undergraduate education.

Development of the Engineering Judgement Tool

The engineering judgement (EJ) survey was developed to address the growing need for a scalable method of assessing how undergraduate engineering students engage with Engineering Modeling Judgment (EMJ) during open-ended modeling problems (OEMP). This tool provides measurement of EJ behaviors without requiring interviews, allowing students to reflect on their own modeling behaviors while giving instructors and researchers a practical way to evaluate students' engagement with EJ. Its structure lowers participation barriers, capturing a broader range of student opinions by minimizing the time and logistical coordination required for interview-based EJ assessments used in previous studies [2], [3], [17], [18].

The survey was developed using a theory-driven, top-down approach based on student engagement with EMJ established in [3]. Following practices outlined in [16], this framework guided the construction of individual items, a strategy often used in engineering education research studies [18], [20], [21]. Because the goal of this instrument is to measure specific features of engineering judgement, Likert-style items were selected over a traditional Likert scale. Likert-style items measure attitudes of respondents [22], [23], over a yes or no which reduces the interpretation possible or open-response questions that tend to dissuade respondent participation. We chose a 7-item Likert-style scale to account for nuanced responses without sacrificing data quality [24], [25], [26], [27], [28], [29], [30]. To ensure clarity and validity, the survey underwent iterative pretesting [16], [31], during which a second response scale was added to improve response interpretation.

This pretesting was conducted in five phases and included engineering judgement researchers, current engineering students, researchers at a peer institution, and instructors who assign OEMPs in their courses. In the first phase, the primary writer drafted items corresponding to each of the 15 EMJ subtypes, using stems like "I made…", "I considered…", and "I changed…" to frame the behaviors discussed in each subtype. This structure was intended to help students interpret items consistently and increase item clarity, both of which are emphasized in survey design and validity literature [16], [31], [32]. Items were reviewed to eliminate double-barreled phrasing, remove redundancy, and identify potential sources of confusion [16], [32]. The question numbers corresponding to each EMJ subtype is noted in Table 1, with the first number indicating if it is from the first or second likert scale and the second number indicating which number on the scale it is.

Although not a subtype noted in the EMJ taxonomy, two additional items were added regarding if students evaluated and prioritized criteria and assumptions. These items were incorporated based on a panel of professional engineers who reviewed and discussed OEMPs and the EMJ taxonomy in their role on the advisory board of this project. The panel emphasized deciding the relative importance of constraints and assumptions in their engineering practice and expressed interest in understanding whether, and if so how, students employ this type of judgement.

The second phase of the survey was made based on feedback from EJ researchers, who evaluated how the items reflected the EMJ taxonomy and if students' understanding of the items would reflect the targeted subtype. The researchers noted that the use of “model” or “assumptions” may be unclear and interpreted differently among students. To address this, we considered how “modeling” behaviors were defined in the taxonomy and how items could be tailored to explicitly match this. Additionally, we developed an initial definition of assumption – assumption means decisions that you made – which was incorporated into subsequent review stages to improve clarity and consistency.

The third phase of our survey review involved establishing a panel of undergraduate engineering students based on [16], [33], [34], [35]. Our goal was to understand how students in our target population will interpret what we are asking them [16]. For this reason, our panel had six students, with four having completed an OEMP in a prior semester. Students first reviewed the survey and identified any questions that were confusing or unclear. Then, they discussed what they thought the survey was about. All students agreed that the survey focused on how students solved OEMPs, used assumptions throughout them, and how their understanding of course material evolved. When asked whether any prompts were too difficult to understand, none were identified as problematic. Next, the panel was asked what questions would only get similar responses. The panelists were then asked to provide recommendations regarding the overall structure of the survey. The primary concern raised was that the survey was too long, which they felt could negatively impact participation rates. While students agreed that the survey was easy to understand, they emphasized that its length was a significant drawback. Throughout the discussion, probing non-directional questions were used [16], [33], [34] to clarify responses and avoid moderator bias. The most consistent feedback was that, although easy to understand, the overall length of the survey posed a barrier to engagement.

We then sent the survey, with all considerations from phases one through three, to be reviewed by engineering judgement researchers at a peer institution for our fourth phase. [16] highlighted the importance of varying perspectives in conducting the survey review process, thus two researchers were asked to take a mock survey and identify strengths and weaknesses. Our fifth phase, the final external phase, included reaching out to instructors that administer OEMPs and asking them for feedback. Feedback from phases four and five were received at a similar time, and thus were considered together. Many small changes were recommended and included, but the two most significant received regarded the scale and the definition of assumption. The first

Likert-style scale originally included quantitative labels such as “I did this a few times” and “I did this often”, but they were altered to be based on section of the OEMP and have numerical values i.e. “I did this sometimes, 1-2 times per section” and “I did this many times, 5+ per section”. The definition of assumption received feedback from many reviewers, and the finalized definition used for the tool is “Engineering assumptions are decisions made to simplify a problem that fill in missing information. Example assumptions are negligible mass, concentrated force, negligible dimensions, steady state, rigid body, and frictionless.”

The final engineering judgement tool can be found in the appendix.

Results

To establish the validity of this survey as a tool for measuring engineering judgement, we administered it to a range of undergraduate students at multiple universities throughout the fall 2025 semester. We chose a subset for the validation process of students who were enrolled in an introductory statics course at a large university. Students were tasked with demonstrating their knowledge of statics through construction of a balsa wood bridge and creation of a technical report. The project required completion of at least one prototype of their design, a SolidWorks model, equations that represent their model, and to assess the cost of the project– material cost and cost for CO_2 emissions. Although OEMPs have not typically involved building physical models [2], [3], this project required students to use all aspects of engineering judgement, with a strong reliance on creating the model before construction and predicting their models behavior, which aligns with the intent of OEMP implementation.

We performed exploratory factor analysis (EFA) because of its ability to quantify trends from Likert-style survey responses [36], [37], [38]. EFA uses statistical modeling to identify trends in responses from individual participants and between participants [36], [37], [38]. The second Likert-style scale was designed to validate the first scale using factor analysis. Additionally, each of the subtypes of EMJ have multiple survey prompts and we expected that the different prompts regarding the same subtype would be correlated together.

This analysis was done with an original size of N=139; after cleaning due to missing responses and from incorrect responses to our attention-check question, we used N=117. Multiple guidelines have been developed including ratios of participants to measured variables (i.e. 5:1, 10:1)[38], [39], a minimum N between 100 to 200 participants [39], and higher communalities require lower sample sizes below 100 [38], [39]. There is no universally agreed-upon threshold for an adequate sample size in EFA [36], [38], [39]. The general form for EFA is shown below.

$$x = \Lambda f + \epsilon$$

EFA identifies patterns in how people respond to questions and groups them by underlying factors, *f*. The number of factors depends on the variance in responses received. Generally, it is better to use less factors than more factors to keep the model simple, but enough to capture meaningful patterns. Each question has a factor loading, Λ, that shows how strongly one factor relates to the responses and how much variability it can explain, with a value of 1 representing a maximum correlation and 0 no correlation. The variability that can not be explained by the factors is represented by ϵ.

EFA works when the data is normally distributed, which means that most of the data is near the middle and few data points are near the extremes. Skewness and kurtosis are used to test the distribution, with skewness less than 2 and kurtosis less than 7 indicating that it is reasonably normally distributed [36], [39]. Multivariate normality is assumed for EFA, which means that all variables tested together follow a normal pattern [36], [39]. This can be tested using the Mardia Test [36], which returns the probability that multivariate normality is reasonable, with a higher value indicating it is more likely to follow a multivariate distribution [40]. If a p-value below 0.05 is returned from the Mardia Test, then it can be corrected with Bartlet's Least Squares Method [37]. The number of factors can be determined using multiple methods, including parallel analysis, scree plot, Kaiser Criterion, and theory [3]. Parallel analysis is done by comparing how the eigenvalues from the real data compare to eigenvalues of simulated data [37]. Scree plots require plotting how much of the variance in the variables could be explained by X number of variables and choosing the number that is one less than where the plot flattens out [36], [37], [38]. The Kaiser Criterion is conducted by computing the eigenvalues of the data and the number of eigenvalues greater than 1 is the number of factors [38], [41], [42]. Each one of these can be used in different scenarios, and each one was evaluated here and the results were compared. Lastly, we determined Cronbach's alpha to understand the internal consistency of the responses received [36], [37], [43]. These were calculated in MATLAB R2025b using built in statistics functions.

Both parallel analysis [44] and scree plot gave 2 factors as appropriate, but the Kaiser Criterion gave 6 factors as appropriate. Multiple studies [42], [45], [46] have noted the tendency of the Kaiser Criterion to overestimate the number of common factors, therefore for this study, 2 factors were selected. The results from skewness, kurtosis, Mardia Test, and Cronbach's alpha are shown in Table 2.

Table 2. Results from tests of data.

Measure	Minimum	Maximum	Criteria
Skewness	-1.38	-0.4	All should be less than 2.
Kurtosis	2.51	5.69	All should be less than 7.
Mardia Test	1.75E-10	0	Values below 0.05 indicate that multivariate distribution is not reasonable.
Cronbach's Alpha	0.83		Scores greater than 0.8 indicate high internal consistency.

All items were initially included in a preliminary EFA. A minimum factor loading of 0.32 and communality threshold of 0.30 were then used to refine the set of items retained for further analysis. The factor loading cutoff of 0.32 was selected because it represents at least 10% shared variance between an item and its factor, a standard applied in similar studies [37]. Within the same item, the smaller factor loading was kept if it was less than half of the larger factor loading to reduce the likelihood of crossloading (where items load onto multiple factors when they may only represent one underlying construct) [36], [37]. Factor loadings are summarized in Table 3, with the question number that can be correlated to Table 1 and the full table in the appendix.

Communality was calculated to understand how much of the question's variability can be explained using the factors. Item communality in social sciences is often considered to be good when between 0.4 and 0.7 [36], [37], and in this study, factor loadings were only retained if they had a communality of at least 0.3 to account for the smaller sample size. EFA was run multiple times until running EFA again returned the same set as was input.

Table 3. Results from factor analysis and communalities.

		Factor		
Number	Question	1	2	Communality
1_5	I made a new assumption or adjusted an assumption after attempting to solve the problem.	0.05	0.57	0.33
1_7	I considered if my assumption was reasonable.	0.11	0.60	0.37
1_8	Based on my answers from previous sections, I considered if my answer made sense.	0.14	0.66	0.46
1_9	I felt that some criteria from the assignment were more important than others.	0.15	0.73	0.55
1_10	I felt that some values (numbers, parameters, dimensions, etc.) were more important than others to designing the product/device.	0.05	0.69	0.48
2_1	I am better at making useful assumptions than I was before the project.	0.62	0.15	0.40
2_2	I am better at judging if my final answer makes sense based on the criteria of the problem.	0.66	0.17	0.47
2_3	I am better at solving for unknowns based on diagrams I drew.	0.53	0.16	0.31
2_4	I am better at solving for unknowns based on diagrams provided.	0.61	0.22	0.42
2_5	I feel more confident in my ability to judge if my diagram is accurate to real-life.	0.74	0.16	0.58
2_6	I am better prepared to solve real problems using math tools.	0.60	0.01	0.36
2_7	I am better at combining concepts from multiple courses to solve a single problem.	0.65	0.14	0.45
2_8	I improved my ability to decide what given information is most relevant.	0.70	0.03	0.49
2_9	Overall, I am better equipped to solve open-ended problems where the steps are not given.	0.76	0.05	0.57

Table 3 can be interpreted by examining which items load strongly on the same factor. Items 2_7, 2_8, and 2_9 have high loadings on Factor 1 (0.65, 0.70, 0.76) and negligible loadings on Factor 2 (0.14, 0.03, 0.05), indicating that they measure the same latent construct and are distinct from items 1_9 and 1_10, which load primarily on Factor 2 (0.73 and 0.69) and little on Factor 1 (0.15, 0.05). This pattern shows that responses to 2_7, 2_8, and 2_9 covary–meaning their responses change similarly– with each other and not with 1_9 and 1_10.

Discussion

This work built on the Emerging Engineering Modeling Judgment Taxonomy (EEMJT) developed in [3], and other work completed in [45], [46] which worked with undergraduate students in introductory statics and dynamics courses to develop the taxonomy. This survey was designed for and has been administered to undergraduate students' across all four years of curriculum. A key goal of this work is to understand how students at varying stages in their undergraduate curriculum engage with EMJ, so the questions were made to be applicable for all levels to account for this.

Limitations of this study include that the survey relies on retrospective self-reporting from students and therefore depends on their memories and understanding of the question. Previous work in EJ has also relied on retrospective student self-reports and not analysis of solely their work [3], [17], [18], [48], [49], [50]. In addition, [51] found that students tended to more accurately assess their growth over a semester through a retrospective survey than a pre-post format. Although it is possible that some students may under report their engagement in EJ if asked directly about it, the high frequency of responses indicating engagement suggests that this limitation did not meaningfully affect the results.

An important consideration for the implementation of this survey is if it is appropriate for the students it is being administered to. For example, it likely would not be appropriate for students working on close-ended problems exclusively, as this was developed for open-ended style problems where they are forced to make judgements to solve problems. Additionally, as introduced in the development section, depending on the OEMP students were assigned some questions may need to be reworded or excluded. Not all OEMPs include the use of mathematical tools like MATLAB, especially if implemented in a first-year course, but the other questions would likely still be appropriate. OEMPs can vary significantly in length depending on course structure and curriculum, and the first scale may need to be adjusted if that is the case.

Conclusions

This tool provides insights into engineering judgement for instructors, students, and researchers that can be demonstrated with future implementation of this tool. Building this survey from an existing behavior model, the EEJMT, gave us a clear structure for our tool. The goal of this work is to create a practical, theory-driven tool that's results can be used to understand EMJ across institutions and course levels.

Through an iterative, multi-phase development and validation process, we produced a survey that captures students' engagement with key elements of engineering judgement while reducing the work load for professors and researchers. This is important for engineering programs interested in measuring engineering judgement to meet their ABET-accredition requirements. Additionally, if a program implemented this across their courses, they could use it to support their ABET-accreditation or to guide continuous improvement of courses based on teaching advances made by other instructors using it. With this tool, work could be done to understand how students taking a statics course without a project designed to teach engineering judgement versus with a project, like an OEMP, to compare how it impacts their engagement with engineering judgement.

For instructors of engineering science courses, this tool provides a way to measure student learning outcomes beyond course tests and assignments. This is especially valuable because few existing tools allow instructors to assess how students engage in engineering judgement. By incorporating this survey, instructors can gain understanding of how their students are approaching questions and how they are making decisions. This information can be used to refine activities, embed judgement-based activities, or compare how different activities impact their learning. Instructors in other engineering courses can also benefit from these results, because students often use similar problem-solving strategies across their coursework. The findings can help instructors anticipate how students are likely approaching tasks in their own classes.

With this reflection tool, students can identify what parts of EJ they are effectively engaging in and where they want to improve. Students that are looking to prepare themselves for internships and early-career positions, can use the results to better articulate their engagement in EJ, communicate professional readiness, and seek out targeted learning experiences.

The implications of this research are expanding our understanding of students' engagement with EJ and how EMJ develops into engineering judgement used by professional engineers. This tool provides researchers with a path towards systematically studying how engineering judgement develops over time and across varying institutions. Use in longitudinal studies could help identify what instructor behaviors, class formats, and pedagogical techniques are particularly effective at developing EJ. This contributes to a broader understanding that researchers have been seeking to gain regarding how EJ develops and how it fits into engineering programs.

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant No. 2313240. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the national science foundation.

Special thanks to the professors and graduate students that assigned open-ended problems and facilitated student participation in the survey. We also thank the students who took time to participate in the survey and the survey focus group and carefully considered and answered our questions.

References

[1] ABET, “Criteria for accrediting engineering programs, 2025–2026.” 2024. [Online]. Available: https://www.abet.org/wp-content/uploads/2024/11/2025-2026_EAC_ Criteria.pdf

[2] A. Johnson and J. Swenson, “Open-Ended Modeling Problems in a Sophomore-Level Aerospace Mechanics of Materials Courses,” in *2019 ASEE Annual Conference & Exposition Proceedings*, Tampa, Florida: ASEE Conferences, Jun. 2019, p. 33146. doi: 10.18260/1-2--33146.

[3] J. Swenson *et al.*, “A taxonomy of emerging engineering modeling judgment in undergraduate engineering courses,” *J. Eng. Educ.*, vol. 114, no. 4, p. e70011, Oct. 2025, doi: 10.1002/jee.70011.

[4] R. A. Francis, M. C. Paretti, and R. Riedner, “Theorizing Engineering Judgment at the Intersection of Decision-Making and Identity,” *Stud. Eng. Educ.*, vol. 3, no. 1, p. 79, Nov. 2022, doi: 10.21061/see.90.

[5] J. Weedon, “Judging for Themselves: How Students Practice Engineering Judgment,” in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana: ASEE Conferences, Jun. 2016, p. 25509. doi: 10.18260/p.25509.

[6] M. Gerhardt, M. Robinson, and B. Faulkner, “Beyond Calculations: Engineering Judgment as Epistemic Cognition in Engineering Education,” in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 55504. doi: 10.18260/1-2--55504.

[7] R. Francis, M. Paretti, and R. Riedner, “Engineering Judgment and Decision Making in Undergraduate Student Writing,” in *2021 ASEE Virtual Annual Conference Content Access Proceedings*, Virtual Conference: ASEE Conferences, Jul. 2021, p. 37066. doi: 10.18260/1-2--37066.

[8] R. Clark, S. Dickerson, and A. Toader, “Defining Engineering Judgment,” in *2025 ASEE Annual Conference & Exposition Proceedings*, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 56213. doi: 10.18260/1-2--56213.

[9] J. Gainsburg, “Abstraction and Concreteness in the Everyday Mathematics of Structural Engineers,” 2003.

[10] José António Mateus De Brito, “Judgement in geotechnical engineering practice,” *Soils Rocks*, vol. 44, no. 2, pp. 1–26, Jun. 2021, doi: 10.28927/SR.2021.063821.

[11] V. Patel and G. Groen, “The general and specific nature of medical expertise: A critical look,” in *Toward a General Theory of Expertise: Prospects and Limits*, A. Ericsson and J. Smith, Eds., Cambridge, UK: Cambridge University Press, 1991, pp. 93–125.

[12] N. M. Bradford and R. Sen, “Engineering Judgment and the Design That Got Away,” *J. Prof. Issues Eng. Educ. Pract.*, vol. 130, no. 2, pp. 84–89, Apr. 2004, doi: 10.1061/(ASCE)1052-3928(2004)130:2(84).

[13] A. Goncher and A. Johri, “Contextual Constraining of Student Design Practices,” *J. Eng. Educ.*, vol. 104, no. 3, pp. 252–278, Jul. 2015, doi: 10.1002/jee.20079.

[14] J. Walther, N. Kellam, N. Sochacka, and D. Radcliffe, “Engineering Competence? An Interpretive Investigation of Engineering Students’ Professional Formation,” *J. Eng. Educ.*, vol. 100, no. 4, pp. 703–740, Oct. 2011, doi: 10.1002/j.2168-9830.2011.tb00033.x.

[15] L. Maykish, O. Johnson, K. Churakos, J. Swenson, and A. Johnson, “BOARD # 427: Preliminary Results of Understanding and Scaffolding the Productive Beginnings of Engineering Judgment in Undergraduate Students (RFE),” in *2025 ASEE Annual Conference*

& Exposition Proceedings, Montreal, Quebec, Canada: ASEE Conferences, Jun. 2025, p. 55805. doi: 10.18260/1-2--55805.
[16] W. Cobern and B. Adams, “Establishing survey validity: A practical guide,” *Int. J. Assess. Tools Educ.*, vol. 7, no. 3, pp. 404–419, Sep. 2020, doi: 10.21449/ijate.781366.
[17] D. Chadha and K. Hellgardt, “A case of conceptualisation: using a grounded theory approach to further explore how professionals define engineering judgement for use in engineering education,” *Eur. J. Eng. Educ.*, vol. 49, no. 2, pp. 348–369, Mar. 2024, doi: 10.1080/03043797.2023.2210517.
[18] J. Swenson, E. Treadway, and K. Beranger, “Engineering students’ epistemic affect and meta‐affect in solving ill‐defined problems,” *J. Eng. Educ.*, vol. 113, no. 2, pp. 280–307, Apr. 2024, doi: 10.1002/jee.20579.
[19] K. Churakos, E. Markham, J. Swenson, and A. Nightingale, “Understanding the Effect of Scaffolding on Introductory Ill-Defined Problems in Engineering Education,” in *2024 IEEE Frontiers in Education Conference (FIE)*, Washington, DC, USA: IEEE, Oct. 2024, pp. 1–9. doi: 10.1109/FIE61694.2024.10893350.
[20] Ö. Korkmaz, M. Kösterelioğlu, and M. Kara, “A Validity and Reliability Study of the Engineering and Engineering Education Attitude Scale (EEAS),” *Int. J. Eng. Pedagogy IJEP*, vol. 8, no. 5, p. 44, Oct. 2018, doi: 10.3991/ijep.v8i5.8667.
[21] I. Benek and B. Akcay, “Development of STEM Attitude Scale for Secondary School Students: Validity and Reliability Study,” *Int. J. Educ. Math. Sci. Technol.*, pp. 32–52, Jan. 2019, doi: 10.18404/ijemst.509258.
[22] A. Joshi, S. Kale, S. Chandel, and D. Pal, “Likert Scale: Explored and Explained,” *Br. J. Appl. Sci. Technol.*, vol. 7, no. 4, pp. 396–403, Jan. 2015, doi: 10.9734/BJAST/2015/14975.
[23] American Educational Research Association, Ed., *Report and recommendations for the reauthorization of the institute of education sciences*. Washington, D.C: American Educational Research Association, 2011.
[24] K. Finstad, “Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales,” vol. 5, no. 3, 2010.
[25] E. C. Aybek and C. Toraman, “How many response categories are sufficient for Likert type scales? An empirical study based on the Item Response Theory,” *Int. J. Assess. Tools Educ.*, vol. 9, no. 2, pp. 534–547, Jun. 2022, doi: 10.21449/ijate.1132931.
[26] E. P. Cox, “The Optimal Number of Response Alternatives for a Scale: A Review,” *J. Mark. Res.*, vol. 17, no. 4, pp. 407–422, 1980, doi: 10.2307/3150495.
[27] M. A. Revilla, W. E. Saris, and J. A. Krosnick, “Choosing the Number of Categories in Agree–Disagree Scales,” *Sociol. Methods Res.*, vol. 43, no. 1, pp. 73–97, Feb. 2014, doi: 10.1177/0049124113509605.
[28] G. M. Russo, P. A. Tomei, B. Serra, and S. Mello, “Differences in the Use of 5- or 7-point Likert Scale: An Application in Food Safety Culture,” *Organ. Cult.*, vol. 21, no. 2, pp. 1–17, 2021, doi: 10.18848/2327-8013/CGP/v21i02/1-17.
[29] J. Dawes, “Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales,” *Int. J. Mark. Res.*, vol. 50, no. 1, pp. 61–104, Jan. 2008, doi: 10.1177/147078530805000106.
[30] I. Kusmaryono, D. Wijayanti, and H. R. Maharani, “Number of Response Options, Reliability, Validity, and Potential Bias in the Use of the Likert Scale Education and Social Science Research: A Literature Review,” *Int. J. Educ. Methodol.*, vol. 8, no. 4, pp. 625–637, Nov. 2022, doi: 10.12973/ijem.8.4.625.

[31] K. E. A. Burns and M. E. Kho, “How to assess a survey report: a guide for readers and peer reviewers,” *CMAJ Can. Med. Assoc. J.*, vol. 187, no. 6, pp. E198–E205, Apr. 2015, doi: 10.1503/cmaj.140545.
[32] B. M. McSkimming, S. Mackay, and A. Decker, “Investigating the usage of Likert-style items within Computer Science Education Research Instruments,” in *2021 IEEE Frontiers in Education Conference (FIE)*, Lincoln, NE, USA: IEEE, Oct. 2021, pp. 1–8. doi: 10.1109/FIE49875.2021.9637198.
[33] T. Yan, F. Kreuter, and R. Tourangeau, “Evaluating Survey Questions: A Comparison of Methods,” *J. Off. Stat.*, vol. 28, pp. 503–529, 2012.
[34] R. Levenstein, F. Conrad, J. Blair, R. Tourangeau, and A. Maitland, “The effect of probe type on cognitive interview results: A signal detection analysis,” 2007.
[35] E. Harrison, M. Mizota, H. Daly, and E. Falkenburger, “Community-Engaged Surveys: From Research Design to Analysis and Dissemination,” 2024.
[36] A. Mazzurco, B. K. Jesiek, and A. Godwin, “Development of Global Engineering Competency Scale: Exploratory and Confirmatory Factor Analysis,” *J. Civ. Eng. Educ.*, vol. 146, no. 2, p. 04019003, Apr. 2020, doi: 10.1061/(ASCE)EI.2643-9115.0000006.
[37] A. Godwin, “The Development of a Measure of Engineering Identity,” in *2016 ASEE Annual Conference & Exposition Proceedings*, New Orleans, Louisiana: ASEE Conferences, Jun. 2016, p. 26122. doi: 10.18260/p.26122.
[38] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, “Evaluating the Use of Exploratory Factor Analysis in Psychological Research”.
[39] P. J. Curran and S. G. West, “The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis”.
[40] A. Trujillo-Ortiz and R. Hernandez-Walls, *Mskekur: Mardia’s multivariate skewness and kurtosis coefficients and its hypotheses testing.* (2003). MATLAB. [Online]. Available: https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Rmardia
[41] H. F. Kaiser, “The application of electronic computers to factor analysis,” *Educ. Psychol. Meas.*, vol. 20, pp. 141–151, 1960, doi: 10.1177/001316446002000116.
[42] J. Braeken and M. A. L. M. Van Assen, “An empirical Kaiser criterion.,” *Psychol. Methods*, vol. 22, no. 3, pp. 450–466, Sep. 2017, doi: 10.1037/met0000074.
[43] M. Wilson-Fetrow, V. Svihla, and A. Olewnik, “Confirmatory factor analysis of the framing agency survey,” in *2023 ASEE Annual Conference & Exposition Proceedings*, Baltimore , Maryland: ASEE Conferences, Jun. 2023, p. 42737. doi: 10.18260/1-2--42737.
[44] L. E. Garrido, “Horn’s parallel analysis method with polychoric correlations.” Accessed: Jan. 13, 2026. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/64213-horn-s-parallel-analysis-method-with-polychoric-correlations
[45] C. E. Lance, M. M. Butts, and L. C. Michels, “The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?,” *Organ. Res. Methods*, vol. 9, no. 2, pp. 202–220, Apr. 2006, doi: 10.1177/1094428105284919.
[46] M. Auerswald and M. Moshagen, “How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions.,” *Psychol. Methods*, vol. 24, no. 4, pp. 468–491, Aug. 2019, doi: 10.1037/met0000200.
[47] J. Swenson, A. Johnson, M. Rola, and H. Koushyar, “Making Assumptions and Making Models on Open-ended Homework Problems,” in *2020 ASEE Virtual Annual Conference*

Content Access Proceedings, Virtual On line: ASEE Conferences, Jun. 2020, p. 34940. doi: 10.18260/1-2--34940.
[48] J. Swenson, A. Johnson, T. Chambers, and L. Hirshfield, “Exhibiting Productive Beginnings of Engineering Judgment during Open-Ended Modeling Problems in an Introductory Mechanics of Materials Course,” in *2019 ASEE Annual Conference & Exposition Proceedings*, Tampa, Florida: ASEE Conferences, Jun. 2019, p. 32786. doi: 10.18260/1-2--32786.
[49] J. Luo and C. K. Y. Chan, “Exploring the process of evaluative judgement: the case of engineering students judging intercultural competence,” *Assess. Eval. High. Educ.*, vol. 48, no. 7, pp. 951–965, Oct. 2023, doi: 10.1080/02602938.2022.2155611.
[50] E. Dringenberg, A. Kramer, and A. R. Betz, “Smartness in engineering: Beliefs of undergraduate engineering students,” *J. Eng. Educ.*, vol. 111, no. 3, pp. 575–594, Jul. 2022, doi: 10.1002/jee.20463.
[51] Y. Lewis and S. K. Nani, “Comparing Retrospective and Pre-Post Survey Measurement of First-Year African Engineering Students’ Skill and Perception Changes through a PBL Course”.

Appendix

Attention-check question		
The following questions pertain to the balsa wood bridge project exclusively. Only respond based on your experience completing the problem this semester. Assumption Definition: Engineering assumptions are decisions made to simplify a problem that fill in missing information. Example assumptions are negligible mass, concentrated force, negligible dimensions, steady state, rigid body, and frictionless. Select yes to show you read this.		
Scale		
Yes	Maybe	No

Instructions given to students					
Answer the following items using the scale provided based on the frequency that you completed them during each section of the balsa wood bridge project open-ended modeling problem (OEMP). Assumption Definition: Engineering assumptions are decisions made to simplify a problem that fill in missing information. Example assumptions are negligible mass, concentrated force, negligible dimensions, steady state, rigid body, and frictionless.					
Scale					
I never did this.	I did this once per section.	I did this sometimes, 1-2 times per section.	I did this often, 3-4 times per section.	I did this many times, 5+ times per section	I don't know/I'm unsure

Number	Question	EMJ Code
1_1	I made an assumption with manufacturability in mind.	1b
1_2	I made an assumption after considering the user or customer of the product/device.	1b

1_3	I made an assumption based on my personal experience with a same (or similar) product/device.	1c
1_4	I made an assumption to make the problem easier to solve.	1f
1_5	I made a new assumption or adjusted an assumption after attempting to solve the problem.	2a
1_6	I realized I didn't need an assumption after solving the problem.	1g
1_7	I considered if my assumption was reasonable.	2a
1_8	Based on my answers from previous sections, I considered if my answer made sense.	2b
1_9	I felt that some criteria from the assignment were more important than others.	PEP
1_10	I felt that some values (numbers, parameters, dimensions, etc.) were more important than others to designing the product/device.	PEP
1_11	I changed a number after calculating it with safety in mind.	3b
1_12	I changed a number after calculating it with manufacturability in mind.	3b
1_13	I changed a number after calculating it because I received instructor feedback.	2c, 3b
1_14	I used a technology tool to help understand and solve a part of the project (ex. MATLAB, Excel, SolidWorks, ANSYS, etc.)	4a
1_15	I considered if the result from the technology tool was reasonable.	4b

Instructions given to students							
Choose the option that best describes your opinion after completing the balsa wood bridge project OEMP this semester. Assumption Definition: Engineering assumptions are decisions made to simplify a problem that fill in missing information. Example assumptions are negligible mass, concentrated force, negligible dimensions, steady state, rigid body, and frictionless.							
Scale							
Strongly disagree	Somewhat disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Somewhat agree	Strongly agree	I don't know/I'm unsure

Number	**Question**	**EMJ Code**
2_1	I am better at making useful assumptions than I was before the project.	1
2_2	I am better at judging if my final answer makes sense based on the criteria of the problem.	2, PEP
2_3	I am better at solving for unknowns based on diagrams I drew.	All
2_4	I am better at solving for unknowns based on diagrams provided.	All
2_5	I feel more confident in my ability to judge if my diagram is accurate to real-life.	All
2_6	I am better prepared to solve real problems using math tools.	4
2_7	I am better at combining concepts from multiple courses to solve a single problem.	All

2_8	I improved my ability to decide what given information is most relevant.	PEP
2_9	Overall, I am better equipped to solve open-ended problems where the steps are not given.	All

PEP=Question sourced from professional engineer panel.