Abstract

Background: It is essential that engineering students have well-developed professional skills to meet the demands of a highly interdisciplinary and globalized workforce. Students' professional skills have been assessed in a variety of ways; however, many are limited by measuring self-reported attainment, not rooted in developmental approaches to learning and plagued by self-report bias.

Purpose: The purpose of this study is to examine the validity evidence for using the Professional Skills Opportunities survey to assess undergraduate engineering students' opportunities to practice professional skills. Using Dall'Alba's "ways of being," we developed the PSO survey, where engineering students self-report their opportunities to practice four professional skills: shared leadership, problem-solving, communication, and business and management skills.

Method: We distributed the PSO to 13 institutions using a probabilistic stratified sampling approach, focusing on four institution types. We established validity and fairness for the PSO through three studies: confirmatory factor analysis to establish the internal consistency of the factors, measurement invariance to establish similar interpretation across groups, and classical test theory methods to provide evidence of fairness for groups with small sample sizes.

Results: Results from the CFA found that the PSO can be scored by opportunities to practice an individual skill and total opportunities, and measurement invariance and classical test theory methods establish that the PSO can be used fairly across various student groups.

Conclusion: Based on the evidence of validity and fairness established, we recommend the PSO to researchers, faculty, and program administrators interested in understanding differences in opportunities to develop professional skills.

Keywords: *professional skills, validity, assessment, multi-institution*

A Validation Study for Measuring Opportunities to Learn Engineering Professional Skills

It is well-established that employers seek engineers who are not only technically prepared, but also adept in skills such as teamwork, leadership, and communication. Since the 1990s, implementing *professional skills* into engineering education has been a topic of discussion between industry professionals, professional societies, and academic institutions. Institutions such as the Accreditation Board for Engineering and Technology (ABET) and the National Academies of Engineering (NAE) have called attention to the need for professional skills and implemented engineering education standards emphasizing professional skills development (National Academy of Engineering, 2005; Volkwein et al., 2004). These organizations have a common understanding of professional skills as cognitive, non-technical skills needed to complement students' technical abilities, and we adopt this high-level understanding in our work. Recent reports indicate that employers seek professional skills in 81% of engineering job openings in the United States (America Succeeds et al., 2021), and a survey of the engineering profession shows that 30% of employers believe employees do not have the right professional skills for the workplace and 50% of employers believe there is an increasing shortage of professional skills in the workplace (Burner et al., 2019). In order to understand what professional skills engineering students are specifically developing and where the needs are, there needs to be assessment tools that enable large-scale studies across multiple institutions and contexts.

Professional skills have been assessed in a variety of ways with each method having implementation challenges. The main categories of professional skills assessment are development of reasoning (e.g., Engineering Ethical Reasoning Instrument, Zhu, et al., 2014), third-party evaluation (e.g., the Comprehensive Assessment for Team-Member Effectiveness, Ohland et al., 2012), behavior-based (e.g., the Global Engineering Competency-Situational Judgement Test (Jesiek, et al., 2020), self-rating scales (e.g., Global Engineering Competency, Mazzurco, et al., 2020), and professional preparedness (anonymized). Assessment tools that measure actual competency are extremely valuable, and development of reasoning, third-party evaluation, observation, and behavior provide a more direct assessment of actual skill than self-reports. Yet testing students across multiple domains of professional skills in a given program or course in these manners would be very expensive and cumbersome. For these reasons, researchers turn to self-assessment.

While simple to implement and valuable method of data collection, self-assessment of skill attainment is limited by students' perception by their competence. Students may not be able to discern where they are currently at in the continuum of skill development (Berry et al., 2022; Shuman et al., 2005). For example, students with intermediate levels of knowledge often report the highest level of confidence, and folks from collectivist cultures—cultures that value the needs of the community over the individual—commonly rate themselves more modestly than individualistic cultures (Atwater et al., 2009; Cullen et al., 2015; anonymized). Thus, one could mistakenly interpret lower mean scores to indicate students from collectivist cultures have lower attainment of professional skills. Yet, the problem is not inherently with surveying students or that the approach is a form of self-report, but rather that assessments have strong evidence establishing validity for how the results are used (Kane, 1992).

To be truly useful, measures of professional skills must be aligned with how researchers understand professional skills to be developed, learning theory, and how students can demonstrate their learning (anonymized; Pellegrino, 2002). Dall'Alba's (2009) ontological framework suggests that engineering students develop professional skills over time through multiple opportunities to practice, receive feedback, and reflect. Learning to be a professional comes from repetitively engaging and learning from others in the field, such as cocurricular's apprentice-style learning (Dall'Alba, 2009). For example, studies show that co-curricular and extracurricular settings provide intrinsically motivated, long-term, repeated learning opportunities in spaces tailored to encourage professional development (Fakhretdinova et al., 2021; Hinkle & Koretsky, 2019; Polmear et al., 2023). While many available assessments measure students' engagement or measure their attainment of professional skills, no assessment to date has measured how access to opportunities can build students' professional skills.

To open new avenues of research on the development of professional skills, we have developed the Professional Skills Opportunities survey. This assessment instrument was designed to measure engineering undergraduate students' perception of their opportunities to practice their professional skills based on the students' report of the frequency they practiced specific behaviors (related to the respective skills). Our novel approach to assessing professional skills can guide researchers' understanding of how students access opportunities to learn. Understanding learning as a developmental process that requires repetitive opportunities is consistent with national-level conversations stating that employees need more opportunities to practice teamwork, interdisciplinary skills and change-management skills (National Academies of Sciences, Engineering, and Medicine, 2022).

The purpose of this study is to examine the validity evidence for using the Professional Skills Opportunities (PSO) survey to assess undergraduate engineering students' opportunities to practice four professional skills. We seek to enable researchers, engineering programs, and student organizations to evaluate and enhance the learning opportunities available to their students. Specifically, we ask:

- To what extent do the PSO items function as conceptualized? This evidence is crucial to support the interpretation of summed items together as intended factors.
- To what extent does the PSO measure the same construct across binary gender identities and school year? This evidence is crucial to support the fairness of scores when used across undergraduate years and with men and women.
- To what extent does the PSO measure differences in opportunities to practice professional skills across gender identity, school year, racial identity, and institution type? This evidence is crucial to ensure that the variance is being captured, while not excluding groups due to sample size requirements.

Literature review

Professional Skills in Engineering

Professional skills have been operationalized in a variety of ways by different entities. ABET (2021), the National Academy of Engineering (2005), and the National Society of Professional Engineers (2022) have operationalized professional skills by providing lists of skills necessary for the formation of engineers. Despite differences, these lists overlap significantly, for example with respect to professional skills, such as problem-solving, communication, and teamwork skills. Research backs these emphases, such as evidence that employers and engineering professors view problem-solving skills as fundamental to the engineering profession and see them as vital because of how they work in conjunction with other professional skills, such as teamwork and communication (Passow & Passow, 2017). Likewise, multiple studies have identified communication skills, including team, cross-cultural, and interdisciplinary communication as among the most crucial skills for engineers to have (Barrett, 2006; Leandro Cruz & Saunders-Smits, 2022). Strong communication skills have been tied to an engineer's success in the workforce, with cross-cultural communication and technical communication skills being a substantial part of employability and effectiveness in the workforce (Norback et al., 2009). Similar to communication skills, teamwork skills are essential for being an effective engineer, with engineers needing the skills to be members of interdisciplinary and cross-cultural teams (Azmi et al., 2018; Leandro Cruz & Saunders-Smits, 2022). Work experiences such as internships have provided engineering students with enhanced analytical, problem-solving, and teamwork skills (Chan et al., 2017; Nogueira et al., 2021).

Both the engineering curriculum and cocurricular and extracurricular activities (e.g., professional societies and engineering clubs) are sources for opportunities to practice professional skills. Engineering undergraduates practice skills in capstone curriculum, Engineering Projects in Community Service (EPICS), professional societies, and engineering clubs and teams (e.g., Baja SAE, a program of the Society of Automotive Engineers [SAE]; Coyle et al., 2005; Dalrymple & Evangelou, 2006; Immekus et al., 2005; Olson, 2018). Courses such as capstone design promote the development of team skills as students work together on design projects, often with an industry sponsor (Dym et al., 2005). These courses are then supplemented by cocurricular activities that provide self-motivated opportunities and developing students' social capital through relationships with their peers (Garrett et al., 2021).

Opportunities to engage in authentic learning, such as within co-curricular and extracurricular activities, holistically develop students' professional skills since students best develop professional skills when they are integrated with technical skills in authentic engineering problems. Professional skills are most readily attained when taught in technical contexts, as professional and technical skill attainment may be connected (Hissey, 2000; R. Martin et al., 2005). For instance, EPICS and Formula Student SAE provide students with opportunities to practice multiple professional skills, such as teamwork, communication, business management, and ethics, while engaged in authentic engineering problems over semesters-long projects (Gadola & Chindamo, 2019; Hissey, 2000; Huff et al., 2016; Passow & Passow, 2017). While co-curricular and extracurricular spaces are an essential part of professional skill development for many engineering students, little has been done to assess the impact of these opportunities on professional skill development.

Differences in Access to Professional Skills

Not all engineering students have the same access to opportunities to develop social and technical professional skills (Polmear et al., 2023). Chilly organizational climates have been found to limit access to professional development opportunities for marginalized students; research has well established that such climates impact interest and persistence in engineering among these students (Camacho & Lord, 2011; Walton et al., 2015). For example, women have high levels of engagement in social out-of-class activities specific to engineering, such as engineering fraternities and sororities, engineering outreach support, and service and identity-based organizations; in contrast, men are more involved in technical out-of-class activities, such as professional societies, design competition teams, and professional experiences (Simmons et al., 2018). Design competitions and engineering-specific professional societies are predominately white and masculine spaces that reinforce stereotypes that Black and Brown students and women are not “technical enough” for technical engineering spaces (Polmear et al., 2023; Simmons et al., 2018; Smit & Fuchsberger, 2020). Racially minoritized students in engineering are more likely to be involved in undergraduate research and activities tied to identity-based organizations, such as outreach programs and professional societies (Polmear et al., 2023; Simmons et al., 2018). Identity-based organizations like minority engineering programs, the National Society of Black Engineers, and the Society of Women Engineers were developed to provide emotional and professional support to underrepresented and minoritized students in engineering (Harley, 2022). These organizations provide direct access to support and opportunities to develop professional skills, such as leadership and management, that might not be available to minoritized students in traditional, predominately white, and masculine engineering spaces (Garrett et al., 2021). Thus, to better prepare engineering students to become engineering professionals, researchers need strategies that allow for a holistic understanding of what skills are being developed and a tool able to capture differences in access to professional skill development opportunities across student groups.

Assessing Professional Skills

Professional skills are commonly assessed through the assessment of a single professional skill, either through asking students (or others) to rate their skill or through content analysis where instructors rate students demonstrating a particular professional skill. In a review of 28 studies on generic competence development in problem-based learning, Boelt et al. (2022) found that all but two of the studies utilized surveys or questionnaires that measured some form of perceived skill attainment across 14 generic competences or skills. For example, Beagon et al. (2019) performed an intervention-based study with pre- and post-testing of students' self-evaluation and reflection on their communication skills during a problem-based learning module. Most leadership skill assessments utilize self-rating scales (see Ahn et al., 2014; Knight & Novoselich, 2017; Park et al., 2022) with limited work exploring leadership skill development, behaviors or professional preparedness based assessments (see Özgen et al., 2013). Other skill, such as teamwork, are commonly assessed through instruments that have students perform group work then have their teammates rate their performance (see Ohland et al., 2012, Reid et al., 2016). Some skills rely more heavily on content analysis, such as problem-solving skills being commonly assessed through classroom assessments such as test essays and complex problem scenario tests (Ijtihadi & Vidákovich, 2022; Price et al., 2022) and also frequently through self-rating scales (see Chan et al., 2017; Phillips et al., 2019). Ethics assessments are typically assessed from a scenario or case study approach (see Zhu & Jesiek, 2017) or is assessed from self-rating scales (see Odom & Zoltowski, 2019; Zhu et al., 2014). While common across all professional skills, many of these self-rating tool assess students' perception of their professional skill competence. New avenues of research—such as measuring opportunities to practice—may enable researchers to speak to professional skill development with evidence (PSO scores) that speaks to the developmental nature of learning made possible through repeat opportunities (Use of PSO).

Interpretive Use Argument Approach to Validity

Our validation studies and research questions are based on the fundamental understanding that validity refers to how assessment instrument scores are interpreted and used (AERA/APA/NCME, 2014; Kane, 1992). Validity is not a checklist of analyses that assessment developers must all follow (Songer & Ruiz-Primo, 2012; Douglas & Purzer, 2015). The type of data collected, and analyses conducted are to be supportive of testing the plausibility of a particular use and interpretation of resulting scores. Following a similar rationale as scientific argumentation, validation studies are to be centered on establishing evidence to guide how resulting assessment scores can be justifiably used and interpreted. Kane's Interpretive Usage Argument (IUA) validation framework (1992; 2013a; 2013b) has become a well-established approach by likening assessment validation to legal argumentation, where validity is both a practical and scientific activity. This perspective requires assessment developers to identify the claims they plan to make about an assessment score (i.e., the score measures opportunities to practice communication skills) and then provide evidence for that claim (i.e., factor analysis).

Similar to previous validation approaches guided by Kane's validation framework (anonymized; Jorion et al., 2015), we summarized the PSO's use case and developed corresponding desired claims. Following the procedures outlined by Authors (anonymized), we summarized the purpose, use and inferences for the PSO in Figure 1 for researchers and administrators to utilize when determining if the PSO is suited for their needs. Succinctly, the purpose of the PSO is to assess undergraduate engineering students' opportunities to practice professional skills so that researchers and program administrators may use PSO scores to assess and grow engineering students' opportunities. From this summary and desired use, we then developed the following desired claims:

- **Desired Claim #1 – Overall opportunity:** Students' PSO scores can be used to indicate undergraduate engineering students' opportunities to develop professional skills. A single summated score would be reported as an indicator of the overall level of opportunities that students have to develop professional skills.
- **Desired Claim #2 – Opportunity for individual skills:** PSO sub-scores can be used as indicators of students' level of opportunities to develop skills in each of five categories of professional skills: shared leadership, communication, problem-solving, business and management principles, and ethics and professional responsibilities. This claim is about how well the factor scores in the survey represent opportunities that would foster skill development in each of those skills.
- **Desired Claim #3 – Subgroup comparisons:** Students' PSO scores (i.e., total score) and sub-scores can be used in reference to each other and to compare subgroups within US engineering schools, regardless of type of institution or student-level differences in school year and gender. This will allow administrators to make programmatic decisions to bolster professional skills training opportunities and to advance equitable opportunities across student programs. Researchers will be able to advance our understanding of engineering students' opportunities to develop professional skills nationally and use the PSO in conjunction with other instruments to establish relationships between opportunities and other constructs.

To establish evidence that supports our desired claims, we adapt Jorion et al.'s (2015) analytic framework for evaluating validity. This framework allows researchers to clearly map out how psychometric analyses can provide evidence for one or more desired claims, thus, generating a holistic view of how the current validation study can guide the interpretation of instrument scores. Using this analytical framework, our methodological approach is focused on critically thinking about what evidence speaks to the plausibility of a desired claim. We aligned the desired claims with the analysis (column 1) to the hypotheses associated with the each of the desire claims. Table 1 depicts how the analyses included in this study can support our desired claims for the PSO instrument.

Figure 1

Articulation of intended use for the PSO survey for assessing undergraduate engineering students' opportunities to practice professional skills

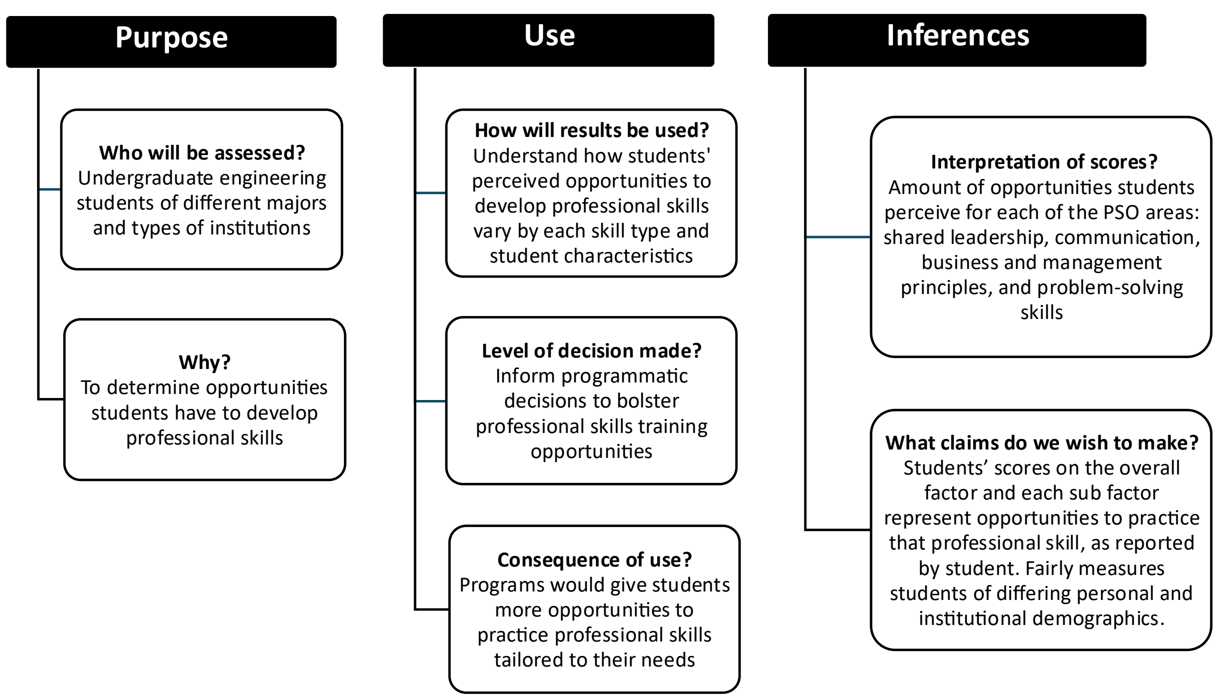


Table 1

Overview of how analyses conducted can support the desired claims of the PSO survey.

Analysis	**Claim 1 – Overall opportunity**	**Claim 2 – Opportunity for individual skills**	**Claim 3 – Subgroup comparison**
Confirmatory factor analysis (CFA)	Provides a structure of one overall second-order PSO factor that can explain intercorrelations between first-order professional skills.	Evaluates how well the proposed structure of five latent professional skills, including *shared leadership*, *ethics*, *business and management principles, and communication,* can explain the correlation among items present in the data.	--
Cronbach's alpha	Provides an estimation of total score reliability that reflects consistency in student scores across repeat adminstrtations of the survey.	--	--
Multi-group CFA (for measurement invariance)	--	--	Evaluates how factor structures have the equivalent organization, loadings, and intercepts, or residual within across different student groups that reflect the degree to which scores can be compared across student groups.
Item difficulty	Provides the average score on each item, indicating the average level of opportunities students have to practice different professional skills.	Provides item functioning information for each item within a professional skill.	Provides differences in item functioning information for each item across various student groups.

Item discrimination	Provides the correlation between item scores and the total PSO score. Results will show the items that are most effective in differentiating students with high opportunity levels from low opportunities levels to practice professional skills.	Provides item functioning information for each item within a professional skill.	Provides differences in item functioning information for each item across various student groups.

Note. For additional information on the larger project's instrument development, data collection, and data management, please refer to (anonymized). The final version of the PSO instrument is shown in the appendices (Appendix C).

Developing the Professional Skills Opportunities Survey

Selecting Professional Skills

We selected five professional skills (shared leadership, communication, problem-solving, business and management principles, and ethics and professional responsibilities) for assessment based on their established importance by the National Academy of Engineering, and ABET (ABET EAC, 2021; National Academy of Engineering, 2005; Passow & Passow, 2017) and prior work (anonymized) that indicated these particular skills were significantly influenced by opportunities in cocurricular programs. This prior work (anonymized) focused on the processes by which African American engineering students developed Engineer of 2020 traits (NAE, 2004) via ethnic-specific cocurricular activities such as the National Society for Black Engineers (NSBE). Participation in these cocurricular organizations provided ample opportunities for students to learn and to practice the key traits identified by the NAE. Serving in a leadership role often involved significant responsibility to peers and others in the organization; they had to satisfy multiple stakeholders, resolve conflicts, manage time and resources, and communicate with multiple audiences. Leadership positions came with responsibilities for the organization's reputation and finances that engendered professional and ethical responsibility (anonymized). Table 2 provides the professional skill and definition established from the literature.

Table 2

PSO construct definition of professional skill, with citations.

Professional Skill	Definition
Shared leadership	Engineering students' capacity to demonstrate commitment to learning, drive for excellence, integrity, and result orientation while working with others, translating to the abilities to treat others with good intention and respect, motivate others, assist in others' development, encourage others to stay on goals, and take responsibility of continuous self-improvement (Özgen et al., 2013; Wang et al., 2011).
Communication	Engineering students' development of written and oral skills to convey information and express opinions to audiences, and tailor their communication according to different situations using a variety of communication formats, including presentations, emails, letters, reports, via digital platforms, etc. (Cegala et al., 1998; Kim et al., 2001; Suter et al., 2009).
Problem-solving	Engineering students' development of the ability to generate, conceptualize, implement, and optimize original and applicable solutions using cognitive skills, including problem finding, ideation, evaluation, convergent thinking, divergent thinking, constraint analysis, and optimization (Basadur & Gelade, 2003; Charyton & Merrill, 2009).
Business & management principles	Engineering students' development of skills related to executing tasks to meet the priorities established by management, translating to the ability to manage financial, human resources, and time appropriately, demonstrate basic knowledge of the laws and regulations associated with the engineering design process and products, understand various stakeholders' needs, and analyze future needs that might emerge from stakeholders and the market (Cather et al., 2001; Mejtoft, 2016).
Ethics and professional responsibilities	Engineering students' development of personal awareness of ethical and professional obligations to their organization, customers, and society (i.e., mindfulness of reputation and their impacts, and accountability for long-term results) and social considerations during the engineering problem-solving process, translating to the ability to analyze social issues from professional perspectives and engage in professional activities objectively and truthfully (Canney & Bielefeldt, 2016; National Society of Professional Engineers, 2022).

Selecting a Theoretical Framework - Ways of Being and Opportunities to Practice Professional Skills

Dall'Alba's (2009) "Ways of Being" framework provides the theoretical basis for our assessment by emphasizing the developmental (ontological) aspect of professional identity alongside knowledge and skills (epistemological). Vu and Dall'Alba (2024) stressed that students need frequent practice of their skills as they learn to become professionals, and posit that professional formation occurs through repeated opportunities to learn. This mirrors how professional skills, like leadership, are acquired: through iterative practice and feedback, not just theoretical learning (Knight & Novoselich, 2017). Engineering students often engage in such apprenticeship style practice through co-curricular activities (Fakhretdinova et al., 2021; Hinkle & Koretsky, 2019; Polmear et al., 2023).

While it is impractical to measure the development of professional skills at a large-scale and the quality of opportunities provided, Dall'Alba posits that multiple opportunities are required to practice becoming professionals. Thus, we focus on how to quantify students' opportunities. Different from asking students to assess their opportunities or skills, we seek to collect a composite of opportunities in terms of reported frequency of various aspects of each professional skill. The frequency scale directly operationalizes Dall'Alba's (2009) framework by quantifying practice opportunities. Authors (anonymized) successfully employed this framework, using a frequency scale to measure doctoral students' opportunities for professional development in research, demonstrating strong validity and reliability (Cronbach's alpha: α = 0.87-0.90). Notably, Authors (anonymized) found disparities in these opportunities, with racially minoritized doctoral students reporting fewer instances of opportunities to develop professional skills. This highlights the importance of measuring opportunities, so programs can take steps to ensure all students have opportunities to develop professional skills.

Instrument Design & Previous Work

Different from self-assessments where the respondent would be asked to rate how often they had opportunities to practice each type of professional skill, we sought to assess the level of opportunities based on students' report of the frequency they performed specific behaviors. Conceptually, our goal was to create a survey that resulted in factor scores that would be used by researchers and program administrators to make inferences about opportunities undergraduate engineering students have to practice professional skills.

We designed and iteratively developed the PSO according to the four steps proposed by Netemeyer et al. (2003) in *Scaling Procedures.* These steps are 1) construct definition, 2) judging measurement items, 3) designing and conducting pilot studies for development, and 4) finalizing scale. This manuscript reports the finalizing step of the scale (according to the intended uses and inferences) with data which has not been published elsewhere.

The first three steps have been reported in conference proceedings (anonymized), thus we will briefly summarize here. To define each construct, we conducted a comprehensive literature review to identify relevant professional skills and the activities that develop them in undergraduate engineering education (anonymized). Although existing instruments assess students' self-reported skills or measure single skill domains, we found a gap in tools that capture students' opportunities to develop these skills. Recognizing that literature-based definitions may differ from how students understand these constructs, we treated each category as a latent construct, using students' endorsement of specific behaviors to infer their opportunities for skill development. From this framing, our aim was not to measure every possible activity or behavior but to create items representative of each construct. For example, instead of asking about opportunities to practice "problem-solving," students rated the frequency of behaviors practiced such as "*optimize your solution(s) when working on a project*" and "*evaluate the feasibility of ideas generated when working on a project.*" All items were rated on a 7-point Likert-type scale (1 = not at all frequent; 7 = very frequent).

For judging measurement items, we organized a panel of experts in both engineering education and psychometric assessment for expert review and judging of the construct definition and item alignment (anonymized; Table 1 & Table C in Appendix C). Based on their feedback, we revised items and then conducted cognitive think-aloud interviews with engineering students from diverse backgrounds to determine how well students would interpret the items as we had intended (anonymized). We revised items again after findings from the think alounds. We designed and conducted pilot testing for the purpose of examining how well the items factored according to the constructs they had been written through exploratory factor analysis (EFA). While most of the items functioned quite well, some items in the ethics, leadership, teamwork factors performed in ways unanticipated (anonymized). We discussed the findings and relevant literature within our team and our advisory board, which resulted in redefinining the constructs of leadership and team into one construct of shared leadership. We revised ethics items with poor performance, seeking to improve the fit without losing coverage the construct.

Methods

Positionality Statement

Our positionalities, inclusive of our background, experiences and interests, have motivated the work we present here. Throughout the process, we frequently reflected as a team on the influence of our positionalities on how we understand the value of this work. The first author, a white man who is a professor of engineering practice, has conducted research on graduate students' professional development and has a vested interest in how students develop professional skills. His experience as a graduate student motivated him to examine the research experiences of engineering Ph.D. students across engineering disciplines. His work demonstrated that understanding students' opportunities to practice is essential for understanding their professional development. The second author, a white woman who is an associate professor in engineering education, has engaged in developing assessment instruments for more than 10 assessments used in engineering education. In the development process, we drew on her experience as a researcher who has iteratively developed numerous assessment instruments and her expertise in the imperfect processes that comprise data collection, assessment validation, and assessment usage. Similarly, two authors, graduate students in engineering education who identify as a white nonbinary person and a Chinese woman, respectively, have considered the impact of their positionalities when understanding how we engage in developing evidence of validity and fairness in engineering education assessment instruments. The team's experiences with assessment development have motivated us to share the process of developing the PSO assessment instrument, including its limitations, recommendations for appropriate usage, and evidence of validity. Lastly, one author, a white woman who is a professor in engineering education, is an expert on how engineering students develop and leverage their social capital. Her perspectives on student development and broadening participation have motivated our team to focus on the social networks and opportunities embedded in those networks that influence students' opportunities to practice professional skills.

Research Setting & Data Collection

Selection of Data Sites and Data Collection

In April 2022, we administered the revised survey to students at 13 U.S. higher education institutions. These 13 institutions were selected via probabilistic stratified sampling, which involves selecting sub-groups from different strata to strive for equal representation of each stratum in the sample (Blair & Blair, 2014). We did so because institution type (strata) bore direct relevance to our research questions and there are important differences within and between strata in terms of cost of data collection and existent information about each stratum. Likewise, we expected to find differences between institution type because of students' differences in terms of opportunities to develop professional skills based on factors such as their social network and the activities accessible to them, and their levels of engagement in those activities. Probabilistic stratified sampling offers equal representation of students across different types of higher education institutions at which students have differing opportunities to practice professional skills.

In this study, we determined the strata (institution type) based on the Carnegie Classification (Indiana University Center for Postsecondary Research, n.d.):

- Research—doctoral universities with very high research activity
- Undergraduate—teaching-focused with exclusive or very high undergraduate populations

- Hispanic serving and minority serving institutions (HSI/MSIs)—institutions with at least 25% Hispanic enrollment
- Historically Black colleges and universities (HBCUs)—institutions founded with a mission to educate African American students

We categorized every U.S. institution with an ABET-accredited engineering program according to the institution type. (While Carnegie made changes to the classification system in 2025, our study was already underway at the time of their publication.) Then, we randomly selected three institutions within each institution type and contacted them for data collection. Administrators at all 12 institutions assented, so we administered the PSO survey via Qualtrics to all currently enrolled engineering undergraduate students in the selected institutions. The number of participants from each stratum is available in Table 3. Each survey respondent was awarded an incentive of $10 if they completed the survey. Having received fewer responses from HBCUs than the other institution types, we contacted an additional HBCU institution and increased the incentive to $20—increasing our total data collection to 13 institutions. In total, we received 2,246 responses.

Data Preprocessing & Research Participants

Before analysis, we performed data cleaning and preprocessing to prepare our dataset for latent analyses. We screened the survey responses using two criteria. The first criterion looks at the survey completion rate. We deleted all the responses with less than a 50% completion rate, which excluded 658 responses. We filtered the remaining sample using a filter question asking respondents to select “Not at all” as their response. All responses that did not pass the filter question were excluded from the data. This criterion reduced 354 responses from the sample. The final dataset has 1,234 survey responses. Table 3 contains the demographic information of the participants included in the final data set.

Table 3

Demographic Information and Descriptive Statistics for Research Participants

Measure	n	%
Gender		
Women	522	42
Men	678	55
Other[a]	34	3
Race/Ethnicity		
White	648	52
Asian	206	17
African American	123	10
Hispanic/Latino	172	14
Other[b]	85	7
Institution Type		
Research	460	37
UG	336	27
MSI/HSI	310	26
HBCU	128	10
School year		
First year	317	26
Second year	239	19
Third year	305	25
Fourth year	273	22
Fourth year+	93	7
N/A	7	1

Note. n = 1,234. We collected participant demographics in the survey.

[a] Other gender includes students who indicated nonbinary or N/A as their responses.

[b] Other race/ethnicity includes students who identified as multi-racial, Native Americans, Pacific Islanders, Arabic/Middle Eastern, and other as their race/ethnicities.

Study 1 – Confirmatory Factor Analysis

Data Fitness for Confirmatory Factor Analysis

The first study examines whether PSO items written to measure different professional skills have empirical internal consistency. We conducted confirmatory factor analysis (CFA) to investigate this hypothesis. As a multivariate analysis, CFA allows researchers to analyze "how measured variables reflect certain latent variables" using a pre-specified measurement model (Thompson, 2004, p. 110).

Prior to conducting CFA, we conducted several analyses to see if any items should be omitted from the factor analysis (full set of items shown in Appendix C). We computed the descriptive statistics of scores for each item and determined the skewness and kurtosis were within acceptable ranges (skewness within ±+/- 3 kurtosis within +/- 10) for factor analysis, however noted that the means were not normally distributed. Additionally, we computed the bivariate correlation matrix of the theoretically hypothesized factors separately and looked for poorly correlated items. We used the threshold of less than 0.3 to judge the correlation coefficient. The only factor that required closer inspection was the ethics factor. The correlation matrix of the ethics items is shown in Table 4. We decided not to exclude any of these items at this stage, as Q23, Q24, and Q25, were modestly correlated.

Table 4

Bivariate Correlation Matrix of Ethics Items

	Q22	Q23	Q24	Q25
Q22	-	-	-	-
Q23	0.33	-	-	-
Q24	0.33	0.34	-	-
Q25	0.29	0.28	0.51	-

Next, we looked at the Cronbach's alpha to determine the internal consistency of the PSO survey, treating 0.7 or larger as an indication of survey reliability (Nunnally, 1978; Taber, 2018). We first looked at the Cronbach's alphas for the entire survey and then calculated the Cronbach's alphas with every item removed. All the Cronbach's alpha values were within the acceptable range (ranging from α = 0.73 to α = 0.87). Therefore, no items were excluded from the CFA

Confirmatory Factor Analysis

We performed the CFA using R's *lavaan* package. Although CFA is a confirmatory analysis, it can be used when researchers have some knowledge of the underlying theoretical structure of the measurement and plan to utilize both the underlying theories and the empirical data to test hypothesized models (Thompson, 2004). As the PSO survey questions use a 7-point Likert-scale inquiring about frequency levels and our data is ordinal in nature, we used the weighted least square with mean and variance adjusted estimator (WLSMV). We chose the WLSMV because the data is ordinal, not truly continuous and non-normal. WLSMV is robust to deviations of mean normality by adjusting to bias in standard errors (Muthén & Muthén, 2010).

We examined both absolute and relative goodness of fit indices to evaluate the fitness between the empirical data and the proposed measurement models. The absolute fitness indices evaluated the fitness of the model itself without comparing it to a reference baseline model (Xia & Yang, 2019). The absolute fitness indices we examined included normed Chi-square (χ^2/df), the Bayesian information criterion (BIC), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). Relative fitness indices, on the other hand, evaluate the model fitness by comparing the specified measurement model to a baseline reference where all measured variables are assumed to be unrelated (Xia & Yang, 2019). The relative fitness indices we looked at were the Tucker-Lewis Index (TLI) and comparative fit index (CFI). According to Thompson (2004), a smaller normed Chi-square indicates a better model fit. SRMR and RMSEA values should be as close to zero as possible as they measure error of approximation and residual, with a value lower than .08 for both indices indicating acceptable fit (Hu and Bentler, 1999). The CFI and TLI values are deemed acceptable if they are larger than .9 (Hu and Bentler, 1999). Lastly, a model's BIC should be negative, and a smaller value indicates better parsimony (Boykin et al., 2023).

Study 2 – Measurement Invariance Test

In Study 2, we examined whether the PSO survey measures the same constructs across students from diverse backgrounds (H_2). Our investigation of this hypothesis will inform the proper use of the instrument and interpretations of the survey results—that is, if scores can be compared across student groups fairly. It also has the potential to point out potential differences in how students define, interpret, and conceptualize professional skills based on empirical evidence. We used the factor model determined as plausible based on the CFA results and conduct a measurement invariance test to test H_2. More specifically, we conducted measurement invariance tests with our data among different student groups, including gender, school year, and institution type (i.e., strata).

The measurement invariance test reveals whether invariance is achieved on several levels and provides information on the extent to which an instrument evaluates the same latent constructs across different groups of users (Vandenberg & Lance, 2000). With measurement invariance testing, progression to the next level only occurs if invariance in the previous level is confirmed. If invariance is not achieved on a certain level, it is not necessary to proceed to the next level, as the next invariance level uses the previous level as an underlying assumption. The order of the invariance levels we tested is shown in Table 5.

For a first-order model, measurement invariance should be tested in a hierarchical order, starting from the most basic level, the configural invariance. Achieving configural invariance means a stable factor structure is realized between items and the latent variables, indicating that the instrument measures similar latent constructs across different groups of respondents. The next step is to test for metric invariance, which translates to equal factor loadings across different respondent groups. Then, scalar invariance is tested to investigate whether equal factor loading and equal item intercepts can be observed across groups. Lastly, a test on residual invariance is recommended to examine whether equal item residuals are evident in the data across groups. For models with a second-order factor, additional tests for the second-order variable need to be included (Chen et al., 2005). In this paper, we follow the measurement invariance test order recommended by Chen et al. (2005), a study where they demonstrated the necessary procedure of conducting such analysis with instrument results with a second-order factor model.

Table 5

Order of Measurement Invariance Test for the Second-Order Factor Model in a PSO Survey

Model	Invariance tested
1	Configural invariance
2	First-order metric invariance
3	First- and second-order metric invariance
4	First- and second-order metric and scalar invariance

5	First- and second-order metric and scalar invariance, and first-order factors invariant
6	First- and second-order metric, scalar, and disturbances of first-order factors invariant
7	First- and second-order metric, scalar, disturbances of first-order factors, and residual variances of measured variables invariant

We conducted the measurement invariance tests using R's *semTools* package and the *sem()* function to specify the different models that need to be compared during the measurement invariance tests as recommended by Chen et al. (2005) for second-order factor structures. We continue to use the same estimator in Study 1, WLSMV, during model specification. Apart from evaluating the same indices used in CFA result interpretation including chi-square value, RMSEA, and CFI, one other index was also included—the difference of chi-square value between the different levels of invariance. To compare the goodness of fit between adjacent models, we used the *compareFit()* command in R.

Prior to the analysis, we examined our data more to evaluate its fitness for the measurement invariance test. Among the different genders, school years, and institution types, some of the subgroup sizes in our data were small. These smaller groups include students who identified as nonbinary or "other" as their gender (n = 39), students who study in HBCUs (n = 128), and undergraduate students who have been studying at their institution for more than four years (i.e., Fourth year+; n = 94). Past literature indicated that a small sample size (n = 100-200) could negatively impact the statistical power and precision of the measurement invariance tests (Meade et al., 2008; Meade & Bauer, 2007). Our sample allowed measurement invariance tests on the following groups:

1. Gender—women and men; data for students who self-identify as non-binary or of other genders were not included in Study 2 (see Study 3).
2. School year—first year, second year, third year, and fourth year and higher (a combination of fourth year and fifth year+ undergraduate students, as students from these two years are near enough to completing their undergraduate engineering programs that they likely interpret the professional skills measured in the PSO survey in similar ways).

To avoid overrepresentation within the dataset, we identified the smallest sub-group size from each grouping category and randomly selected a group of equal size from the larger groups. Thus, the dataset for different gender groups contains 1,244 data points (622 responses from women and men). The dataset for different years-in-school contains 1,392 data points (348 responses from each of the four years).

Study 3 – Fairness and Group Based-Mean Differences With Classical Test Theory

To consider evidence of fairness for students with minoritized identities (e.g., students with minoritized racial and gender identities), we used CTT techniques to perform basic item analysis including item difficulty and item discrimination. CTT methods yield valuable information about the instrument analyzed, especially when sample/group size does not meet the requirements for more complex psychometric analyses. CTT is not typically used for examining evidence of fairness; however, it is commonly used to evaluate item functioning for evidence of validity at a whole group level and for smaller sample sizes (e.g., classrooms). Thus, from the understanding that fairness is a property of validity, where unfairness is considered construct irrelevance (AERA/APA/NCME, 2014), we evaluated how an item functions on subgroups within our sample with established criteria for retaining or deleting an item (DeVellis, 2006). Fundamentally, the evidence we are seeking is how well the items function in terms of difficulty and discrimination for subgroups in our sample, where sample sizes are less than ideal. We used the entire dataset (*n* = 1,592), with a focus on student groups who were excluded from Study 2.

We first computed the item difficulty parameter by calculating the item mean across different student groups. In the context of the PSO survey, item difficulty can be interpreted as the extent of endorsement respondents exhibited towards the survey items (DeVellis, 2006). A higher item mean indicates a lower item difficulty in the PSO (i.e., students on average percieve many opportunities to practice that behavior). As the CTT is unable to distinguish true differences from measurement bias, we interpret the item difficulty parameter to have two potential meanings: as evidence of fairness or bias and as evidence of true differences in students' perceived opportunity to practice professional skills. Our adjusted interpretation is guided by existing findings showing that gender, racial, and socio-economic groups tend to differ in their professional skill development (Polmear et al., 2023; Simmons et al., 2018). Little guidance is provided on how to judge polytomous item difficulty using CTT. For our study, we are using guidance for dichotomous items, where difficulty values should fall within about 30% to 80% of the available range (Allen & Yen, 2001). Thus, for PSO items, difficulty values should fall between 2.1 and 5.6. Items outside of this ideal difficulty range should be reviewed holistically to determine if items should be revised or removed. In other words, items with group means outside of the acceptable range should be evaluated in comparison to prior literature and empirical data to determine whether the out-of-bound difficulty is due to item bias or true difference in students' experiences.

Item discrimination, as an indicator of item reliability, refers to the strength of association of one item with the other items in the instrument (DeVellis, 2006). As a result, item discrimination is calculated in terms of correlation. Thus, in this study, we use corrected item-total polyserial correlation of individual items as indicators of item discrimination best suited for ordinal data (Guo et al., 2022; Olsson et al., 1982). We computed a series of corrected total PSO scores excluding every item and calculated the polyserial correlation coefficient between the excluded item score and the corresponding corrected total PSO score. Generally, a higher correlation between the item and the rest of the instrument indicates that such an item is more apt at discerning between respondents with various levels of survey scores. The acceptable range for the correlation coefficient is above 0.3 (Ferketich, 1991).

Results

Study 1 – Confirmatory Factor Analysis

Based on the convergence issues due to the *ethics* factor, we excluded the ethics items from the CFA factor structure. The CFA for Model 1, a first order model with all five factors, did not converge when tested with the current dataset. To determine which items caused the issue of non-convergence, we systematically removed items to test whether the measurement model converged. The results showed that the ethics items were the cause of the issue; the model converged if it included no *ethics* factor or an *ethics* factor with more than two items. Among the four ethics items, Q22 and Q23 would not load with Q24 or Q25. Thus, models with only Q22 and Q23 or with only Q24 and Q25 would converge. A closer look at the bivariate correlation matrix of the ethics items reveals the marginally acceptable correlation coefficient between Q22, Q23, and Q24 and the highly acceptable correlation coefficient between Q24 and Q25, which may explain why the factor with all ethics items failed to converge.

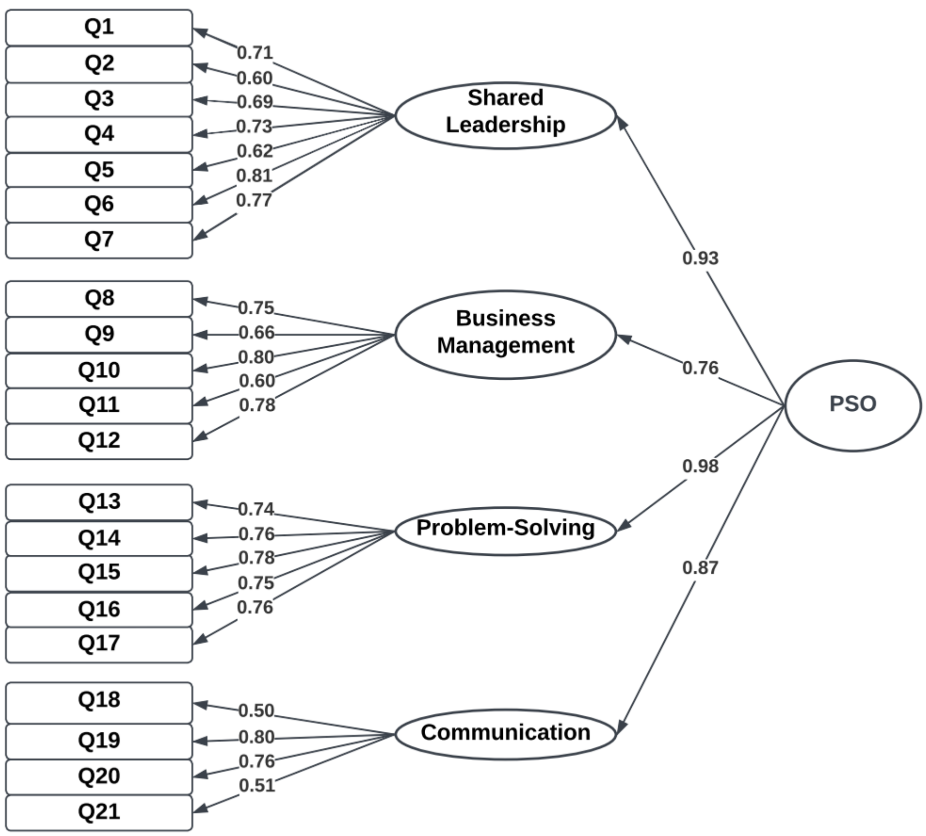
As documented in our previous work and the CFA results, the ethics factor in the PSO survey underwent the most revisions because the items failed to yield acceptable results in the factor analyses. Thus, instead of retaining a factor with only two ethics items, which is highly unlikely to represent the complexity of engineering ethics, we decided to exclude the ethics factor in our later analyses. Certainly, it is vital that students practice engineering ethical conduct; unfortunately our study does not suggest we achieved an effective measurement of it. Thus we propose to establish validity evidence for the more stable factors in the PSO survey while continuing to investigate how to create more representative and better-performing ethics items.

Our second CFA model, a first-order factor structure with four factors, converged and produced acceptable goodness of fit indices. Using Model 2, we also specified covariances between items with similar wordings, as literature suggests that respondents tend to rate these items similarly (Netemeyer et al., 2003), and compared the goodness of fit indices for structures with and without covariances. The results showed that adding the covariance between two similarly worded items did not significantly improve the model fit. Thus, we decided to proceed with the model without any covariances. Finally, we added a higher-order PSO factor to represent students' overall opportunities to practice multiple professional skills (Model 3).

We conducted a third CFA model with a second order factor, *opportunities to practice professional skills*, which also converged with acceptable fit indices. Although adding the second-order PSO factor decreased the model fitness slightly compared to Model 2's statistics, we considered it necessary to have an overall construct to reflect students' opportunities to practice professional skills holistically. The goodness of fit indices for Models 2 and 3 appears in Table 6, and the final path diagram for the PSO survey (sans ethics factor) appears in Figure 2.

Figure 2

Path Diagram for PSO Survey With No Ethics Factor



Note. Factor loadings are standardized by the standard deviation of both the factor and the items.

Table 6

Comparison of CFA Fit Indices for PSO Survey

Model	χ^2	*df*	χ^2/df	CFI	TLI	RMSEA
Null						
Unstructured reference	37103.87	210	176.69	-	-	-
Model 2 – no 2nd order factor						
No covariance between any items	421.21	183	2.30	.994	.993	.036
Covariance between Problemsolving5 and Management5	414.16	182	2.28	.994	.993	.036
Model 3 – with 2nd order factor						
No covariance between any items	445.06	185	2.41	.993	.992	.037

Note. n = 1592. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation.

Study 2 – Measurement invariance test

The measurement invariance test (i.e., multi-group CFA) results indicated that the PSO items have metric invariance for the first order factor structure across gender and school year. Using the CFA results, we performed measurement invariance testing on the first order factor structure, excluding the ethics factor. Our data revealed that metric invariance in the 1st order factor structure was achieved between undergraduate men and women in engineering. This result indicated that students similarly interpret the PSO items regardless of gender, attributing the items to the skills to a similar degree. The same trend can be observed among students from different school years, in all four groups. Table 7 contains the results from our measurement invariance tests.

Table 7

Measurement Invariance Results With Randomly Selected Sample From Larger Sub-Groups

Model	χ^2	*df*	RMSEA	CFI	$\Delta\chi^2$
Gender (*n* = 1553)					
Configural	470.7	364	.036	.989	-
1st order metric	518.1	381	.036	.989	16.64
1st & 2nd order metric	571.68	388	.037	.988	18.36*
School year (*n* = 1585)					
Configural	559.5	728	.036	.990	-
1st order metric	688.9	779	.034	.990	48.49
1st & 2nd order metric[a]	-	-	-	-	-

Note. * = $p < 0.05$.

[a] Models for at least one subgroup did not converge.

Study 3 – Classical Test Theory

Understanding that human-subjects data often violates CTT's assumptions of a true score, our difficulty scores have two possible interpretations: the first is evidence of fairness or bias (i.e., interpreting the difficulty scores as intended) and the second is evidence of actual difference in opportunities (i.e. assuming differences in scores due to socio-environmental influences). Thus, we share both an item-level performance interpretation of the scores through difficulty and discrimination and an interpretation of the group-based means as differences in opportunities.

Difficulty

Generally, item difficulty parameters were within the desired range of difficulty, with some items being overly easy to endorse (see Table 8). The item difficulty parameters for each item were calculated using the mean score for each group, with ideal difficulty parameters ranging between 2.1 to 5.6 out of 7 (Allen & Yen, 2001). Items in the *shared leadership, problem-solving, and communication* factors had moderate to low levels of difficulty (i.e., moderately to high levels of opportunity); the item response difficulty parameters ranged from 4.51 to 6.24. Four items had difficulty parameters outside of the ideal difficulty range: items Q5, *accept the responsibility for your personal growth*, Q15, *identify a problem that needs to be solved related to a project,* Q18, *adapt to the mode of communication*, and Q21, *use written formats of communication*. These four items with elevated means have been flagged for potential revision, as the mean scores for these items indicate that they do not capture students' true differences in opportunities to practice professional skills. All item difficulties are shown in Appendix A.

Discrimination

Overall, three factors had strong discrimination parameters, indicating items can distinguish between students with low levels of opportunity from high levels of opportunity (see Appendix B). Item-level discrimination values were calculated as a correlation between the score on that item and the total score, with discrimination parameters above .3 being acceptable (Ferketich, 1991). Hence, large discrimination parameters can be used to determine items that are highly correlated with the total score (i.e., a strong item) and poorly correlated with the total score (i.e., a poor item). Items in the *Shared Leadership, Problem-Solving,* and *Business and Management Principles* factors had high discrimination parameters ranging from .56 to .75. One factor, *Communication,* had two items with discrimination parameters near non-discriminating levels. Specifically, items Q18, *adapt to the mode of communication*, and Q21, *use written formats of communication*, had discrimination values of .45 and below (Table 8).

These items, Q18 and Q21, have low levels of difficulty and poor discrimination parameters, indicating they should be examined for potential revision.

Group based-mean differences in levels of opportunity

Group based-means at the item level can also be interpreted as true differences in students' access to opportunities. We analyzed group based-mean differences in access to opportunities by factor and demographic group. We found differences in the levels of perceived opportunities (by .3 and .5 mean point differences) across a single factor, institution type, school year, gender, and race and ethnicity (see Appendix A).

At the factor level, we found most items in the *Business and Management Principles* factor had differences in perceived opportunity across institution types. For example, students at MSI/HSI institutions reported the lowest level of perceived opportunities, with three items with a mean point difference of 0.5 or more and three items with a difference of 0.3 or more. Additionally, smaller differences (mean point differences of 0.3 or greater) were found within HBCU's and research institutions.

At the demographic level, we found evidence that some demographic groups had fewer opportunities across all types of professional skills. Across school year, the majority of items had differences between first year students and fourth year and above students, where often first year students reported lower levels of perceived opportunities than fourth+ year students. Seven items had differences of 0.5 and above and an additional 10 items had differences of 0.3 and above. Across the majority of items, nonbinary students reported lower levels than men and women. Seven items had a difference of 0.5 and greater, in addition to five items having a difference of 0.3 or greater. Within racial and ethnic identities, Black students reported having a lower level of perceived opportunities by a difference of 0.5 and greater for two items and 0.3 and greater for three items.

Table 8

Difficulty and Discrimination Values for Four Items With Low Difficulty, With Darker Color Indicating Lower Difficulty and Discrimination Parameters

Demographics	**Difficulty**				**Discrimination**			
	Q5	**Q15**	**Q18**	**Q21**	**Q5**	**Q15**	**Q18**	**Q21**
Total	5.74	5.54	5.83	5.98	0.59	0.74	0.45	0.47
Institution Type								
HBCU	5.88	5.38	5.91	5.95	0.51	0.68	0.42	0.44
MSI/HSI	5.77	5.39	5.85	5.80	0.52	0.77	0.37	0.57
Research	5.66	5.48	5.63	6.14	0.65	0.73	0.47	0.43
UG	5.75	5.65	5.90	5.98	0.59	0.76	0.5	0.46
Gender								
Female	5.66	5.51	5.74	5.85	0.62	0.74	0.39	0.45
Male	5.89	5.59	6.00	6.18	0.57	0.74	0.48	0.47
Nonbinary & Other	5.38	5.44	5.49	5.72	0.55	0.82	0.46	0.64
Race								
Asian	5.6	5.38	5.82	5.78	0.6	0.74	0.52	0.54
Black	5.84	5.36	5.98	5.92	0.53	0.72	0.36	0.41
Hispanic/Latino	5.79	5.48	5.97	5.92	0.58	0.8	0.33	0.61
Other	5.77	5.62	5.80	6.04	0.61	0.79	0.5	0.6
White	5.73	5.65	5.72	6.17	0.6	0.72	0.46	0.4
School year								
First year	5.69	5.37	5.38	5.65	0.61	0.72	0.39	0.42
Second year	5.68	5.49	5.82	5.92	0.61	0.77	0.48	0.44
Third year	5.74	5.53	5.96	6.10	0.56	0.75	0.48	0.5
Fourth year+	5.87	5.77	6.15	6.24	0.57	0.73	0.37	0.48

Discussion

Prior work on professional-skills assessment in engineering has largely fallen into two categories: measures of perceived competency and measures of specific skills. The PSO addresses the need for approaches to holistically assess students' opportunities to practice multiple, interacting professional skills (e.g., Barrett, 2006; Leandro Cruz & Saunders-Smits, 2022; Norback et al., 2009). Specifically, the PSO measures opportunities to practice in shared leadership, business and management principles, problem-solving, and communication—skills consistently identified as essential by research on professional-skills attainment, the National Academy of Engineering, and ABET (ABET EAC, 2021; National Academy of Engineering, 2005; Passow & Passow, 2017).

During the instrument's development, we sought multiple forms of validity evidence, such as cognitive interviews, expert reviews, EFA, to refine construct–item alignment, ensure fairness and clarity, and evaluate the emerging factor structure (anonymized). This study extends that validity argument of previous work by examining the PSO's factor structure and group differences using confirmatory factor analysis, measurement invariance testing (i.e., multi-group CFA), and classical test theory methods

Answering RQ1, our first-order and second-order CFA results indicate that the PSO instrument can be scored both in terms of specific, individual professional skills (i.e., *shared leadership, business and management principles*, *problem-solving*, and *communication*) and the overall opportunities to develop professional skills. For example, our CFA results confirm both our prior validation efforts (anonymized) and current literature finding that leadership and teamwork skills are highly correlated (Park et al., 2022). Thus, we loaded teamwork and leaderships skill items into a *shared leadership* factor and loaded the *shared leadership* factor (and the additional three first order factors) into a total *PSO* second order factor. This evidence provided by the CFA model results supported our desired claims of using the PSO for measuring students' total opportunities for practicing professional skills (desired claim #1) and opportunities to practice a single skill, such as *shared leadership* (desired claim #2).

During the CFA, we decided to exclude the *ethics* factor from this version of the PSO because the items did not converge in the first order CFA model. It is not uncommon to encounter validation issues when assessing latent traits (e.g., perceived opportunities) and it is understood that high-quality assessments take many iterations (anonymized). In our case, we found that the ethics items could not capture the true variance in students' opportunities to practice ethics skills. Based on previous studies showing that quantitatively measuring ethics is challenging due to broad and nuanced perceptions of ethics, it was challenging to capture ethics as a clear factor (Jesiek et al., 2022; McLeod et al., 2016) and thus decided to remove it until future work could establish more robust measurement. While it is common in validation studies to only report models that converge, we chose to share the iterative process of assessment instrument development and validation as we find it fundamental to our positionality as assessment development researchers.

Answering RQ2, we found that the PSO factor structure had measurement invariance across binary gender identity and students' school year, indicating that students interpret the PSO similarly across these groups. Past research on engineering students has established that participation in professional skill building opportunities (see Polmear et al., 2023; Simmons et al., 2018) and perceived professional skill attainment differ across student groups, such as gender (see Sperling et al., 2024) and school year (see Phillips et al., 2019). However, validity evidence supporting the comparison across groups is not common within engineering education (Douglas et al., 2016), thus limiting researchers' ability to make supported claims about differences in group scores. The evidence provided by the measurement invariance testing supported our desired claim (desired claim #3) that researchers using the PSO can compare total and individual professional skill scores across binary gender groups (i.e., men and women) and students' in-school year (e.g., comparing first and second year students) and guide interventions to support more equitable access to opportunities.

We found that most items fairly assessed students' opportunities to practice professional skills across gender, institutions, and race and ethnicity. We utilized CTT methods, item difficulty and discrimination, in addition to measurement invariance testing, to examine item performance for groups with smaller, uneven sample sizes. For most of the factors, item difficulty and discrimination parameters were within acceptable ranges, save for a few items that reported low levels of difficulty and discrimination. The communication factor included two items with low difficulty and discrimination parameters, indicating these items were easy to endorse even with low levels of opportunity and may not accurately capture students with low levels or opportunities from high levels of opportunity. The low difficulty in accessing communication practice opportunities is unsurprising as communication skills are commonly practiced in all types of assignments and fundamental to engineering and other professional skills such as leadership and teamwork (Fleming et al., 2024; Norback et al., 2009). These items have been examined holistically and retained, because they can capture the low levels of opportunity more likely to be reported by first year students. Similar to the study 2, these results supports comparing student scores across demographic groups (desired claim #3), but researchers may need to holistically consider students' opportunities to practice communication skills when evaluating scores from the *communication* subscore.

We interpreted the group-based means difference in opportunity as indicating that certain student groups had fewer opportunities to engage in professional skill development, and this was true across nearly all skills. Unsurprisingly, first year students reported the lowest level of opportunity across all factors, which aligns with previous studies asserting that first years may struggle to build the social capital needed to access opportunities through their social network (Brouwer et al., 2016). Marginalized student groups, such as nonbinary and Black students, reported fewer opportunities across many factors, aligning with previous studies that have found that minoritized students encounter unwelcoming environments that prevent them from engaging in opportunities and developing the social networks needed to develop their professional skills (Polmear et al., 2023; Simmons et al., 2018; Skvoretz et al., 2020; Smit & Fuchsberger, 2020; Smith et al., 2021). Additionally, our results point to different access to opportunities based on type of institution, specifically students at MSI/HSI institutions having fewer opportunities. This may be in part that students at HSI/MSI institutions reported lower levels of instrumental social capital, an established predictor of opportunities to practice professional skills (anonymized). The *Business and Management Principles* factor had more items with group-based differences based on institution type—potentially indicating that these skills are being deliberated fostered by specific institutions (e.g., undergraduate institutions).

We encourage researchers, faculty, and administrators to focus on the practical applications of this study, which center the use and scoring of the PSO. Our desired claims and corresponding validity evidence supports the PSO's use in several ways to measure engineering undergraduates' professional skills opportunities: 1) determine where students need more opportunities to practice, implementing programmatic, curricular, or co-curricular activities; 2) determine where students have ample opportunities to practice professional skills in order to emphasize those activities; 3) compare groups of students to understand which groups are accessing more/fewer opportunities. We note two practical implications regarding the scoring of the PSO: 1) the PSO is not intended as an individual diagnostic for any particular student's opportunities – scores should only be scored and reviewed in aggregate and in terms of average scores; 2) low average scores on the PSO should be interpreted very carefully – low scores on the PSO do not denote a deficiency of students' professional skills, but denote the potential for helping students find more opportunities to practice their professional skills. Please refer to the scoring guide (anonymized) for additional details.

Limitations

Due to the small and uneven size of the nonbinary students', the Black students', and the HBCU and MSI/HSI student groups, we were unable perform metric invariance testing to determine equal interpretation across groups. However, despite the limitations of our sample, we collected data across 13 institutions and even over-sampled, with increased participant stipends for select groups. The resulting small sample of racially minoritized students and nonbinary students in this study is a symptom of broader problem in U.S. engineering education and one that many engineering education researchers face. Engineering students as a population are so heavily dominated by few groups (Roy, 2019), so it is unlikely that many researchers will have a sample size large enough to robustly examine whether items are biased toward students who are underrepresented. Thus, studies typically do not include any evidence of fairness regarding students from minoritized backgrounds and may cite small samples as the rationale. Instead, we utilized CTT methods, specifically polytomous difficulty and discrimination parameters ideal for small sample sizes, to provide additional validity evidence.

CTT methods have limitations, such as not being sample invariant, meaning we are measuring both the difficulty of the items and the sample-specific characteristics of the group. For example, when we utilize the difficulty parameter to claim that certain professional skills items are more discriminating for certain groups, we cannot distinguish between the difficulty of the item and the level of opportunities to practice professional skills. As differences were found and we sought to understand whether it was actual differences or measurement bias, we turned to the literature and rationale. Following from the argument-based approach to validity, a lack of data to provide evidence of fairness (or bias) does not equate moving forward as though the instrument is indeed measuring similar constructs for different groups that have diverse cultures and social understandings (Kane, 1992, 2013a, 2013b). Thus, it is essential that researchers critically examine how assessment items function for subpopulations of students, rather than continue to exclude or ignore potential differences in their experiences. We encourage other researchers to include small, minoritized groups in their analysis by employing statistical testing methods that are inclusive of small sample sizes and approaching the evidence and claims made with the limitations of the analysis in mind.

Conclusion and Future Work

The PSO and similar surveys such as the [instrument name] (anonymized) provide a novel approach to understanding students' preparation for their professional careers. In this study, we approached the assessment of professional skills using a learning-centered and developmental perspective where skills are developed through opportunities for intentional and repetitive practice. It is well established that current methods for quantitatively measuring professional skills are limited in their ability to measure professional skills, given their reliance on subjective reporting of attainment and narrow focus on single skills. By shifting the assessment's focus from measuring students' confidence or others' confidence in their attainment of a single skill to measuring students' opportunities to engage, this paper encourages researchers, practitioners, and administrators to think strategically about how to increase access to professional skill building opportunities.

Results from previous work, including cognitive interviews and EFA (anonymized), and the evidence presented here establish evidence of the PSO's ability to assess students' opportunities to build shared leadership, business management and principles, problem-solving, and communication professional skills. Our CFA findings provide evidence for the scoring of engineering students' opportunities to practice individual professional skills (e.g., shared leadership) and overall, holistic professional skills. Additionally, based on our measurement invariance testing, the PSO can be used fairly to assess opportunities to practice professional skills across binary gender and student school year. Thus, users of the PSO can be confident when inferring meaning from students' opportunities to practice professional skills scores.

Based on our strong evidence of validity, we recommend the PSO to those interested in understanding the differences in groups of students' opportunities to practice professional skills. We recommend that users utilize the PSO as a research tool to assess groups of engineering students' opportunities—such as first year students, students in cocurricular programs, or women—to practice professional skills, rather than as a diagnostic tool to assess individual students' opportunities. From the results, researchers can recommend interventions to increase access to skill-building opportunities for particular university types or demographic groups.

We recommend scoring the PSO as an individual score for single skills, *shared leadership*, *business and management principles*, *problem-solving*, and *communication*, and as a total score (excluding the *ethics* factor). The scores for an individual skill were calculated by summing the scores for every item assessing that skill and normalizing the score with the number of items in that skill. As a result, each respondent has four scores (one for each professional skill), and all the scores are on a scale of 1 to 7. The overall PSO score is calculated by taking the sum of the five individual skill scores. Thus, the overall PSO score ranges from 4 to 28. Low PSO total score (i.e., a total score less than 14) indicates that students in a sample infrequently have opportunities to engage in activities that develop professional skills. Such low scores should not be used to infer student deficits, but rather deficits in their opportunities to grow their professional skills. High PSO scores (total score over 22) indicate students have many opportunities to engage in professional skills. For more details on scoring, please refer to the scoring guide (anonymized).

The PSO can also complement other measures, such as the ones we reviewed in the literature. Assessing students' professional skills often requires multiple measures, some of which are best done quantitatively and others qualitatively, or sometimes together. The PSO provides a new and unique way to frame the assessment of professional skills, solving the longheld concern with focusing on students reporting their skills attainment and instead, focusing on the opportunities to practice those skills. There will always be debate on “how well” students are at performing these complicated professional skills, and the PSO is a step towards answering those questions.

Future work could focus on how students’ professional skill opportunities promote their professional development and how other factors, such as their social capital, promote their access to these opportunities. Additionally, future work could reassess the ethics items by revisiting the literature, conducting additional expert reviews, and implementing cognitive interviews.

Acknowledgement

This material is based upon our work supported by the National Science Foundation under Grants No. (anonymized). Any opinions, findings, and conclusions or recommendations expressed in this material come from us as the authors of the paper and do not necessarily reflect the views of the National Science Foundation.

References

ABET. (2021, October). *Criteria for accrediting engineering programs*. https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2022-2023/

Ahn, B., Cox, M. F., London, J., Cekic, O., & Zhu, J. (2014). Creating an Instrument to Measure Leadership, Change, and Synthesis in Engineering Undergraduates. *Journal of Engineering Education, 103(1)*, 115–136. https://doi.org/10.1002/jee.20036

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.

America Succeeds & Emsi Burning Glass. (2021). The high demand for durable skills. https://americasucceeds.org/wp-content/uploads/2021/04/AmericaSucceeds-DurableSkills-NationalFactSheet-2021.pdf

Atwater, L., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with the relationship between self and others' ratings of leadership? *Journal of Applied Psychology*, *94*(4), 876. https://doi.org/10.1037/a0014561

Azmi, A. N., Kamin, Y., Noordin, M. K., & Nasir, A. N. M. (2018). Towards industrial revolution 4.0: Employers' expectations on fresh engineering graduates. *International Journal of Engineering & Technology*, *7*(4.28), 267–272.

Barrett, D. J. (2006). Strong communication skills a must for today's leaders. *Handbook of Business Strategy*, *7*(1), 385–390. https://doi.org/10.1108/10775730610619124

Basadur, M., & Gelade, G. (2003). Using the creative problem solving profile (CPSP) for diagnosing and solving real-world problems. *Emergence*, *5*(3), 22–47.

Beagon, Ú., Niall, D., & Ní Fhloinn, E. (2019). Problem-based learning: student perceptions of its value in developing professional skills for engineering practice. *European Journal of Engineering Education*, *44*(6), 850-865. https://doi.org/10.1080/03043797.2018.1536114

Berry, F. C., Huang, W., & Exter, M. (2022). Improving accuracy of self-and-peer assessment in engineering technology capstone. *IEEE Transactions on Education*, *66*(2), 174–187. https://doi.org/10.1109/TE.2022.3208219

Blair, E., & Blair, J. (2014). *Applied survey sampling*. Sage Publications.

Boykin A. A., Ezike N. C., Myers A. J. (2023) Model-data fit evaluation: posterior checks and Bayesian model selection. *International Encyclopedia of Education: Elsevier Science*, 279-289.

Boelt, A. M., Kolmos, A., & Holgaard, J. E. (2022). Literature review of students' perceptions of generic competence development in problem-based learning in engineering education. *European Journal of Engineering Education*, *47*(6), 1399-1420. https://doi.org/10.1080/03043797.2022.2074819

Brouwer, J., Jansen, E., Flache, A., & Hofman, A. (2016). The impact of social capital on self-efficacy and study success among first-year university students. *Learning & Individual Differences*, *52*, 109–118. https://doi.org/10.1016/j.lindif.2016.09.016

Burner, T., Supinski, L., Zhu, S., Robinson, S. and Supinski, C. (2019). *The global skills shortage: bridging the talent gap with education, training, and sourcing*. Society for Human Resource Management. https://www.shrm.org/content/dam/en/shrm/topics-tools/news/employee-relations/SHRM-Skills-Gap-2019.pdf

Camacho, M. M., & Lord, S. M. (2011). “Microaggressions” in engineering education: Climate for Asian, Latina and White women. In *2011 Frontiers in Education Conference (FIE)* (pp. S3H-1–S3H-6). https://doi.org/10.1109/FIE.2011.6142970.

Canney, N. E., & Bielefeldt, A. R. (2016). Validity and reliability evidence of the engineering professional responsibility assessment tool. *Journal of Engineering Education*, *105*(3), 452–477. https://doi.org/10.1002/jee.20124

Cather, H., Morris, R., & Wilkinson, J. (2001). *Business skills for engineers and technologists* (1st edition). Butterworth-Heinemann.

Cegala, D. J., Coleman, M. T., & Turner, J. W. (1998). The development and partial assessment of the medical communication competence scale. *Health Communication*, *10*(3), 261–288. https://doi.org/10.1207/s15327027hc1003_5

Chan, C. K., Zhao, Y., & Luk, L. Y. (2017). A validated and reliable instrument investigating engineering students’ perceptions of competency in generic skills. Journal of Engineering Education, 106(2), 299-325. https://doi.org/10.1002/jee.20165

Charyton, C., & Merrill, J. A. (2009). Assessing general creativity and creative engineering design in first year engineering students. *Journal of Engineering Education*, *98*(2), 145–156. https://doi.org/10.1002/j.2168-9830.2009.tb01013.x

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher’s corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7

Coyle, E. J., Jamieson, L. H., & Oakes, W. C. (2005). EPICS: Engineering projects in community service. *International Journal of Engineering Education*, *21*(1), 139–150.

Cruz, M. L., Saunders-Smits, G. N., & Groen, P. (2020). Evaluation of competency methods in engineering education: A systematic review. *European Journal of Engineering Education*, *45*(5), 729–757. https://doi.org/10.1080/03043797.2019.1671810

Cullen, K. L., Gentry, W. A., & Yammarino, F. J. (2015). Biased self-perception tendencies: Self-enhancement/self-diminishment and leader derailment in individualistic and collectivistic cultures. *Applied Psychology*, *64*(1), 161–207. https://doi.org/10.1111/apps.12026

Dall'Alba, G. (2009). Learning professional ways of being: Ambiguities of becoming. *Educational Philosophy and Theory*, *41*(1), 34–45. https://doi.org/10.1111/j.1469-5812.2008.00475.x

Dalrymple, O., & Evangelou, D. (2006). The role of extracurricular activities in the education of engineers. In *International Conference on Engineering Education* (pp. 24–30). ICEE.

DeVellis, R. F. (2006). Classical test theory. *Medical Care*, *44*(11), S50–S59.

Douglas, K. A., & Purzer, Ş. (2015). Validity: Meaning and relevancy in assessment for engineering education research. *Journal of Engineering Education*, *104*(2), 108-118. https://doi.org/10.1002/jee.20070

Douglas, K. A., Rynearson, A., Purzer, S., & Strobel, J. (2016). Reliability, validity, and fairness: A content analysis of assessment development publications in major engineering education journals. *The International journal of engineering education*, *32*(5), 1960-1971.

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education*, *94*(1), 103–120. https://doi.org/10.1002/j.2168-9830.2005.tb00832.x

Fakhretdinova, G., Osipov, P., & Dulalaeva, L. (2021). Extracurricular activities as an important tool in developing soft skills. In M. Auer & T. Rüütmann (Eds.), *Educating engineers for future industrial revolutions: Proceedings of the 23rd International Conference on Interactive Collaborative Learning (ICL2020)* (Vol. 2, pp. 480–487). Springer. https://doi.org/10.1007/978-3-030-68201-9_47

Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. *Research in Nursing & Health*, *14*(2), 165–168. https://doi.org/10.1002/nur.4770140211

Fleming, G. C., Klopfer, M., Katz, A., & Knight, D. (2024). What engineering employers want: An analysis of technical and professional skills in engineering job advertisements. *Journal of Engineering Education*, *113*(2), 251-279. https://doi.org/10.1002/jee.20581

Gadola, M., & Chindamo, D. (2019). Experiential learning in engineering education: The role of student design competitions and a case study. *International Journal of Mechanical Engineering Education*, *47*(1), 3–22. https://doi.org/10.1177/030641901774958

Garrett, S. D., Martin, J. P., & Adams, S. G. (2021). Developing nontechnical professional skills in African American engineering majors through co-curricular activities. *IEEE Transactions on Education*, *65*(3), 394–401. https://doi.org/10.1109/TE.2021.3120210

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative methods for item parameter estimation: From CTT to IRT. *ETS Research Report Series*, *2022*(1), 1–16. https://doi.org/10.1002/ets2.12355

Harley, R. (2022). *Black beyond measure: An antideficit exploration of cultural capital within a National Society of Black Engineers (NSBE) chapter at a predominantly white institution (PWI)* (Publication no. 931). [Doctoral dissertation, University of San Diego]. Digital USD.

Hinkle, C. M., & Koretsky, M. D. (2019). Toward professional practice: Student learning opportunities through participation in engineering clubs. *European Journal of Engineering Education*, *44*(6), 906–922. https://doi.org/10.1080/03043797.2018.1477119

Hissey, T. W. (2000). Education and careers 2000. Enhanced skills for engineers. *Proceedings of the IEEE*, *88*(8), 1367–1370. https://doi.org/10.1109/5.880089

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Huff, J. L., Zoltowski, C. B., & Oakes, W. C. (2016). Preparing engineers for the workplace through service learning: Perceptions of EPICS alumni. *Journal of Engineering Education*, *105*(1), 43–69.

Indiana University Center for Postsecondary Research. (n.d.) Carnegie classification of institutions of higher education. https://carnegieclassifications.acenet.edu/

Ijtihadi, K. A., & Vidákovich, T. (2022). Assessment in STEM Problem-Solving: A Systematic Review. *The International Journal of Assessment and Evaluation, 29*(2), 63-80. https://doi.org/10.18848/2327-7920/CGP/v29i02/63-80

Immekus, J., Maller, S., Tracy, S., & Oakes, W. (2005). *Evaluating the outcomes of a service learning based course in an engineering education program: Preliminary results of the assessment of the engineering projects in community service – EPICS.* Paper presented at the 2005 ASEE Annual Conference, Portland, OR. https://peer.asee.org/evaluating-the-outcomes-of-a-service-learning-based-course-in-an-engineering-education-program-preliminary-results-of-the-assessment-of-the-engineering-projects-in-community-service-epics

Jesiek, B., Claussen, S., Kim, D., Stepback, L., & Zoltowski, C. (2022). Exploring perceptions of ethics and social responsibility among engineering students and professionals: Research highlights and implications for the field. In *2022 ASEE Annual Conference & Exposition*. https://doi.org/10.18260/1-2--41360

Jesiek, B., Woo, S. E., Parrigon, S., & Porter, C. M. (2020). Development of a situational judgement test for global engineering competency. *Journal of Engineering Education*, *109*(3), 470–490. https://doi.org/10.1002/jee.20325

Jorion, N., Gane, B. D., James, K., Schroeder, L., DiBello, L. V., & Pellegrino, J. W. (2015). An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, *104*(4), 454-496. https://doi.org/10.1002/jee.20104

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

Kane, M. T. (2013a) The argument-based approach to validation. *School Psychology Review, 42*(4), 448-457, https://doi-org/10.1080/02796015.2013.12087465

Kane, M. T. (2013b). Validation as a pragmatic, scientific activity. *Journal of Educational Measurement*, *50*(1), 115-122, https://doi.org/10.1111/jedm.12007

Kim, S. C., Boren, D., & Solem, S. L. (2001). The Kim Alliance Scale: Development and preliminary testing. *Clinical Nursing Research*, *10*(3), 314–331.

Knight, D. B., & Novoselich, B. J. (2017). Curricular and co-curricular influences on undergraduate engineering student leadership. *Journal of Engineering Education*, *106*(1), 44–70. https://doi.org/10.1002/jee.20153

Leandro Cruz, M., & Saunders-Smits, G. N. (2022). Using an industry instrument to trigger the improvement of the transversal competency learning outcomes of engineering graduates. *European Journal of Engineering Education*, *47*(1), 30–49. https://doi.org/10.1080/03043797.2021.1909539

Martin, R., Maytham, B., Case, J., & Fraser, D. (2005). Engineering graduates' perceptions of how well they were prepared for work in industry. *European Journal of Engineering Education*, *30*(2), 167–180. https://doi.org/10.1080/03043790500087571

Mazzurco, A., Jesiek, B., & Godwin, A. (2020). Development of Global Engineering Competency Scale: Exploratory and confirmatory factor analysis. *Journal of Civil Engineering Education*, *146*(2). https://doi.org/10.1061/(ASCE)EI.2643-9115.0000006

McLeod, M. S., Payne, G. T., & Evert, R. E. (2016). Organizational ethics research: A systematic review of methods and analytical techniques. *Journal of Business Ethics*, *134*, 429–443. https://doi.org/10.1007/s10551-014-2436-9

Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 611–635. https://doi.org/10.1080/10705510701575461

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Mejtoft, T. (2016). *Integrating business skills in engineering education: Enhancing learning using a CDIO approach.* Paper presented at the 12th International CDIO Conference, Turku, Finland.

Muthén, L. K., & Muthén, B. O. (2010). *M*plus *user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.

National Academies of Sciences, Engineering, and Medicine. (2022). *Imagining the future of undergraduate STEM education: Proceedings of a virtual symposium* (K. Brenner, A. Beatty, & J. Alper, Eds.). The National Academies Press. https://doi.org/10.17226/26314

National Academy of Engineering. (2005). *Educating the engineer of 2020*. The National Academies Press. https://nae.edu/25677/Educating-the-Engineer-of-2020

National Society of Professional Engineers. (2022). *Code of ethics for engineers*.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Sage Publications.

Nogueira, T., Magano, J., Fontão, E., Sousa, M., & Leite, Â. (2021). Engineering students' industrial internship experience perception and satisfaction: Work experience scale validation. *Education Sciences*, *11*(11), 671. https://doi.org/10.3390/educsci11110671

Norback, J. S., Leeds, E. M., & Forehand, G. A. (2009). Engineering communication—Executive perspectives on the necessary skills for students. *International Journal of Modern Engineering*, *10*(1), 11–19.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Ohland, M. W., Loughry, M. L., Woehr, D. J., & Bullard, L. G. (2012). The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation. *Academy of Management Learning & Education*, *11*(4), 609–630. https://doi.org/10.5465/amle.2010.0177

Olson, S. (Comp.). (2018). *Engineering societies and undergraduate engineering education: Proceedings of a workshop*. National Academies Press. Washington, DC. https://doi.org/10.17226/24878.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*(3), 337–347. https://doi.org/10.1007/BF02294164

Özgen, S., Sánchez-Galofré, O., Alabart, J. R., Medir, M., & Giralt, F. (2013). Assessment of engineering students' leadership competencies. *Leadership and Management in Engineering*, *13*(2), 65–75. https://doi.org/10.1061/(ASCE)LM.1943-5630.0000168

Passow, H. J., & Passow, C. H. (2017). What competencies should undergraduate engineering programs emphasize? A systematic review. *Journal of Engineering Education*, *106*(3), 475–526. https://doi.org/10.1002/jee.20171

Pellegrino, J. W. (2002). Knowing what students know. *Issues in Science and Technology*, *19*(2), 48–52.

Phillips, A. R., & Lambie, C. (2019). Assessing civil engineering students perceptions of their problem solving ability. *International Journal of Engineering Education*, *35*(5), 1551-1560.

Polmear, M., Chance, S., Hadgraft, R. G., & Shaw, C. (2023). Informal learning as opportunity for competency development and broadened engagement in engineering. In A Johri (Ed.), *International handbook of engineering education research* (pp. 312–335). Routledge.

Price, A., Salehi, S., Burkholder, E., Kim, C., Isava, V., Flynn, M., & Wieman, C. (2022). An accurate and practical method for assessing science and engineering problem-solving expertise. *International Journal of Science Education*, *44*(13), 2061-2084. https://doi.org/10.1080/09500693.2022.2111668

Roy, J. (2019). Engineering by the numbers. In *American Society for Engineering Education* (pp. 1-40). American Society for Engineering Education.

Shuman, L. J., Besterfield-Sacre, M., & McGourty, J. (2005). The ABET “Professional Skills” — Can they be taught? Can they be assessed? *Journal of Engineering Education*, *94*(1), 41–55. https://doi.org/10.1002/j.2168-9830.2005.tb00828.x

Simmons, D. R., Van Mullekom, J., & Ohland, M. W. (2018). The popularity and intensity of engineering undergraduate out-of-class activities. *Journal of Engineering Education*, *107*(4), 611–635. https://doi.org/10.1002/jee.20235

Skvoretz, J., Kersaint, G., Campbell-Montalvo, R., Ware, J. D., Smith, C. A. S., Puccia, E., Martin, J. P., Lee, R., MacDonald, G., & Wao, H. (2020). Pursuing an engineering major: Social capital of women and underrepresented minorities. *Studies in Higher Education*, *45*(3), 592–607. https://doi.org/10.1080/03075079.2019.1609923

Smit, D., & Fuchsberger, V. (2020). Sprinkling diversity: Hurdles on the way to inclusiveness in makerspaces. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping experiences, shaping society* (pp. 1–8). Association for Computing Machinery.

Smith, C. A. S., Wao, H., Kersaint, G., Campbell-Montalvo, R., Gray-Ray, P., Puccia, E., Martin, J. P., Lee, R., Skvoretz, J., & MacDonald, G. (2021). Social capital from professional engineering organizations and the persistence of women and underrepresented minority undergraduates. *Frontiers in Sociology*, *6*, 671856. https://doi.org/10.3389/fsoc.2021.671856

Songer, N. B., & Ruiz-Primo, M. A. (2012). Assessment and science education: Our essential new priority?. *Journal of Research in Science Teaching*, *49*(6), 683-690. https://doi.org/10.1002/tea.21033

Sperling, J., Mburi, M., Gray, M., Schmid, L., & Saterbak, A. (2024). Effects of a first-year undergraduate engineering design course: Survey study of implications for student self-efficacy and professional skills, with focus on gender/sex and race/ethnicity. *International Journal of STEM Education*, *11*(1), 8. https://doi.org/10.1186/s40594-024-00467-6

Suter, E., Arndt, J., Arthur, N., Parboosingh, J., Taylor, E., & Deutschlander, S. (2009). Role understanding and effective communication as core competencies for collaborative practice. *Journal of Interprofessional Care*, *23*(1), 41–51. https://doi.org/10.1080/13561820802338579

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. American Psychological Association.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research.

Organizational Research Methods, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Volkwein, J. F., Lattuca, L. R., Terenzini, P. T., Strauss, L. C., & Sukhbaatar, J. A. (2004). Engineering change: A study of the impact of EC2000. *International Journal of Engineering Education*, *20*(3), 318-328.

Vu, T.T. & Dall'Alba, G. (2024). Becoming Authentic Professionals: Learning for Authenticity. In: Dall'Alba, G. (eds) Being and Becoming Through Higher Education. Debating Higher Education: Philosophical Perspectives, vol 13. Springer, Singapore. https://doi.org/10.1007/978-981-97-2062-0_5

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology*, *107*(2), 468. https://doi.org/10.1037/a0037461

Wang, J., Imbrie, P. K., & Lin, J. J. (2011). Work in progress—A feedback system for peer evaluation of engineering student teams to enhance team effectiveness. In *2011 Frontiers in Education Conference* (pp. S4C-1–S4C-5). IEEE.

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*, 409–428. https://doi.org/10.3758/s13428-018-1055-2

Zhang, M., Kranov, A. A., Beyerlein, S. W., McCormack, J. P., Pedrow, P. D., & Schmeckpeper, E. (2015). Investigating a scenario-based performance assessment of engineering professional skills. In *2015 IEEE Integrated STEM Education Conference* (pp. 230–235). IEEE Xplore.

Zhu, Q., Zoltowski, C. B., Feister, M. K., Buzzanell, P. M., Oakes, W. C., & Mead, A. D. (2014). *The development of an instrument for assessing individual ethical decisionmaking in project-based design teams: Integrating quantitative and qualitative methods.* Paper presented at the 2014 ASEE Annual Conference & Exposition, Indianapolis, IN.

Appendix A

Difficulty Values for all Items (Q1 – Q12)

Table A

Difficulty values for all items by demographic group.

Demographics	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Total	5.37	5.47	4.79	5.37	5.74	5.28	5.27	4.87	3.92	4.31	3.69	4.31
Institution Type												
HBCU	5.43	5.11	4.65	5.52	5.88	5.22	5.48	4.91	3.67	4.38	3.8	4.16
MSI/HSI	5.19	5.34	4.83	5.34	5.77	5.19	5.21	4.65	3.8	3.9	3.43	3.71
Research	5.31	5.37	4.69	5.27	5.66	5.13	5.14	4.67	3.92	4.13	3.73	4.21
UG	5.45	5.61	4.84	5.41	5.75	5.39	5.3	5.03	4.01	4.54	3.75	4.6
Gender												
Female	5.31	5.45	4.72	5.27	5.66	5.2	5.16	4.82	3.9	4.27	3.63	4.29
Male	5.46	5.52	4.88	5.55	5.89	5.43	5.42	4.97	3.99	4.41	3.78	4.34
Nonbinary & Other	5.13	4.97	4.87	5.05	5.38	4.95	5.23	4.59	3.46	3.92	3.64	4.18
Race												
Asian	5.23	5.33	4.78	5.3	5.6	5.26	5.17	4.84	4.11	4.3	3.77	4.22
Black	5.37	5.28	4.6	5.5	5.84	5.08	5.33	4.75	3.36	4.12	3.61	3.97
Hispanic/Latinx	5.38	5.44	4.92	5.33	5.79	5.3	5.26	4.75	3.88	4.21	3.51	4.05
Other	5.43	5.55	4.78	5.38	5.77	5.32	5.28	4.93	3.95	4.37	3.72	4.45
White	5.21	5.47	4.92	5.44	5.73	5.31	5.37	4.86	3.88	4.37	3.55	4.33
School year												
First-year	5.12	5.25	4.51	5.13	5.69	5.11	5.11	4.64	3.44	4	3.43	3.99
Second year	5.34	5.45	4.81	5.38	5.68	5.33	5.29	4.94	4.11	4.46	3.68	4.47
Third year	5.37	5.58	4.91	5.38	5.74	5.29	5.28	4.9	3.87	4.33	3.67	4.23
Fourth year+	5.63	5.59	4.91	5.58	5.87	5.38	5.37	5	4.2	4.42	3.95	4.5

Appendix A cont.

Difficulty Values for all Items (Q13- Q21)

Table A

Difficulty values for all items by demographic group.

Demographics	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21
Total	5.35	5.37	5.54	5.36	5.4	5.83	5.3	5.38	5.98
Institution Type									
HBCU	5.2	5.21	5.38	5.23	5.24	5.91	5.33	5.38	5.95
MSI/HSI	5.21	5.26	5.39	4.99	5.11	5.85	5.14	5.31	5.8
Research	5.14	5.24	5.48	5.34	5.37	5.63	5.28	5.31	6.14
UG	5.51	5.5	5.65	5.54	5.55	5.9	5.37	5.43	5.98
Gender									
Female	5.33	5.39	5.51	5.32	5.34	5.74	5.19	5.31	5.85
Male	5.4	5.36	5.59	5.44	5.49	6	5.47	5.47	6.18
Nonbinary & Other	5.05	5.36	5.44	5.08	5.23	5.49	5.18	5.26	5.72
Race									
Asian	5.26	5.26	5.38	5.24	5.26	5.82	5.28	5.36	5.78
Black	5.26	5.15	5.36	5.15	5.19	5.98	5.24	5.36	5.92
Hispanic/Latinx	5.28	5.27	5.48	5.16	5.21	5.97	5.34	5.37	5.92
Other	5.4	5.46	5.62	5.49	5.52	5.8	5.29	5.35	6.04
White	5.45	5.43	5.65	5.32	5.4	5.72	5.43	5.64	6.17
School year									
First-year	5.07	5.16	5.37	5.03	5.29	5.38	5.04	5.15	5.65
Second year	5.37	5.42	5.49	5.32	5.43	5.82	5.28	5.3	5.92
Third year	5.38	5.39	5.53	5.44	5.42	5.96	5.39	5.49	6.1
Fourth year +	5.56	5.51	5.77	5.65	5.44	6.15	5.5	5.57	6.24

Appendix B

Discrimination Values for all Items (Q1 - Q13)

Table B

Corrected polyserial correlation coefficient between separate items and total PSO score as item discrimination for all demographics.

Demographics	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12
Total	0.66	0.56	0.66	0.68	0.59	0.75	0.71	0.67	0.6	0.72	0.56	0.72
Institution Type												
HBCU	0.65	0.44	0.67	0.53	0.51	0.71	0.67	0.66	0.51	0.71	0.52	0.59
MSI/HSI	0.66	0.58	0.72	0.74	0.52	0.8	0.77	0.64	0.61	0.71	0.59	0.71
Research	0.66	0.56	0.61	0.66	0.65	0.73	0.7	0.66	0.58	0.69	0.53	0.71
UG	0.67	0.6	0.72	0.73	0.59	0.76	0.72	0.7	0.67	0.76	0.6	0.77
Gender												
Female	0.63	0.56	0.67	0.67	0.62	0.76	0.71	0.68	0.61	0.72	0.57	0.74
Male	0.67	0.56	0.64	0.68	0.57	0.74	0.71	0.65	0.59	0.71	0.54	0.69
Nonbinary & Other	0.78	0.67	0.82	0.72	0.55	0.73	0.72	0.75	0.73	0.79	0.77	0.85
Race												
Asian	0.72	0.6	0.7	0.7	0.6	0.77	0.75	0.74	0.67	0.73	0.62	0.73
Black	0.57	0.48	0.72	0.58	0.53	0.73	0.67	0.64	0.5	0.67	0.55	0.66
Hispanic/Latinx	0.6	0.47	0.67	0.71	0.58	0.82	0.74	0.64	0.66	0.73	0.55	0.72
Other	0.68	0.74	0.72	0.76	0.61	0.82	0.76	0.69	0.64	0.8	0.63	0.76
White	0.67	0.56	0.63	0.67	0.6	0.72	0.69	0.64	0.57	0.7	0.53	0.71
School year												
First-year	0.65	0.61	0.68	0.69	0.61	0.76	0.77	0.65	0.61	0.68	0.69	0.61
Second year	0.69	0.56	0.66	0.68	0.61	0.75	0.72	0.69	0.56	0.66	0.68	0.61
Third year	0.68	0.55	0.67	0.75	0.56	0.78	0.74	0.68	0.55	0.67	0.75	0.56
Fourth year+	0.58	0.51	0.61	0.59	0.57	0.72	0.63	0.58	0.51	0.61	0.59	0.57

Appendix B cont.

Discrimination Values for all Items (Q13 -Q21)

Table B

Corrected polyserial correlation coefficient between separate items and total PSO score as item discrimination for all demographics.

Demographics	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21
Total	0.69	0.72	0.74	0.72	0.73	0.45	0.7	0.67	0.47
Institution Type									
HBCU	0.67	0.74	0.68	0.73	0.66	0.42	0.57	0.61	0.44
MSI/HSI	0.68	0.7	0.77	0.74	0.76	0.37	0.72	0.62	0.57
Research	0.68	0.72	0.73	0.7	0.72	0.47	0.72	0.69	0.43
UG	0.73	0.74	0.76	0.73	0.73	0.5	0.7	0.7	0.46
Gender									
Female	0.7	0.77	0.74	0.73	0.74	0.39	0.69	0.68	0.45
Male	0.68	0.68	0.74	0.7	0.71	0.48	0.7	0.65	0.47
Nonbinary & Other	0.79	0.84	0.82	0.8	0.81	0.46	0.76	0.81	0.64
Race									
Asian	0.71	0.72	0.74	0.78	0.73	0.52	0.74	0.69	0.54
Black	0.63	0.76	0.72	0.73	0.66	0.36	0.54	0.61	0.41
Hispanic/Latinx	0.67	0.71	0.8	0.76	0.81	0.33	0.75	0.63	0.61
Other	0.78	0.76	0.79	0.73	0.78	0.5	0.76	0.75	0.6
White	0.68	0.71	0.72	0.67	0.7	0.46	0.69	0.67	0.4
School year									
First-year	0.7	0.72	0.72	0.72	0.73	0.39	0.76	0.66	0.42
Second year	0.66	0.74	0.77	0.74	0.74	0.48	0.7	0.7	0.44
Third year	0.71	0.72	0.75	0.71	0.74	0.48	0.66	0.66	0.5
Fourth year+	0.68	0.7	0.73	0.68	0.73	0.37	0.67	0.65	0.48

Appendix C.

Professional Skills Opportunities Final Survey

Table C

PSO Factors and Associated Items

Question stem: *How often in your undergraduate engineering experiences did you:*	
Shared Leadership	
Q1	*support team members then they faced a challenge?*
Q2	*share the workload among the members throughout the project?*
Q3	*work to resolve conflicts within the team?*
Q4	*encourage others to focus on achieving goals?*
Q5	*accept the responsibility for your personal growth?*
Q6	*support others to development skills or improve performances?*
Q7	*motivate others to produce quality work?*
Business and Management Principles	
Q8	*plan the order of competing tasks based on stakeholder priorities?*
Q9	*manage available financial resources?*
Q10	*anticipate possible future stakeholder needs?*
Q11	*consider possible legal constraints?*
Q12	*evaluate whether different stakeholder needs are satisfied?*
Problem-Solving	
Q13	*analyze the constraints of potential solutions?*
Q14	*optimize your solutions?*
Q15	*identify a problem that needs to be solved related to a project?*
Q16	*evaluate the feasibility of ideas generated?*
Q17	*generate whether different stakeholder needs are satisfied?*
Communication	
Q18	*adapt to the mode of communication?*
Q19	*adjust the content of your communication based on audience?*
Q20	*changed the styles of your communications according to different situations?*
Q21	*use written formats of communication?*
Professional and Ethic Responsibilities	
Q22	*consider possible negative consequences of your design?*
Q23	*report undesirable results truthfully?*
Q24	*consider the impacts of your professional conducts?*
Q25	*reflect how your decisions can impact your organization’s reputation?*

Author Biographies

[anonymized for review]