Research Experiences Instrument: Validation Evidence for an Instrument to Assess the Research Experiences of Engineering Ph.D. Students' Professional Practice Opportunities

Eric A. Holloway[1], Kerrie A. Douglas[1], David F. Radcliffe[1], William C. Oakes[1]

[1]School of Engineering Education, Purdue University

Correspondence
Eric Holloway,
School of Engineering Education, Purdue University, West Lafayette,
516 Northwestern Avenue, West Lafayette, IN 47907, USA.
Email: eahollow@purdue.edu

Abstract

Background. There are long-held concerns about how graduate research programs prepare engineering Ph.D. students for professional practice. Suitable instruments are lacking to effectively assess how research experiences contribute to the success of graduate students becoming professionals.

Purpose. The purpose of this work is to examine evidence of internal reliability and validity of using the Research Experiences Instrument (REI) scores as a measure of engineering Ph.D. students' professional practice opportunities in their research experiences.

Method. REI was constructed using an ontological framework. REI was administered twice to engineering Ph.D. students, once to a single university (n = 236) and once to multi-universities (n = 215). Psychometric analyses were conducted related to validity and reliability evidence, including exploratory and confirmatory factor analyses, and group score comparisons between genders, race/ethnicity, and engineering disciplines.

Results. Results of both factor analyses aligned with the theoretical five-factor structure, with a second-order *Opportunity* factor. Mean scores were the same between women and men, and for three engineering disciplines, but significantly lower for racially/ethnically minoritized groups. Factor scores indicate that students often lack opportunities to engage with professionals, a likely cause of students' professional practice struggles.

Conclusions. Evidence of validity has been provided to justify the use of REI to assess the unique research experiences of engineering Ph.D. students as these relate to preparedness for professional practice. REI scores can be used for gender and race/ethnicity comparisons and are generalizable across engineering disciplines. Where students' REI scores show a lack of opportunities, remedial strategies can be implemented.

Keywords: professional practice, research experiences, assessment, engineering Ph.D. students

1. Introduction

Engineering Ph.D. students currently face many challenges upon graduation. These challenges include limited tenure-track positions in academia (Kindsiko & Baruch, 2019), economic and job uncertainty due to the COVID-19 pandemic (Godinic et al., 2020), and questions from employers about skills versus credentials (Brown & Souto-Otero, 2020). Upon entering the workforce, students are often surprised by the difficulty of applying their knowledge and skills obtained during their Ph.D. research to real-world problems (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018). Likewise, employers often perceive that recent engineering Ph.D. graduates are unprepared to practice professionally (Holloway & Radcliffe, 2018; Mosyjowski et al., 2017; NASEM, 2018).

Perceptions that engineering Ph.D. students are unprepared for professional practice are not new. For almost forty years, the National Academies have highlighted concerns about engineering Ph.D. students' preparation for professional practice (NASEM, 2018; National Academy of Engineering [NAE], 1983, 1995, 2017). The National Science Foundation (NSF) cited better engineering graduate student preparation as a motivating factor when creating Engineering Research Centers (ERCs) (NAE, 1983). Concerns about Ph.D. students' preparedness are also a global issue, as reports from Canada, Europe, and Australia (Borrell-Damian et al., 2015; Edge & Munro, 2015; McGagh et al., 2016) echo the National Academies. All of these reports have a common theme about engineering Ph.D. students' lack of preparation: students struggle to translate their knowledge into impact in practice.

For most engineering Ph.D. students, their formative professional practice preparation in the academy comes from their research experiences. Students' research experiences provide them with their identities as researchers as future professionals (Choe & Borrego, 2020).

Graduate student educational reform has also focused on students' research experiences to foster better preparation. For example, NSF ERCs strive "to improve engineering research so that U.S. engineers will be better prepared to contribute to engineering practice" (Parker, 1997, p. 3). ERCs typically offer different research experiences for students, requiring more applied research projects, more significant interaction with industry and government sponsors, and diverse student skill sets (Kannankutty et al., 1999). However, very little is known about how students' research experiences prepare them (Crede & Borrego, 2012).

Graduate engineering students' research experiences can vary considerably based on contexts, yet there is a need for assessments capable of generating insight into the professional preparation. One reason for this lack of insight on their preparation is the perception that students' research experiences are unique and diverse, or as Thune explained, a heterogeneous phenomenon (Thune, 2009, 2010). The few assessments of graduate students' research experiences that exist have focused on their knowledge, skills, and abilities (Holloway et al., 2020). NSF-funded assessments of ERCs experiences have shown significant impacts on the students' outcomes from improving industry's perceptions of their preparedness for professional practice (Parker, 1997; Roessner et al., 2004). While helpful, these studies lacked scales developed with theoretical frameworks for their constructs, nor reported evidence of reliability or validity (Parker, 1997; Roessner et al., 2004), making them difficult to reuse reliably. Other assessments tend to focus directly on graduate students' research skills, such as developing research proposals, and how those skills changed over time (Timmerman et al., 2013), rather than holistic professional practice preparation. Studies focused on graduate students self-assessing their own research skills have led to inflated scores (Gilmore & Feldon, 2010). No studies were found that assess Ph.D. students' professional practice preparation at the institutional level.

2. Research Purpose

The purpose of this research paper is to study aspects of validity to evaluate the use of the Research Experiences Instrument (REI), as a measure of students' reports of opportunities to practice being a professional in their research experience. A recent report on the state of professional development for STEM graduate students cites the "lack of evidence about what forms of professional development are most effective and most worthy of investment" (Denecke et al., 2017, p. 9). For engineering Ph.D. students, this arises because there are no instruments that measure characteristics related to how their research experiences are preparing them for professional practice. As such, there is a need to assess how students are being prepared for practice by their research experiences. Opportunities to practice being a professional are a significant contributor to Ph.D. students' being prepared for actual professional practice upon graduation (Dall'Alba, 2009; Dall'Alba & Sandberg, 2010).

The development of the REI followed recommendations for instrument development by Netemeyer and colleagues (2003), focused on construct design and initial evidence of validity. While there is no such thing as a perfectly valid assessment (Douglas & Purzer, 2015), and one manuscript could not do justice to the ongoing nature of increasing measurement precision, four phases of initial validation are presented, following the recommendations from Netemeyer and colleagues (2003), shown in Table 1 below. The goal is to establish a strong rationale based on evidence for using the REI as an assessment tool in engineering education research.

Table 1 Four phases of the study

Phase #	Study phase name	Netemeyer et al. (2003) 4-step development process
1	Research Experiences Instrument (REI) development	1) construct definition. 2) generating and expert review of items.
2	Pilot study of factor structure	3) designing and conducting studies to develop the instrument.
3	Confirmatory factor structure	3) designing and conducting studies to develop the instrument.
4	Group score comparison	4) finalizing the instrument.

At each phase of the study, validity evidence was collected and evaluated. Validity in instrument development is commonly understood as the degree to which the measure is actually measuring the construct it is intended to measure, supported by the evidence (Messick, 1995; Netemeyer et al., 2003). This study specifically utilized Messick's Unified Theory of Validity (Messick, 1995), in which all sources of validity evidence support aspects of construct validity in an accumulation of evidence and subsequent justification of the evidence. Messick (1995) identified six aspects of validity that support construct validity in a unified concept: content, substantive, structural, generalizability, external, and consequential. Others advocate this evidence accumulation and justification approach (Douglas et al., 2016; Kane, 1992). Messick's framework is aligned with the latest strategy taken by the joint committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, which controls the Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014). This strategy views validity not as a checklist of tasks to be completed or as a property of the instrument, but as a summary of the evidence and the justification of the uses and interpretations of the instrument. Or, as Messick summarized, "what is required is a compelling argument that the available evidence justifies the test interpretation and use" (Messick, 1995, p. 774).

This study was fundamentally an instrument development and evaluation. Following the 4-step process specified by Netemeyer et al. (2003), this process was iterative, where evidence was collected at each step, with item refinement, leading to the next step. Aligning with Messick's (1995) approach, the validity evidence was accumulated and justified at each step. As such, this strategy was utilized to determine how the REI is intended to be used and interpreted, and how to evaluate if those were appropriate. It was determined that if the REI measures students' opportunities to practice being a professional, then 1) items will be internally consistent (i.e., reliability); 2) the factor structure will be aligned to the theoretical design (i.e., structural aspect of validity); 3) scores between demographic groups (gender and racially/ethnically minoritized groups) will be aligned to literature (i.e., external aspect of validity); 4) scores between engineering disciplines indicate generalizability to other situations (i.e., generalizability aspect of validity).

The following research questions were asked in support of the determinations made of the REI: To what extent do the REI scores demonstrate evidence of reliability? To what extent do the REI scores support the theoretical factor structure? To what extent do the REI scores for gender and racially/ethnically minoritized groups align with literature? To what extent do the REI scores for engineering discipline indicate generalizability to other situations? What do the REI scores indicate about students' opportunities to practice being a professional in their research experiences?

2.1 Positionality Statement

Three of the four authors have spent many years in industry working as engineers. Our approach to this research was influenced by observing that engineers, including those with graduate degrees, frequently struggle to translate their technical knowledge and skills acquired in academia into the practice of engineering when they first enter industry. This research was also shaped by an awareness of the variety of experiences that Ph.D. students have in diverse laboratory settings across campus and between institutions. These research experiences include professional matters (e.g., being an effective team member) as well as technical matters. We reasoned that the term ‘research experiences’ needed to be as broad as possible, encompassing whatever technical and professional opportunities an individual student has access to as part of their graduate research endeavors.

3. Phase 1: Research Experiences Instrument Development

Phase 1 was focused on defining the REI constructs, developing initial items, and providing an indication of how well students understood the items. This goal was accomplished by utilizing a theoretical framework to situate students’ research experiences to define the constructs and develop initial items, and conducting a small study to evaluate the items by experts and students.

3.1 Construct Definition

The critical first step in construct definition is to arrive at a suitable theoretical framework from which to view the phenomena in question (Cronbach & Meehl, 1955). The search for a framework (Holloway et al., 2020) centered on frameworks that help explain why students find it difficult to make the transition from being a graduate student to being a professional, as the literature indicates this is a significant problem for Ph.D. students (NASEM, 2018; Ronfeldt & Grossman, 2008; Shulman, 1998; Taylor, 2007).

Three frameworks were considered. These are listed in Table 2 in terms of how each explains why students find it difficult to make the transition from graduate student to being a professional.

Table 2 Three frameworks considered for why students' find it difficult to become a professional

Framework	Approach to why Ph.D. students' find it difficult to become a professional	Framework applied to becoming a professional
Chi's coherence framework (Chi et al., 2012)	Some concepts are difficult for Ph.D. students by utilizing an approach that suggests there are missing ontological categories in students' knowledge, which are typically emergent. According to Chi, students are making a category mistake "when a concept has been assigned inappropriately to a lateral or alternative ontological category" (Chi, 2008, p. 65).	This framework does not include the ontological aspect of what it means to become a professional. In Chi's language, the process of becoming a professional is likely an emergent process, indicating it would be difficult for students. For Chi, the mistake students would be making is to assume the knowledge they are learning will directly apply in a professional setting. The students will need to make an ontological shift to apply their knowledge in a professional setting.
Säljö's sociocultural framework (Säljö, 1999)	Some concepts are difficult for Ph.D. students by utilizing an approach that suggests that the difficult concepts are discursive and not situated in a social or cultural meaning until students can provide meaning to the concepts.	This framework focuses on the limited social or cultural understanding that Ph.D. students have about being a professional. This context can be helpful to understand why students have difficulty situating the knowledge and skills they develop during their research experience into professional settings.
Dall'Alba's Ways of Being framework (Dall'Alba, 2009)	Becoming a professional is difficult for Ph.D. students by adding an ontological (being) aspect onto the typical epistemological (knowing) aspect in the process of becoming a professional. Adding the ontological aspect allows the focus to be on what it means to develop professional ways of being (ontological) and not just focus on the acquisition of knowledge and skills (epistemological).	This framework clearly delineates between the epistemological focus on Ph.D. students' training and experiences in academia and the overlooked ontological aspect of what it means to become a professional. Overlooking the ontological aspect makes it necessary for students to make the required adjustments to fully understand and experience what it means to become a professional.

Dall'Alba's Ways of Being framework (Dall'Alba, 2009) was chosen as the framework used for this research, as it is essentially a specific example of Chi's coherence framework and Säljö's sociocultural framework. While Chi and Säljö are somewhat generic frameworks, Dall'Alba's Ways of Being framework is very specific for the application of becoming a professional. In fact, Dall'Alba references Säljö (but not Chi) several times in her work. A fundamental way to compare the Chi and Säljö frameworks with that of Dall'Alba is that the Chi and Säljö frameworks are complementary to Dall'Alba, but Dall'Alba's Ways of Being framework provides the better overall depth and specificity for what it means to be and become a professional.

Whereas most studies in higher education about students' outcomes focus on their knowledge and skill acquisition (an epistemological approach), the Ways of Being approach (Dall'Alba & Sandberg, 2010, p. 106) integrates knowledge and skill acquisition into the processes of students being and becoming a professional. Students have difficulty translating their knowledge to impact in practice because they do not understand what it means to be a professional. Further, students' Ph.D. research experiences do not explicitly provide sufficient opportunities for them to practice being a professional and to reflect upon these. Dall'Alba's Ways of Being framework also explains what it means for students <u>be/become a professional</u>, as it "involves transformation of the self through embodying the routines and traditions of the profession in question" (Dall'Alba, 2009, p. 37) and <u>learning to be/become a professional</u> as it "involves not only what we know and can do, but also who we are (becoming). It involves integration of knowing, acting, and being in the form of professional ways of being that unfold over time" (Dall'Alba, 2009, p. 34). Dall'Alba warns that "practice-based approaches commonly adopt an epistemological focus that neglects the ontological dimension central to learning. In other words, these approaches emphasize the knowledge or activities that are learned, at the expense of attention to who learners are becoming and what this process of becoming involves" (Dall'Alba & Sandberg, 2010, p. 104).

Dall'Alba's framework was utilized as a lens to review the literature of engineering Ph.D. students' research experiences and identify six key characteristics that were essential to students' professional development in their research experience (Holloway et al., 2020). Each of the six key characteristics was evaluated per Dall'Alba's recommendations (Dall'Alba, 2009; Dall'Alba & Sandberg, 2010) for higher education to help students be and become professionals. This evaluation was performed to determine which of these characteristics had an ontological (being a professional) dimension to utilize for measurement. Table 3 below describes the six key characteristics, examples of each of the six, and the results of the ontological analysis using Dall'Alba's recommendations. The results of the ontological analysis led to the identification of five ontological dimensions that are likely present or absent to various amounts in students' research experiences, shown in Table 3 below.

Table 3 Key characteristics and ontological dimensions of students' research experiences

#	Key characteristics of students' research experiences	Examples of characteristics	Ontological dimensions
1	Size of the research group	Small (less than 5 students), medium (5 to 20 students), large (more than 20 students)	Not a direct ontological dimension; Influences how the research group is organized.
2	How the research group is organized	Work mostly individually, work mostly in a team	1. Opportunity for students to work as a team member on research.
3	Engineering disciplines	Discipline-specific (i.e., Electrical, Mechanical, Chemical, etc.)	Not an ontological dimension.
4	Types of research	Basic research, applied research, educational research	Not an ontological dimension.
5	Types of collaborators	Government, industry, research center	2. Opportunity for students to be exposed to their collaborator's form of practice. 3. Opportunity for students to have relevant exposure to professional practice based on later employment.
6	Types of equipment	Experimentation, modeling, limited equipment	4. Opportunity for students to gain experiences with modeling and simulation tasks. 5. Opportunity to gain experiences with hands-on and troubleshooting tasks.

Note. The literature review underpinning the six key characteristics of engineering Ph.D. students' research experiences is extensive and led to a novel way of conceptualizing the students' research experiences for their professional practice development (Holloway et al., 2020).

Each of these five ontological dimensions in Table 3 was operationalized for measurement by defining each from the literature. Table 4 below assigns a name to each dimension, describing the types of opportunities to practice being a professional that students might experience during their research degree. The principal sources from the literature that underpin each of these dimensions of the construct are also listed.

Table 4 Dimensions of the Research Experiences Instrument (REI) construct

Dimensions	Opportunities to practice being a professional in students' research experiences
Teamwork	Two or more engineering Ph.D. students who share a commitment to common research goals that are part of a larger research group in which members have differentiated skill sets, roles, and responsibilities in which they make decisions and coordinate tasks to accomplish research goals while exhibiting interdependencies with respect to workflow, goals, and outcomes (adapted from Fernandez et al. (2008)).
Collaboration	An engineering Ph.D. student exhibiting professional actions and behaviors with practicing engineers in the act of working in close collaboration together on research through regular interaction (adapted from Aldridge (1994)).
Networking	An engineering Ph.D. student exhibiting professional actions and behaviors with practicing engineers through their wider research experiences, such as conferences, internships, and professional societies where professional networks are developed (adapted from Aldridge (1994)).
Modeling	An engineering Ph.D. student demonstrating the process of identifying engineering problems, specifying constraints and assumptions in order to design and develop a mathematical model, often utilizing sophisticated engineering tools. The process continues with the verification and optimization of the model by evaluating the simulated performance of the system in an iterative process that utilizes refinement of the constraints, assumptions, and the model itself, facilitated by knowledge and discovery (adapted from Magana and Coutinho (2017); (Radcliffe, 2014)).
Experimentation	An engineering Ph.D. student performing engineering research tasks in which the student plans, uses, and/or deploys physical equipment or instrumentation, ensuring proper operation, collection, and interpretation of data, and troubleshooting and repair of the equipment or instrumentation to support the research endeavor (adapted from Lumpe and Oliver (1991); (Rivera-Reyes & Boyles)).

The overall REI construct is composed of these five dimensions with the aim of measuring the extent to which students' research experiences contributed to the opportunities to practice being a professional. When using these one-word descriptors, it is critical not to lose sight of the specifics and the subtle differences that lie behind each of these shorthand terms. For example, a distinction is drawn between what it means for *Teamwork* (i.e., students working with their fellow students as part of a team) and what it means for *Collaboration* (i.e., students working with professionals). While both may be part of a team, students' opportunities to practice being a professional are much different between these two examples.

3.2 Item Generation and Evaluation

Over an 18 month period, prospective items (i.e., survey questions) were generated, evaluated, and refined that assessed the REI measurement construct (Holloway et al., 2019). The item generation consisted of writing Likert-type questions for each of the five ontological dimensions on a frequency scale of 1 to 6, Never to Very Frequently, for each dimension of the REI. This type of scale is commonly used to measure the frequency of occurrence (Vagias, 2006). In addition, through the ontological framework, the frequency scale is how students' opportunities to practice being a professional are translated to empirical numbers (e.g., students 'occasionally' have the opportunity to practice *Teamwork*). In total, twenty-nine Likert-type items were developed for the REI. Next, to provide evidence of the content aspect of validity of each item, feedback was gathered from eleven faculty and graduate students with expertise in assessment development or in students' research experiences. These experts assessed the degree of alignment of each item with the definition of the relevant dimension, including providing written feedback of any concerns about the wording of items and/or ways an item could be improved. Items were modified based on misalignment with the definition. In some cases, revisions were made to the definitions based on the feedback.

Next, to ensure items were understood as intended and to provide evidence of the substantive aspect of validity, think-alouds were conducted with twelve engineering Ph.D. students that were diverse in terms of race/ethnicity, gender, and discipline (Holloway et al., 2019). A think-aloud process is a widely used technique in instrument development where participants "think-aloud," verbalizing their thinking as they completed the assessment (Blair et al., 2013). Student participants were provided a paper copy of the assessment, and allowed and encouraged to ask questions at the beginning of the interview. Students then read the items out-loud, including their responses and their thinking about the responses. The interviewer asked follow-up questions if a response was unclear. This process allowed for determining whether or not the students cognitively understood the items.

The think-aloud process also ensured that each student demonstrated that they understood what was encompassed by the term ‘research experiences.’ Rather than being prescriptive, students were encouraged to answer the survey questions in the context of their own experiences performing research, as it is clear from the literature that students’ research experiences vary greatly (Crede & Borrego, 2012; Thune, 2009, 2010).

The think-alouds were conducted in two rounds with students. In the first round with seven students, six items were modified due to cognitive issues in students’ understanding of the items, and twenty-two questions were modified to clarify a single word or phrase. In the second round with five students, all items were cognitively understood by students, and twelve questions were modified to clarify a single word or phrase. At the conclusion of this Phase 1 process, the REI had a total of twenty-nine prospective items. See Table 18 in Appendix C for a list of the REI items.

4. Phase 2: Pilot Study of Factor Structure

A pilot study of the REI was conducted to provide an initial indication of how well the items functioned relative to the overall research questions. Specifically, how scores initially demonstrated evidence of reliability and to what extent the REI scores formed a factor structure. This goal was accomplished using an exploratory factor analysis (EFA).

4.1 Methods

4.1.1 Setting and Participants

The 29-item REI was administered in the summer of 2019 to 1988 engineering Ph.D. students at a large, primary white institution (PWI) in the Midwest. Students completed the REI using an online survey tool by clicking a link in an email and agreeing to participate in the study.

Overall, 466 responses were received. Responses were removed based on the criteria of those who did not complete 100% of the survey, and those who did not answer a filter question correctly. No patterns were observed in those who did not provide complete data. This process led to the removal of 210 responses, leaving a total of 236 complete responses used for analysis, leaving a sample appropriate for EFA (Fabrigar et al., 1999; Netemeyer et al., 2003). The demographic distribution of participants in the pilot study is presented in Table 5.

Table 5 Demographic distribution of participants in the pilot study

Group	Sub-group	Total *n* (%)
Gender	Men	165 (69.9%)
	Non-binary	1 (0.4%)
	Women	70 (29.7%)
Race/ethnicity	Minoritized groups	29 (12.3%)
	Non-minoritized groups	207 (87.7%)

Note. *n* = 236. *Note on Race/Ethnicity*. See below for an explanation of racially/ethnically minoritized group compositions.
See Table 17 in Appendix B for a complete list and breakdown of all engineering disciplines.

Minoritized groups of students were comprised of those who identified as Black/African American students, Hispanic or Latino students, Native American students, Pacific Islander students, or a multiplicity of any of these, following NSF's designations of racial/ethnic minority student groups (NSF, 2019c). Non-minoritized groups of students were comprised of Asian students, White students, Other students, or a multiplicity of any of these, and those that preferred not to respond, again following NSF convention. Each of these groups was looked at individually, but in order to meet the journal length requirements, groups were aggregated.

As shown in Table 5, percentages of women (29.7%) and men (69.9%) in this sample were similar to the population at this institution (25.1% women and 74.9% men based on available institution data). The percentage of minoritized students (12.3%) and non-minoritized students (88.7%) in the sample were similar to the population at this institution (5.2% minoritized students and 94.8% non-minoritized students based on available institution data).

4.1.2 Data Preprocessing

The data was analyzed for appropriateness with EFA following common procedures (Spector, 1992). Means, standard deviations, and normality of scores were calculated. Inter-item correlation coefficients were calculated to check for consistency of the construct, and any low inter-item correlations of less than .30 were discarded. Finally, Cronbach's alpha, a common method used in psychometrics to provide a measure of reliability, was calculated for the overall scale and without each item. Three items did not satisfy the criteria to be included in the EFA and were discarded from further analysis.

4.1.3 Exploratory Factor Analysis Procedures

An EFA was conducted with a focus on providing initial evidence of the structural aspect of validity and the extent REI scores formed a factor structure. The purpose of EFA is to reduce the data to a summary understanding by identifying the latent constructs (i.e., previously mentioned construct dimensions that become factors when measured) that make up the larger measurement construct (Thompson, 2004). Guidelines for conducting an EFA were followed (Fabrigar et al., 1999; Thompson, 2004). The R statistical programming language v 3.6.1 (R Core Development Team, 2020) was utilized for the analysis.

First, a parallel analysis was performed to determine how many factors to extract. Next, the EFA was conducted with the number of factors to extract based on the parallel analysis results. The factor rotation was set to an oblique promax rotation, which allows factors to be correlated, and was appropriate for educational research where some correlation between factors was expected. A maximum likelihood (ML) estimation solution was utilized, as the data processing results indicated the scores were not normally distributed. Once factors were extracted, standardized pattern coefficients (sometimes called factor loadings) were examined, as these values indicate how strongly an item loads to a factor (Thompson, 2004). Only items with a standardized pattern coefficient greater than .40 were retained as recommended by Costello and Osborne (2005) and Floyd and Widaman (1995).

4.2 Results

4.2.1 Descriptive Statistics

Eight of the twenty-nine items had elevated means (> 4.2 on a 6-point scale), the skewness ranged from −1.08 to 0.64, and kurtosis ranged from −1.34 to 2.35, indicating the data are not normally distributed, but met the respective thresholds of 3.0 and 10.0 established by Kline (2015) for EFA. The inter-item correlation values for items 4, 14, and 17 were less than .30 and were discarded from the remaining analysis. The overall Cronbach's alpha for the remaining 26 items was calculated ($\alpha = .89$) and was greater than the recommended benchmark of .80 for new scale development (Clark & Watson, 1995).

4.2.2 Results of EFA

EFA was conducted with the remaining 26 items. The results of the parallel analysis indicated that five factors should be extracted. EFA was conducted with five factors extracted, with results shown in Table 6 below. Table 6 lists the standardized pattern coefficients and Cronbach's alpha for each factor. Pattern coefficients greater than .40 are bolded.

Table 6 Summary of exploratory factor analysis pattern coefficients and Cronbach's alpha

		Standardized pattern coefficient				
Factors	Item #	*Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*
Teamwork	1	**.60**	.12	.00	-.08	.12
(α = .87)	2	**.86**	-.08	.15	-.02	-.05
	3	**.83**	.03	-.01	-.01	-.02
	5	**.84**	-.04	-.05	.06	-.04
Collaboration	6	.03	**.79**	-.04	.10	-.01
(α = .88)	7	.08	**.97**	-.29	-.03	.02
	8	-.03	**.73**	.15	-.03	-.03
	9	-.14	**.65**	.13	.02	.05
	10	-.07	**.61**	.23	-.10	-.10
	11	.16	**.59**	.06	.04	.00
Networking	12	.01	.10	**.77**	-.14	.07
(α = .77)	13	.02	.11	**.47**	.03	.06
	15	-.07	.18	**.65**	.02	-.05
	16	.09	-.04	**.55**	.10	-.02
Modeling	18	.05	-.09	.07	**.76**	-.08
(α = .90)	19	.06	.00	.02	**.87**	-.07
	20	-.02	.16	-.11	**.77**	.05
	21	.03	-.04	.01	**.72**	.01
	22	-.05	-.01	.00	**.90**	-.05
	23	.02	.00	.04	**.67**	.12
Experimentation	24	.00	-.06	-.04	-.03	**.88**
(α = .95)	25	.02	.08	-.03	-.02	**.89**
	26	-.03	-.05	.04	-.01	**.92**
	27	-.02	.00	-.06	.02	**.96**
	28	.02	.03	.03	.09	**.79**
	29	-.02	-.05	.11	-.04	**.85**

Note. Items 4, 14, and 17 were removed in the previous analysis.

Standardized correlations between factors are shown in Table 7, indicating a strong correlation between the two factors related to students working with professionals (*Collaboration* and *Networking*). Table 8 shows the model fit statistics, indicating that the 5-factor solution fits the data well.

Table 7 Factor correlations (standardized)

Subscale	*Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*
Teamwork	1.00				
Collaboration	.41	1.00			
Networking	.32	.64	1.00		
Modeling	.03	.18	.31	1.00	
Experimentation	.43	.28	.21	-.02	1.00

Table 8 Exploratory factor analysis model fit summary

Model fit statistic	Value	Criteria
Tucker Lewis Index (TLI)	.944	Rule-of-thumb is that models with TLI > .95 fit the data well (Hu & Bentler, 1999)
Root mean square error of approximation (RMSEA)	. 057	Rule-of-thumb is that models with RMSEA ≤ .06 fit the data well (Hu & Bentler, 1999)

4.2.3 Discussion of EFA Results

The goal of the pilot study of the REI was to provide an initial indication of how scores initially demonstrated evidence of reliability and the extent to which the REI scores formed a factor structure. Based on the results of the EFA, three items were removed (items 4, 14, and 17) due to lower inter-item correlation. The remaining 26 items formed a factor structure across five factors that aligned with the theoretical factor structure; *Teamwork*, *Collaboration*, *Networking*, *Modeling*, and *Experimentation*. The remaining 26 items had high reliability (internal consistency) of $\alpha = 0.89$.

5. Phase 3: Confirmatory Factor Structure

Based on the pilot study results, a confirmatory factor structure study of the REI was conducted to support the overall goal of research questions, specifically: To what extent do the REI scores demonstrate evidence of reliability? To what extent do the REI scores support the theoretical factor structure? This goal was accomplished using a confirmatory factor analysis (CFA).

5.1 Methods

5.1.1 Setting and Participants

The 26-item REI was administered in the fall of 2019 to an unknown number of engineering Ph.D. students at doctoral-granting U.S. universities accessed through listservs at the American Society for Engineering Education (ASEE) and at U.S. universities. Students completed the REI using an online survey tool by clicking a link in an email and agreeing to participate in the study.

Overall, 439 responses were received. Responses were removed based on the criteria of those who did not complete 100% of the survey, those who did not answer a filter question correctly, and those who were not engineering Ph.D. students. No patterns were observed in those who did not provide complete data. This process led to the removal of 224 responses, leaving a total of 215 complete responses used for analysis, leaving a sample appropriate for CFA (Floyd & Widaman, 1995; Netemeyer et al., 2003). The demographic distribution of participants in the confirmatory factor structure study is presented in Table 9.

Table 9 Demographic distribution of participants in the confirmatory factor structure study

Group	Sub-group	Total *n* (%)
Gender	Men	129 (60.0%)
	Non-binary	1 (0.5%)
	Women	85 (39.5%)
Race/ethnicity	Minoritized groups	22 (10.2%)
	Non-minoritized groups	193 (89.8%)

Note. $n = 215$. See Table 5 for racially/ethnically minoritized group compositions.
30 different Universities reported, including none listed. See Table 17 in Appendix B for a complete list and breakdown of all engineering disciplines.

As shown in Table 9, percentages of women (39.5%) and men (60.0%) in this sample were similar to the overall population per ASEE 2018 data (26.3% women and 73.7% men) (Roy, 2018), and NSF 2016 data (24.6% women and 75.4% men) (NSF, 2019c). As shown in

Table 9, the percentage of minoritized students (10.2%) and non-minoritized students (89.8%) in the sample were similar to the overall population per ASEE 2018 data (4.9% minoritized students and 87.7% non-minoritized students) (Roy, 2018), and NSF 2016 data (6.5% minoritized students and 93.5% non-minoritized students) (NSF, 2019c).

5.1.2 Data Preprocessing

The data was analyzed for appropriateness with CFA following common procedures (Spector, 1992). Means, standard deviations, and normality of scores were calculated. Inter-item correlation coefficients were calculated to check for the consistency of the construct. Finally, Cronbach's alpha was calculated for the overall scale and without each item. All items satisfied these criteria and were included in the CFA.

5.1.3 Confirmatory Factor Analysis Procedures

A CFA was conducted with a focus on providing evidence of the structural aspect of validity and the extent REI scores support the theoretical factor structure. The purpose of CFA is to evaluate how well the data fit the theoretical factor structure using goodness-of-fit indices (Thompson, 2004). In conducting the CFA, the sequence of steps recommended by Brown (2015) was followed. These steps focused on 1) establish an "unstructured" reference model for comparison; 2) establish a first-order model that provided a good fit aligned to the theoretical model; 3) evaluate the correlations among the factors in the first-order model; 4) theoretically and empirically fit the second-order model. The MPlus statistical programming language v8.4 (Muthén & Muthén, 1998-2017) was utilized for the analysis.

In order to holistically evaluate the fit of the various models in the CFA across several different criteria, three categories of goodness-of-fit indices were utilized, per Brown (2015): 1) absolute, 2) comparative, and 3) parsimony. Brown (2015) also recommends using at least one index from each category when evaluating and reporting goodness-of-fit. Each of the three categories is briefly explained, along with the selected fit indices utilized in this study and "rules of thumb" for evaluation as recommended per Byrne (2011).

First, absolute fit indices assess the extent that the model being evaluated fits the sample data (Byrne, 2011). The first fit index used to evaluate the absolute fit was the commonly used chi-square statistic (χ^2), along with evaluating chi-square divided by the degrees of freedom (χ^2/df), which provides a better overall indication of fit (lower χ^2/df is better) due to χ^2 sensitivity to sample size (Byrne, 2011). The second fit index used to evaluate the absolute fit was the standardized root mean square residual fit index (SRMR). The SRMR indicates the difference between the measured and expected covariances, with good fitting models having a value of less than .05.

Second, comparative fit indices assess the extent that one model compares to another model, often a baseline model referred to as the "null" or "unstructured reference" model where there is no structure at all between the inputs (Brown, 2015). The first fit index used to evaluate the comparative fit was the comparative fit index (CFI). The CFI indicates a normalized measure of the proportion of incremental improvement of the theoretical model fit to the "unstructured reference" model, ranging from 0 to 1, where a value of above .95 is considered an excellent fit (Byrne, 2011). The second fit index used to evaluate the comparative fit was the Tucker-Lewis fit index (TLI). As with the CFI, the TLI indicates a measure of the proportion of incremental improvement of the fit of the theoretical model to the "unstructured reference" model, but the TLI is non-normalized, and the range can extend above 1 (Byrne, 2011). As with the CFI, TLI values above .95 are considered an excellent fit (Byrne, 2011).

Third, parsimony fit indices assess the extent of the model fit to the data as well as the extent of the complexity of the model and penalize overly complex models (Byrne, 2011). The first fit index used to evaluate the parsimony fit was the root mean square error of approximation (RMSEA) fit index. The RMSEA evaluates model fit and complexity by evaluating the discrepancy between how well the model would fit the population covariance matrix, and is sensitive to model complexity because it considers the degree of freedom (Byrne, 2011). As recommended per Byrne (2011), an RMSEA value less than .05 indicates a good fit, between .05 to .08 indicates a reasonable fit, and above .08 indicates a poor fit. The second fit index used to evaluate the parsimony fit was the Bayesian information criterion (BIC). The BIC provides an indication of model parsimony by calculating an index based on the model fit and degrees of freedom (Byrne, 2011). This index then allows for comparison between models as to which model is more parsimonious, as indicated by the lower BIC value (Byrne, 2011).

5.2 Results

5.2.1 Descriptive Statistics

Eight of the 26 items had elevated means (>4.2 on a 6-point scale), the skewness ranged from −1.32 to 0.62, and kurtosis ranged from −1.27 to 0.54, indicating the data were not normally distributed, but met the respective thresholds of 3.0 and 10.0 established by Kline (2015) for CFA. The inter-item correlation values for all items were less above .30. The overall Cronbach's alpha was calculated ($\alpha = 0.88$), greater than the recommended benchmark of .80 for new scale development (Clark & Watson, 1995). Based on these criteria, all 26 items were retained for the CFA.

5.2.2 Results of CFA

A CFA was conducted with the 26 items. The first model generated (Table 10 – Null) was the “unstructured reference” for comparison. The second model generated (Table 10 – 1st-Order Theory) loaded items onto the five factors only with no higher-order factor. The model produced standardized pattern coefficients for all items in acceptable ranges (.53 to .93) and acceptable fit indices (CFI of 0.94 and a TLI of 0.93). As recommended by Thompson (2004), alternate models were tried in order to improve fit. Therefore, the third model generated (Table 10 – 1st-Order Alternate) was a suggestion from MPlus tool output to improve the fit of the second model by allowing the residuals for items 22 and 21 to co-vary. This suggestion (items 22 and 21 residuals to co-vary) altered the standardized pattern coefficients compared to the second model by at most .01 across the range, and only marginally improved the fit (see Table 10) and was discounted due to parsimony. The fourth and final model generated (Table 10 – 2nd-Order Theory) loaded items onto the five factors with a 2nd-order factor. This model utilized the common practice of setting the second-order factor variance to 1.0 for standardization (Brown, 2015; Byrne, 2011). This solution produced standardized pattern coefficients for all items in the range of .59 to .96 and acceptable model fit compared to the 1st-Order Theory model.

Table 10 Comparison of confirmatory factor analysis fit indices

Model	X^2	*df*	X^2/df	SRMR	CFI	TLI	RMSEA	BIC
Null Unstructured reference	3,782.2*	325	11.64	.29	**	**	.22	21,561
1st-Order Theory 5 theoretical factors, no higher order	502.7*	289	1.74	.06	.94	.93	.06	18.136
1st-Order Alternate Alternate 1st-Order solution allowing I22-21 residuals co-vary	474.4*	288	1.65	.06	.95	.94	.05	18,110
2nd-Order Theory All 5 factors load on a higher order '*Opportunities*' factor	528.0*	295	1.79	.08	.93	.93	.06	18,128

Note. *p < .001 (** not meaningful for the reference model)
SRMR = Standardized root mean square residual; CFI = Comparative fit index;
TLI = Tucker-Lewis fit index, RMSEA = Root mean square error of approximation;
BIC = Bayesian information criterion.

Standardized correlations between factors for the 1st-Order Theory model are shown in Table 11. These values are very similar to the factor correlations in the EFA shown in Table 7, again indicating a strong correlation between the two factors related to students working with professionals (*Collaboration* and *Networking*). As the REI is theorized to be a 2nd-order factor with subcomponents of *Teamwork*, *Collaboration, Networking*, *Modeling*, and *Experimentation*, the 2nd-Order Theory model was selected based on the overall evaluation of model fit, theory alignment, and parsimony.

Table 11 1st-Order theory factor correlations (standardized)

Factors	*Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*
Teamwork	1.00				
Collaboration	.21	1.00			
Networking	.22	.82	1.00		
Modeling	.01	.20	.38	1.00	
Experimentation	.28	.17	.20	-.07	1.00

Figure 1 shows the final CFA 2nd-Order Theory model, along with standardized pattern coefficients. Table 12 lists the standardized pattern coefficient for each factor for the REI and the range of standardized pattern coefficients for each factor. The factors *Teamwork, Collaboration, Networking, Modeling*, and *Experimentation* adequately load onto a second-order factor of *Opportunity*, as shown in Figure 1 and Table 12, as coefficients range from .28 to .97. The pattern coefficients for the items ranged from .59 to .96, indicating all items load on the respective theoretical factors and contribute to the measurement of students' opportunities to practice being a professional.

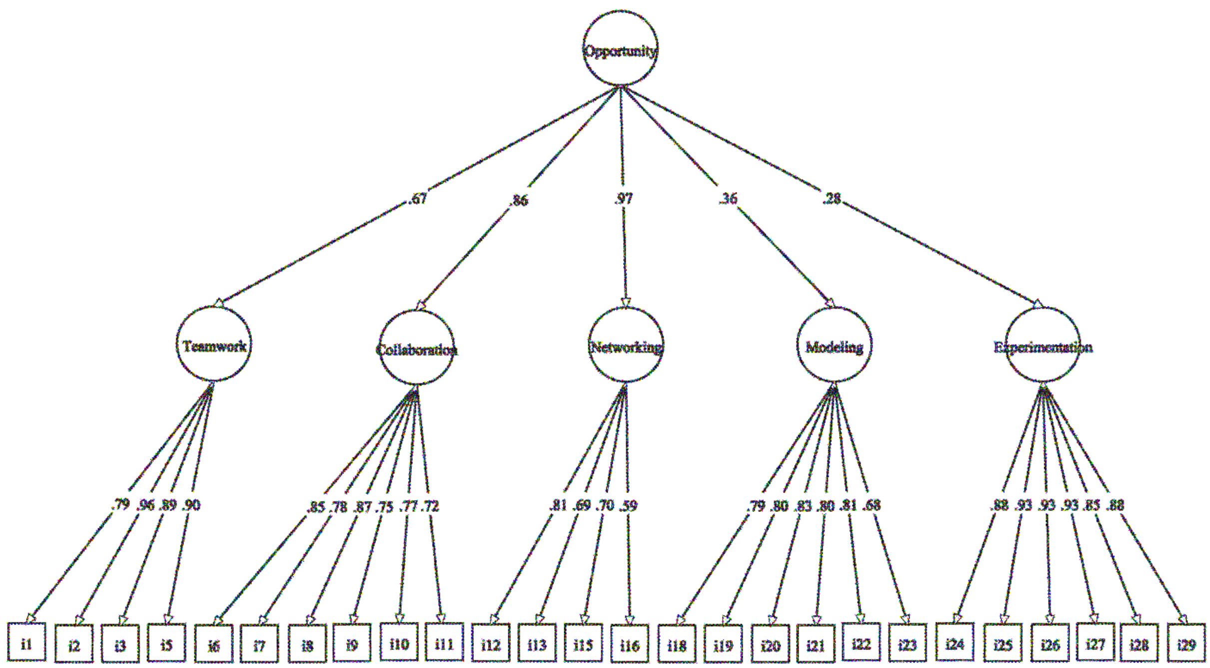

Figure 1 2nd-Order theory confirmatory factor analysis model with standardized loadings.

Table 12 Summary of standardized (std.) factor loadings in the 2nd-order theory model

Factors	# of items	Factor std. pattern coefficient	Range of std. item pattern coefficients
Teamwork	4	.67	.79 - .96
Collaboration	6	.86	.72 - .87
Networking	4	.97	.59 - .81
Modeling	6	.36	.68 - .83
Experimentation	6	.28	.86 - .93

5.2.3 Discussion of CFA Results

The goal of examining the factor structure of the REI was to determine how well items written to comprise a factor actually shared common variance and thus empirically factored together, and to assess the extent to which the factor structure overall was aligned to the theoretical model. Based on the results of the CFA, all 26 items formed a first-order factor structure across five factors that aligned with the theoretical factor structure with acceptable model fit; *Teamwork*, *Collaboration*, *Networking*, *Modeling*, and *Experimentation*. In addition, the CFA results established a second-order factor, *Opportunity*, that aligned with the theoretical factor structure with acceptable model fit. The 26 items had high reliability (internal consistency) of $\alpha = 0.88$.

6. Phase 4: Group Score Comparison

A study of the REI was conducted to assess how REI scores differed among the demographic groups of gender, racially/ethnically minoritized groups, engineering disciplines, and overall scores. This study of the REI was conducted in support of the overall goal of research questions: To what extent do the REI scores for gender and racially/ethnically minoritized groups align with literature? To what extent do the REI scores for engineering discipline indicate generalizability to other situations? What do the REI scores indicate about students' opportunities to practice being a professional in their research experiences? This goal was accomplished by comparing mean REI scores for the entire sample and for demographic groups.

6.1 Methods

6.1.1 Setting and Participants

The data from the EFA (n = 236) and the CFA (n = 215) were combined into one large sample (n = 451), to be simultaneously analyzed specifically for the purpose of examining group differences (Curran & Hussong, 2009). Combining multiple data sets into one data set and analyzing that combined data set (referred to as integrative data analysis, or IDA) is a common practice and has many advantages when done correctly and for the right reasons (Curran & Hussong, 2009). Curran and Hussong (2009) detail the specific advantages of IDA, and relative to this study, these included: 1) the potential to increase the statistical power, as sample size is increased; 2) the potential to increase sample heterogeneity, as "underrepresentation of potentially important subgroups in the population of interest" (2009, p. 86) may appear in the combined sample; 3) the potential to increase the estimation of reported base-rate behaviors; 4) broader psychometric assessment of the constructs, which in this study is across both the EFA and CFA constructs, and provides evidence to determine the REI scoring system. The demographic distribution of the combined participants is presented in Table 13.

Table 13 Demographic distribution of combined participants

Group	Sub-group	Total *n* (%)
Gender	Men	294 (65.2%)
	Non-binary	2 (0.4%)
	Women	155 (34.4%)
Race/ethnicity	Minoritized groups	51 (11.3%)
	Non-minoritized groups	400 (88.7%)

Note. *n* = 451. See Table 5 for racially/ethnically minoritized group compositions. 30 different Universities reported, including none listed. See Table 17 in Appendix B for a complete list and breakdown of all engineering disciplines.

6.1.2 Data Preprocessing

Means, standard deviations, and normality of the combined REI scores were calculated. Next, a simple factor scoring system (Holloway et al., 2021) was developed where a score was assigned to each factor based on the factor mean response. In other words, item responses were averaged according to each factor so that factor scores could be compared. Accordingly, the individual factor scores (*Teamwork*…*Experimentation*) remained on the original scale (Never to

Very Frequently) and are interpreted as such. The overall REI score, *Opportunity*, is the sum of the five factor scores, and indicates students' overall opportunities to practice being a professional (higher scores indicate more opportunities). *Opportunity* is on a 30-point scale: 5-7) never, 8-12) very rarely, 13-17) rarely, 18-22) occasionally, 23-27) frequently, 28-30) very frequently. Next, means, standard deviations, and normality of the REI factor scores were calculated for the entire sample and relative sub-groups.

6.1.3 Group Analysis Procedures

For both the racial/ethnic subgroup and the engineering discipline subgroup, group mean scores in the REI *Opportunity* score were compared based on the results of the individual factor scores. The nonparametric method Mann-Whitney U test for significance (for racial/ethnic subgroup, based on two groups compared) and the Kruskal-Wallis H test for significance (for engineering disciplines, based on three groups compared) was utilized based on the non-normality of the data. Both of these nonparametric tests are ranked-based nonparametric tests. The tests convert the continuous dependent variable (in this case, the REI *Opportunity* score) to ranks, to compare mean ranks between groups (Upton & Cook, 2014). Both tests have as the null hypothesis that there are no differences in the distributions, so if the null was rejected, it indicated there were differences in the distributions of scores (Upton & Cook, 2014). Equal group sizes were created for the racial/ethnic subgroup by randomly sampling the same number of non-minoritized groups' responses (n = 400) to match the total number of minoritized groups' (n = 51) to not bias towards the larger sized group. For the engineering discipline subgroup, the three engineering disciplines compared were chosen based on a statistical power analysis as recommended per Cohen (1988) that indicated a minimum group size of ~50/group. Equal group sizes were created by randomly sampling the same number of Electrical & Computer. Engr.

students (*n* = 82) and Mechanical Engr. students (*n* = 87) to match the total number of Chemical Engr. (*n* = 50) to not bias towards the larger sized group. Both tests were conducted in SPSS version 26 (IBM Corp., 2019), with the significance level set at .05 and the confidence level set at 95%.

6.2 Results

6.2.1 Descriptive Statistics

For the combined data, six of the 26 items had elevated means (> 4.2 on a 6-point scale), the skewness ranged from -1.12 to 0.63, and kurtosis ranged from -1.22 to 0.83, indicating the data were not normally distributed. Means and standard deviations of factor scores for the entire sample are presented in Table 14, along with the interpretation of factor scores. The individual factor scores (*Teamwork…Experimentation*) remained non-normally distributed, as the skewness ranged from -0.75 to -0.02, and kurtosis ranged from -0.87 to -0.13. The overall factor score, *Opportunity*, approached a normal distribution, as the skewness was -0.24, and kurtosis was -0.20. Gender, race/ethnicity, and select engineering discipline mean factor scores are shown in Table 15.

Table 14 Mean factor scores for the entire sample

Factor	*M* (SD)	Score interpretation
Teamwork	4.21 (1.12)	Occasionally
Collaboration	3.63 (1.29)	Between Rarely and Occasionally
Networking	3.16 (1.17)	Rarely
Modeling	4.23 (1.31)	Occasionally
Experimentation	4.15 (1.66)	Occasionally
Opportunity	19.38 (4.05)	Occasionally

Note. Score interpretation refers to how the mean score listed would be interpreted relative to the scale.

Table 15 Mean factor scores for gender, race/ethnicity, and engineering discipline subgroups

		Mean Scores					
Group	Subgroup	*Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*	*Opportunity* *M* (SD)
All	--	4.21	3.63	3.16	4.23	4.15	19.38 (4.05)
Gender	Men	4.22	3.55	3.12	4.34	4.16	19.39 (4.05)
	Women	4.20	3.79	3.23	4.01	4.13	19.38 (4.17)
Race / Ethnicity	Minoritized groups	4.27	3.49	3.06	3.85	3.38	18.06 (4.05)
	Non-minoritized Groups	4.21	3.65	3.17	4.28	4.24	19.55 (3.96)
Engr. Disciplines	Chemical Engr.	4.82	3.43	3.07	4.07	4.72	20.10 (3.35)
	Electrical & Computer Engr.	4.07	3.67	3.36	4.82	4.01	19.93 (3.94)
	Mechanical Engr.	4.02	3.27	2.86	4.47	4.36	18.98 (4.41)

Note. See Table 16 in Appendix A for all scores by race/ethnicity. See Table 17 in Appendix B for all scores by engineering (Engr.) discipline.

6.2.2 Results of Group Analysis

The Mann-Whitney U test results indicated significant differences in the REI *Opportunity* score between racially/ethnically minoritized groups of students and non-minoritized groups of students. Distributions of the REI *Opportunity* scores for minoritized students and non-minoritized students were not similar. The REI *Opportunity* scores for non-minoritized students (mean rank = 57.3) were statistically significantly higher than for minoritized students (mean rank = 45.7), U = 1003.5, z = - 1.988, p = .047. The Kruskal-Wallis H test results indicated no significant differences in the REI *Opportunity* score between the engineering disciplines, as distributions of the REI *Opportunity* scores were similar, $\chi^2(2) = 2.856$, $p = .240$.

6.2.3 Discussion of REI Scores Results

As shown in Table 14, the mean factor scores provide the first indication about students' opportunities, or lack thereof, to practice being a professional in their research experiences. Specifically, the two lowest mean factor scores were *Collaboration* (scored between 'Rarely' and ‘Occasionally’) and *Networking* (scored ‘Rarely’), both related to students having opportunities to work with professionals in their research experiences. In addition, the lowest mean factor score (*Networking*) was approximately one standard deviation lower than scores for *Teamwork*, *Modeling*, and *Experimentation*. This result showed a potential weakness in opportunities for students to engage with working professionals in their research experiences.

Overall *Opportunity* scores and mean factor scores indicate no differences, on average, between women and men students in REI scores. As shown in Table 13, percentages of women (34.2%) and men (65.2%) in the sample were similar to the overall population per ASEE 2018 data (26.3% women and 73.7% men) (Roy, 2018), and NSF 2016 data (24.6% women and 75.4% men) (NSF, 2019c). As shown in Table 15, the mean factor scores between women and men are virtually identical. The overall factor score, *Opportunity*, for women (*M* = 19.38; *SD* = 4.17) and men (*M* = 19.39; *SD* = 3.99) had the same mean. Women had a slightly larger standard deviation, and both women and men matched the overall sample mean (*M* = 19.38). These results indicated there were no differences, on average, between women and men in the opportunities to practice being a professional in their research experiences. These results align with previous research. Borrego et al. (2017) surveyed engineering master’s and Ph.D. students and found no significant gender differences in students’ satisfaction in their research experiences between women and men. Bahnson et al. (2021) studied opportunity structures in engineering Ph.D. students’ experiences and conducted a survey with 1754 participants and found no significant gender differences in students’ identity as a researcher between women and men.

Overall *Opportunity* scores and mean factor scores indicate significant differences, on average, between racially/ethnically minoritized groups of students and non-minoritized groups in REI scores. As shown in Table 13, the percentage of minoritized students (11.3%) and non-minoritized students (88.7%) in the sample were similar to the overall population per ASEE 2018 data (4.9% minoritized students and 95.1% non-minoritized students) (Roy, 2018), and NSF 2016 data (6.5% minoritized students and 93.5% non-minoritized students) (NSF, 2019c). However, the sample in this study also potentially included non-domestic students who identified as minoritized students, whereas ASEE/NSF data only considers domestic students in their data. As shown in Table 15, the overall factor score, *Opportunity*, and standard deviation of the minoritized students (*M* = 18.06; *SD* = 4.50) and non-minoritized students (*M* = 19.55; *SD* = 3.96) were significantly different, as the results of Mann-Whitney U test indicated. The interpretation of the differences in mean scores between minoritized students and non-minoritized students is that, on average, minoritized students had fewer opportunities to practice being a professional. Typically, "less occasionally" than non-minoritized students (minoritized students mean *Opportunity* scores are only 1 point from being in the "rarely" category). Furthermore, the REI indicates which opportunities are less for minoritized students, as their mean factor scores for *Modeling* and *Experimentation* were lower for minoritized students than for non-minoritized students. These results aligned with previous research. The report *Doctoral Initiative on Minority Attrition and Completion* (Sowell et al., 2015) described surveys of racially/ethnically minoritized Ph.D. students, which indicated those students perceived they had fewer opportunities in their Ph.D. experiences. Specifically, the report recommended that "many respondents recommended that institutions and departments be more proactive in providing research and professional development opportunities" (Sowell et al., 2015, p. 48) for minoritized Ph.D. students. As previously mentioned, Bahnson et al. (2021) studied opportunity structures in engineering Ph.D. students' experiences and conducted a survey with 1754 participants, and found significant race/ethnicity differences in students' identity as an engineer between minoritized students and non-minoritized students.

The results of gender and race/ethnicity scores provide evidence of the external aspect of validity, related to the extent that the REI items scores were both similar and different for external groups (i.e., converge or diverge with other variables). Evidence that the REI scores converged, or were similar, was provided by a gender group comparison of women and men. The results indicated that, on average, the REI measured women and men as having the same opportunities in their research experiences. Evidence that the REI scores diverged, or were different, was provided by a race/ethnicity group comparison of minoritized students and non-minoritized students. The result indicated statistically significant differences between minoritized students and non-minoritized students, and on average, the REI measured minoritized students as having fewer opportunities in their research experiences. These results, and the prior research from Bahnson et al. (2021), provide insight on why differences are sometimes measured between some groups and not in others: it depends on what construct is being measured.

Engineering discipline scores for the three engineering disciplines compared indicate no statistically significant differences in REI *Opportunity* scores. While subtle differences can be seen in the factor scores (Chemical Engr. students had higher *Teamwork* and *Experimentation* scores, and Electrical & Computer Engr. had higher *Modeling* scores), all the engineering disciplines followed the general trend of lower scores for *Collaboration* and *Networking*. These results provide evidence of the generalizability aspect of validity in terms of using the REI across engineering disciplines.

7. Discussion

The primary goal of this study was to develop and examine validity evidence to support the use of the Research Experiences Instrument (REI) to assess engineering Ph.D. students' opportunities to practice being a professional. The REI was based on an ontological framework that focused on potential aspects present or absent to various amounts in students' research experiences important to professional practice development (Dall'Alba, 2009; Dall'Alba & Sandberg, 2010). Based on this ontological framework, students' research experiences were conceptualized in a novel way for their professional development, and five ontological dimensions were identified that could be present or absent that were important to their professional development (Holloway et al., 2020): *Teamwork, Collaboration, Networking, Modeling,* and *Experimentation*. Validity evidence was gathered throughout the development process, from construct definition to diverse students' think-alouds, an evaluation that the factor structure was aligned with the theoretical definition, and comparison of group scores to the existing literature on professional opportunities for engineering Ph.D. students (e.g., Sowell et al. (2015)).

The REI was developed over four sequential phases, each one used to collect and evaluate specific evidence of validity that REI scores could be used to measure engineering Ph.D. students' opportunities to practice being a professional in their research experiences. In Phase 1, operational definitions and items were developed for each factor, resulting in twenty-nine initial items on a 6-point scale. The items went through iterations with experts and think-alouds with engineering Ph.D. students to understand their cognitive thought processes (Holloway et al., 2019). In Phase 2, the twenty-nine items were utilized in an initial, single-university pilot study focused on Exploratory Factor Analysis (EFA). The EFA resulted in the elimination of three items due to poor performance, and the remaining twenty-six assessment items supported the theoretical five-factor REI structure.

In Phase 3, the twenty-six items were utilized in a second multi-university study to confirm the factor structure was stable and supportive of the construct definition. The evidence supported a second-order factor, *Opportunities,* along with the five distinct first-order factors to be aligned to the theoretical REI structure. This evidence supports the scoring of each separate factor and an overall score.

In Phase 4, the evaluation of REI scores across gender, race/ethnicity, and engineering discipline groups was concurrent to the existing literature related to the opportunities that racially/ethnically minoritized students obtain in their Ph.D. programs (Sowell et al., 2015). The REI structure was also examined with a sample from multiple universities and fields of engineering. From this and the think-alouds with diverse students, we conclude that REI scores can be used across engineering disciplines, racial/ethnic, and gender groups.

This four-phase process established the validity of the REI to assess engineering Ph.D. students' opportunities to practice being a professional. The evidence also indicates that proper inferences can be made, for example, that higher REI scores indicate more opportunities to practice being a professional while lower scores indicate fewer opportunities to practice being a professional. Specifically, the evidence has shown the overall REI scores (*Opportunity*) indicate the overall opportunities to practice being a professional, while the individual factor scores (*Teamwork*, *Collaboration, Networking, Modeling,* and *Experimentation*) provide an indication of where the total REI scores for peer groups are higher or lower, comparatively.

The validity evidence established through the four-phase process of the REI development also allows for initial consideration of what Messick refers to as the consequential aspect of validity (Messick, 1995). This aspect refers to what outcomes occur when an assessment is used as the developers intended and whether those outcomes are aligned to the original purpose. By considering the consequential aspect of validity, the assessment developer acknowledges the real-world and ethical implications of using the assessment. While a full study of consequential aspects of validity is beyond the scope of the current research (and would require the use of REI outside of the development and initial evidence of validity), we can consider the potential consequences based on the work presented here. First, the purpose, intended use, and inferences to be made with the REI have been made explicit from the beginning of the design, and used consistently throughout the development and implementation process. The intention is that the use of REI would identify aspects of professional preparation that students have received and areas that could be bolstered. The consequence of that use would be more opportunities for professional preparation. Second, the REI scoring guide (Holloway et al., 2021) provides the user the means to practically interpret the resulting score from REI. Low scores on the REI do not indicate anything is wrong with students' research experiences. Rather, low scores indicate the possibilities for improving students' opportunities to practice being a professional, and the REI provides a diagnostic for identifying which opportunities graduate programs can focus on when creating opportunities to better prepare students for professional practice.

When engineering Ph.D. students were assessed about their research experiences with the REI's ontological framework, the results showed that students had fewer opportunities to work with professionals in their research experiences. Literature supports overall low interactions for Ph.D. students to work with professionals in their research experiences (Bruneel et al., 2010; Gemme & Gingras, 2012). Dall'Alba refers to students' working with professionals as their "forms of practice" (Dall'Alba & Sandberg, 2010, p. 105). In other words, the opportunity for students to work with professionals in their research experiences exposes students to certain forms of professional practice that situate students' knowledge and skills and model professional practice behaviors for them. In Dall'Alba's words, students get the opportunity to "take over others' ways of being" as professionals (Dall'Alba & Sandberg, 2010, p. 113). The result that students had fewer opportunities to work with professionals in their research experiences suggests a possible and likely reason why engineering Ph.D. students struggle to apply their knowledge and skills in the workplace. The evidence suggests engineering Ph.D. students are not getting enough opportunities to work with professionals in their research experiences.

Unlike an instrument constructed based on an epistemological framework (i.e., knowledge and skills), the REI does not seek to measure the quantity or quality of a student's research experience. This question would be analogous to asking if a basic research experience is better for students than an applied research experience. The ontological framework of the REI asks instead: What opportunities to practice being a professional might be missing to various degrees in a basic research experience or an applied research experience that faculty or administrators might want to supplement (or change)?

For example, extra efforts might be made to increase the professional experiences of students engaged in basic research, as they tend to be (a) more focused on theory and analysis (Morgan et al., 1996), (b) more likely to be funded by the government with infrequent and limited interactions with engineering practice (Gemme & Gingras, 2012), (c) and more likely to be in a workspace with less equipment where a student works largely by themselves (Crede & Borrego, 2012). The REI would likely produce a mixture of scores for students engaged in basic research. When compared to mean students' scores, their *Teamwork* score would be lower because students work mostly alone, as would their *Collaboration* and *Networking* scores because professional interactions as a collaborator are much less frequent. Their *Experimentation* score would be lower because they are less likely to be working with equipment, and their overall *Opportunity* score would be significantly lower. However, their *Modeling* score would be higher because their focus is on theory and analysis. Overall, these results suggest they should work with their advisor and committee to develop additional opportunities to practice and thereby develop various professional abilities that they are missing in their research experiences.

As an instrument with evidence of validity for its intended use and purpose, the REI has the potential to have a significant impact on the vast majority of engineering Ph.D. students by assisting them in identifying and developing suitable opportunities to better prepare for professional practice. Engineering Research Centers (ERCs) endeavor to provide opportunities in students' research experiences beyond traditional basic research experiences, requiring more applied research projects, more interaction with industry and government sponsors, an emphasis on teamwork, and student skill sets focused on modeling and experimentation (Kannankutty et al., 1999; Morgan et al., 1996; NAE, 2017). However, very few engineering Ph.D. students have access to the ERC experience. For example, from 2013-2017, ERCs averaged 258 graduate student degrees awarded per year (103 master's, 155 Ph.D.) (NSF, 2019b). In that same time frame, approximately 60,000 graduate student degrees were awarded per year in engineering in the U.S. (~50,000 master's, ~10,000 Ph.D.) (NSF, 2019a). Now, with the REI, students, faculty, and administrators can identify where professional opportunities may be falling short, and supplement those with other experiences that they might not be able to get directly in their research experiences.

8. Limitations

This study excludes engineering master's students in the population that was examined. However, the new way of conceptualizing students' research experiences for their professional practice development (Holloway et al., 2020) was developed for all graduate engineering students, not only engineering Ph.D. students. The REI assessment questions were developed for all graduate engineering students as well. In order to use the REI with master's engineering students, the REI stem would need to be changed, and the assessment would need additional evidence of validity for the use case with engineering master's students.

In addition, the targeted recruitment of specific racially/ethnically minoritized groups of students was not feasible in this study. However, the initial study of student think-alouds included students that were diverse in terms of race/ethnicity, gender, and discipline. The two quantitative studies' data collection, as stated, was focused on sample sizes large enough for EFA and CFA analyses, which also allowed for the comparison of some groups of interest. While the data allowed for combining all minoritized students into one large group for comparison, future studies with targeted recruitment and large sample sizes can be done to compare each demographic group individually. There is a need to understand professional practice development for all graduate engineering students.

Finally, this study did not ask for students' domestic/international status. Therefore, there may be international students who self-identified as minoritized students and were thus included in the race/ethnicity group comparison. What is clear from the race/ethnicity scores in Table 16 in Appendix A is that all scores follow the same general trend for each group (generally higher *Teamwork*, *Modeling*, and *Experimentation* scores, and lower *Collaboration* and *Networking* scores), with only subtle differences in magnitudes. Any international students that were included in the race/ethnicity group comparison would not change these trends, only the magnitudes, and it is clear from previous research that minoritized groups perceive to have fewer opportunities in their Ph.D. experiences (Sowell et al., 2015). Future studies can both target specific minoritized groups of students and ask for students' domestic/international status.

9. Conclusions

The purpose of this work was to develop an assessment that could be justifiably used to assess engineering Ph.D. students' opportunities to practice being a professional in their research. The findings of the think-aloud interviews, EFA, CFA, and group analyses show strong support that REI demonstrates evidence of validity for that purpose. As a diagnostic tool, the REI can assist national program officers, researchers, engineering administrators, and engineering faculty, to identify the range of opportunities that students have to *be* and thus *become* a professional. They could also use it to plan to create different or additional research experiences that would enhance the process of professional development of their students. Equally, the REI can be used as a research tool to explore which demographic factors are most at play in determining how different research experiences contribute to fostering professional ways of being amongst a diverse range of students.

Underpinned as it is by an ontological framework, the REI provides a new and potentially more effective approach to understanding and remedying the persistent observation that many Ph.D. graduates are not adequately prepared for professional practice (NASEM, 2018; NAE, 1983, 1995, 2017). By incorporating both technical and professional experiences, the REI provides a means to determine which research experiences provide preparation for engineering Ph.D. students in many more laboratories across the country, similar to that of an ERC research experience. The REI can provide guidance as to which steps can help to provide more opportunities where they might be needed.

The REI scores indicated that opportunities to practice being a professional are similar for women and men in their research experiences, but significantly lower for racially/ethnically minoritized groups. While these results align with previous research and contribute to the external aspect of validity, the lack of perceived opportunities for minoritized groups is cause for concern. As recommended by the report, *Doctoral Initiative on Minority Attrition and Completion* (Sowell et al., 2015), the engineering community should be proactive in creating and involving minoritized groups in research opportunities and other activities that foster their professional development.

The REI scores indicated that opportunities to practice are similar between students in Electrical and Computer Engr., Mechanical Engr., and Chemical Engr., and followed the general trend of overall REI scores. This result indicates that the REI is generalizable across engineering disciplines.

This study also adds to the limited literature on how the research experiences of engineering graduate students prepare them as future professionals (Crede & Borrego, 2012), and provides a tool for which to measure their opportunities for preparation.

Future work remains in which research experiences provide more opportunities for students to practice being a professional, and which professional development opportunities best prepare students. Being a Ph.D. student is an immensely stressful time (Evans et al., 2018; Levecque et al., 2017; Woolston, 2019), and we owe it to them to provide as many opportunities for practice in the best ways possible.

Acknowledgments

This work is part of a dissertation study located here:

https://doi.org/10.25394/PGS.12437108.v1.

References

Aldridge, M. D. (1994). Professional practice: A topic for engineering research and instruction. *Journal of Engineering Education, 83*(3), 231-236. https://doi.org/10.1002/j.2168-9830.1994.tb01108.x

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing 2014 edition*. American Educational Research Association.

Bahnson, M., Perkins, H., Tsugawa, M., Satterfield, D., Parker, M., Cass, C., & Kirn, A. (2021). Inequity in graduate engineering identity: Disciplinary differences and opportunity structures. *Journal of Engineering Education, 110*(4), 949-976. https://doi.org/https://doi.org/10.1002/jee.20427

Blair, J., Czaja, R. F., & Blair, E. A. (2013). *Designing surveys: A guide to decisions and procedures* (3rd ed.). Sage Publications.

Borrego, M., Knight, D. B., & Choe, N. H. (2017). Research group experiences and intent to complete. *Studies in Graduate and Postdoctoral Education, 8*(2), 109-127. https://doi.org/doi:10.1108/SGPE-D-17-00009

Borrell-Damian, L., Morais, R., & Smith, J. H. (2015,). *Collaborative doctoral education in Europe: Research partnerships and employability for researchers*. European University Association, Brussels, Belgium. https://eua.eu/downloads/publications/collaborative%20doctoral%20education%20in%20europe%20report%20on%20doc-careers%20ii%20project%20.pdf

Brown, P., & Souto-Otero, M. (2020). The end of the credential society? An analysis of the relationship between education and the labour market using big data. *Journal of Education Policy, 35*(1), 95-118. https://doi.org/10.1080/02680939.2018.1549752

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed*. The Guilford Press.

Bruneel, J., D'Este, P., & Salter, A. (2010). Investigating the factors that diminish the barriers to university-industry collaboration. *Research Policy*, *39*(7), 858-868. https://doi.org/https://doi.org/10.1016/j.respol.2010.03.006

Byrne, B. M. (2011). *Structural equation modeling with Mplus: Basic concepts, applications, and programming* (1st ed.). Routledge. https://doi.org/10.4324/9780203807644

Chi, M. T. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 61-82). Routledge. https://doi.org/10.4324/9780203154472

Chi, M. T., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive science, 36*(1), 1-61. https://doi.org/10.1111/j.1551-6709.2011.01207.x

Choe, N. H., & Borrego, M. (2020). Master's and doctoral engineering students' interest in industry, academia, and government careers. *Journal of Engineering Education, 109*(2), 325-346. https://doi.org/10.1002/jee.20317

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. https://doi.org/10.1037/1040-3590.7.3.309

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9. https://doi.org/10.7275/jyj1-4868

Crede, E., & Borrego, M. (2012). Learning in graduate engineering research groups of various sizes. *Journal of Engineering Education, 101*(3), 565-589. https://doi.org/10.1002/j.2168-9830.2012.tb00062.x

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/https://doi.org/10.1016/B978-1-4832-0087-3.50007-3

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods, 14*(2), 81-100. https://doi.org/10.1037/a0015914

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

Dall'Alba, G., & Sandberg, J. (2010). Learning through and about practice: A lifeworld perspective. In S. Billett (Ed.), *Learning through practice: Models, traditions, orientations and approaches* (pp. 104-119). Springer. https://doi.org/10.1007/978-90-481-3939-2_6

Denecke, D., Feaster, K., & Stone, K. (2017). *Professional development: Shaping effective programs for STEM graduate students*. Council of Graduate Schools, Washington, DC. Retrieved from https://cgsnet.org/ckfinder/userfiles/files/CGS_ProfDev_STEMGrads16_web.pdf

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/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. *International Journal of Engineering Education, 32*(5), 1960-1971.

Edge, J., & Munro, D. (2015). *Inside and outside the academy: Valuing and preparing PhDs for careers*. The Conference Board of Canada, Ottawa, ON. Retrieved from https://www.conferenceboard.ca/e-library/abstract.aspx?did=7564

Evans, T. M., Bira, L., Gastelum, J. B., Weiss, L. T., & Vanderford, N. L. (2018). Evidence for a mental health crisis in graduate education. *Nature Biotechnology, 36*(3), 282-284. https://doi.org/10.1038/nbt.4089

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299. https://doi.org/10.1037/1082-989x.4.3.272

Fernandez, R., Kozlowski, S. W. J., Shapiro, M. J., & Salas, E. (2008). Toward a definition of teamwork in emergency medicine. *Academic Emergency Medicine, 15*(11), 1104-1112. https://doi.org/doi:10.1111/j.1553-2712.2008.00250.x

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299. https://doi.org/10.1037/1040-3590.7.3.286

Gemme, B., & Gingras, Y. (2012). Academic careers for graduate students: A strong attractor in a changed environment. *Higher Education, 63*(6), 667-683. https://doi.org/10.1007/S10734-011-9466-3

Gilmore, J., & Feldon, D. F. (2010). *Measuring graduate students' teaching and research skills through self-report: Descriptive findings and validity.* Paper presented at the annual meeting of the American Educational Research Association, Denver, CO. Retrieved from https://files.eric.ed.gov/fulltext/ED509407.pdf

Godinic, D., Obrenovic, B., & Khudaykulov, A. (2020). Effects of economic uncertainty on mental health in the COVID-19 pandemic context: Social identity disturbance, job uncertainty and psychological well-being model. *International Journal of Management Science and Business Administration, 6*(1), 61-74. https://doi.org/10.18775/ijied.1849-7551-7020.2015.61.2005

Holloway, E., Douglas, K., & Radcliffe, D. (2019). *Engineering Ph.D. students' research experiences: A think-aloud study.* Paper presented at the Frontiers in Education Annual Conference, Cincinnati, OH. https://doi.org/10.1109/fie43999.2019.9028567

Holloway, E., Douglas, K. A., Oakes, W. C., & Radcliffe, D. (2021). *Research experiences instrument scoring guide*. Purdue University. Retrieved from https://docs.lib.purdue.edu/enepubs/37/

Holloway, E., & Radcliffe, D. (2018). *Review of global trends in knowledge, skills, and abilities (KSA) frameworks applicable to Ph.D. programs in engineering.* Paper presented at the American Society for Engineering Education Annual Conference, Salt Lake City, UT. https://doi.org/10.18260/1-2--30938

Holloway, E., Radcliffe, D. F., Douglas, K. A., & Oakes, W. C. (2020). *Assessing engineering Ph.D. students' research experiences: What is important to assess?* Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Virtual on line. https://doi.org/10.18260/1-2--34175

Hu, L. T., & Bentler, P. M. (1999, 1999/01/01). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0.* . IBM Corp.

Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin, 112*(3), 527–535. https://doi.org/10.1037//0033-2909.112.3.527

Kannankutty, N., Morgan, R. P., & Strickland, D. E. (1999). University-based engineering research in the United States. *Journal of Engineering Education, 88*(2), 213-222. https://doi.org/10.1002/j.2168-9830.1999.tb00437.x

Kindsiko, E., & Baruch, Y. (2019). Careers of PhD graduates: The role of chance events and how to manage them. *Journal of Vocational Behavior, 112*, 122-140. https://doi.org/10.1016/j.jvb.2019.01.010

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. The Gilford Press.

Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., & Gisle, L. (2017, 2017/05/01/). Work organization and mental health problems in PhD students. *Research Policy, 46*(4), 868-879. https://doi.org/https://doi.org/10.1016/j.respol.2017.02.008

Lumpe, A. T., & Oliver, J. S. (1991). Dimensions of hands-on science. *The American Biology Teacher, 53*(6), 345-348. https://doi.org/10.2307/4449322

Magana, A. J., & Coutinho, G. S. (2017). Modeling and simulation practices for a computational thinking-enabled engineering workforce. *Computer Applications in Engineering Education, 25*(1), 62-78. https://doi.org/doi:10.1002/cae.21779

McGagh, J., Marsh, H., Western, M., Thomas, P., Hastings, A., Mihailova, M., & Wenham, M. (2016). *Review of Australia's research training system*. Australian Council of Learned

Academies, Melbourne, Australia. Retrieved from https://acola.org/wp-content/uploads/2018/08/saf13-review-research-training-system-report.pdf

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066x.50.9.741

Morgan, R. P., Kannankutty, N., & Strickland, D. E. (1996). *Forging links between engineering education and industry: The research connection.* Paper presented at the American Society for Engineering Education Annual Conference, Washington, D.C. https://doi.org/10.18260/1-2--6059

Mosyjowski, E. A., Daly, S. R., Peters, D. L., Skerlos, S. J., & Baker, A. B. (2017). Engineering PhD returners and direct-pathway students: Comparing expectancy, value, and cost. *Journal of Engineering Education, 106*(4), 639-676. https://doi.org/10.1002/jee.20182

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide eighth edition*. Muthén & Muthén. Retrieved from https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

National Academies of Sciences, Engineering, and Medicine. (2018). *Graduate STEM education for the 21st century*. The National Academies Press. https://doi.org/doi:10.17226/25038

National Academy of Engineering. (1983). *Guidelines for Engineering Research Centers*. The National Academies Press. https://doi.org/doi:10.17226/19472

National Academy of Engineering. (1995). *Forces shaping the U.S. academic engineering research enterprise*. The National Academies Press. https://doi.org/doi:10.17226/4933

National Academy of Engineering. (2017). *A new vision for center-based engineering research*. The National Academies Press. https://doi.org/doi:10.17226/24767

National Science Foundation. (2019a). *Data tables: Degrees awarded*. National Science Foundation. Retrieved from https://ncses.nsf.gov/pubs/nsf19304/data

National Science Foundation. (2019b). *Engineering Research Centers 2019 end-of-year slides*. National Science Foundation. Retrieved from https://erc-assoc.org/sites/default/files/2019_ERC_EOY_Slides_Final_080720.pptx

National Science Foundation. (2019c). *S&E graduate students, by field, sex, citizenship, ethnicity, and race: 2016*. National Science Foundation. Retrieved from https://ncses.nsf.gov/pubs/nsf19304/data

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Sage. https://doi.org/10.4135/9781412985772

Parker, L. (1997). *The Engineering Research Centers (ERC) program: An assessment of benefits and outcomes*. National Science Foundation. Retrieved from http://www.nsf.gov/pubs/1998/nsf9840/nsf9840.htm

R Core Development Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Radcliffe, D. (2014). Multiple perspectives on engineering design. In M. Fosmire & D. Radcliffe (Eds.), *Integrating information into the engineering design process* (pp. 7). Purdue University. Retrieved from https://docs.lib.purdue.edu/purduepress_ebooks/31/

Rivera-Reyes, P., & Boyles, R. E. (2013). *Training in troubleshooting problem-solving: Preparing undergraduate engineering students for industry.* Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Atlanta, GA. https://doi.org/10.18260/1-2--22646

Roessner, J. D., Cheney, D. W., & Coward, H. (2004). *Impact on industry of interactions with Engineering Research Centers: Repeat study: Summary report*. Retrieved from http://erc-assoc.org/sites/default/files/studies_reports/Impact%20on%20Industry%20of%20ERC%20Interactions_SRI_12-04.pdf

Ronfeldt, M., & Grossman, P. (2008). Becoming a professional: Experimenting with possible selves in professional preparation. *Teacher Education Quarterly*, *35*(3), 41-60. Retrieved from http://www.jstor.org/stable/23478980

Roy, J. (2018). *Engineering by the Numbers*. American Society for Engineering Education. Retrieved from https://ira.asee.org/wp-content/uploads/2019/07/2018-Engineering-by-Numbers-Engineering-Statistics-UPDATED-15-July-2019.pdf

Säljö, R. (1999). Concepts, cognition and discourse: From mental structures to discursive tools. In W. Schnotz, S. Vosniadou, & M. Carretaro (Eds.), *New perspectives on conceptual change. Advances in learning and instruction series* (pp. 81-90). Pergamon.

Shulman, L. S. (1998). Theory, practice, and the education of professionals. *The Elementary School Journal*, *98*(5), 511-526. https://doi.org/10.1086/461912

Sowell, R., Allum, J., & Okahana, H. (2015). *Doctoral initiative on minority attrition and completion*. Council of Graduate Schools, Washington, DC. Retrieved from https://cgsnet.org/ckfinder/userfiles/files/Doctoral_Initiative_on_Minority_Attrition_and_Completion_2015.pdf

Spector, P. E. (1992). *Summated rating scale construction: An introduction (no. 82)*. Sage. https://doi.org/10.4135/9781412986038

Taylor, A. (2007). Learning to become researching professionals: The case of the doctorate of education. *International Journal of Teaching and Learning in Higher Education, 19*(2), 154-166. Retrieved from https://files.eric.ed.gov/fulltext/EJ901293.pdf

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. American Psychological Association. https://doi.org/10.1037/10694-000

Thune, T. (2009). Doctoral students on the university-industry interface: A review of the literature. *Higher Education*, *58*(5), 637-651. https://doi.org/10.1007/s10734-009-9214-0

Thune, T. (2010). The training of "triple helix workers"? Doctoral students in university-industry-government collaborations. *Minerva, 48*(4), 463-483. https://doi.org/10.1007/s11024-010-9158-7

Timmerman, B. C., Feldon, D., Maher, M., Strickland, D., & Gilmore, J. (2013). Performance-based assessment of graduate student research skills: Timing, trajectory, and potential thresholds. *Studies in Higher Education*, *38*(5), 693-710. https://doi.org/10.1080/03075079.2011.590971

Upton, G., & Cook, I. (2014). Kruskal–Wallis test. In *A dictionary of statistics*. Retrieved from https://www.oxfordreference.com/view/10.1093/acref/9780199679188.001.0001/acref-9780199679188-e-873

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management*. Retrieved from https://media.clemson.edu/cbshs/prtm/research/resources-for-research-page-2/Vagias-Likert-Type-Scale-Response-Anchors.pdf

Woolston, C. (2019). PhDs: The tortuous truth. *Nature*, *575*(7782), 403. https://doi.org/10.1038/d41586-019-03459-7

Appendix A: Race/Ethnicity Scores

Table 16 Mean factor scores for all race/ethnicity subgroups

Race/ethnicity group	*n*	%	Mean Scores *Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*	*Opportunity* *M* (SD)
All	451	100	4.21	3.63	3.16	4.23	4.15	19.38 (4.05)
Asian students	159	35.3	4.25	3.83	3.31	4.53	4.25	20.17 (3.94)
Black / African American students	9	2.0	3.94	2.80	2.91	3.19	2.96	15.81 (5.11)
Hispanic or Latino students	26	5.8	4.20	3.84	3.39	3.97	3.40	18.81 (4.25)
Native American students	0	0.0	0.00	0.00	0.00	0.00	0.00	0.00
Other students	12	2.7	4.25	3.83	3.31	4.53	4.25	20.17 (3.94)
Pacific Islander students	0	0.0	0.0	0.00	0.00	0.00	0.00	0.00
Prefer not to respond	9	2.0	4.08	3.30	2.75	4.20	4.31	18.65 (3.53)
White students	202	44.8	4.24	3.53	3.08	4.08	4.27	19.21 (3.99)
All minoritized groups	51	11.3	4.27	3.49	3.06	3.85	3.38	18.06 (4.05)
Multiple selected minoritized groups	16	3.5	4.58	3.31	2.61	4.03	3.57	18.10 (4.41)
Multiple selected non-minoritized groups	18	4.0	4.07	3.67	3.32	4.26	4.24	19.56 (3.53)
Non-minoritized groups	400	88.7	4.21	3.65	3.17	4.28	4.24	19.55 (3.96)

Appendix B: Engineering Discipline Scores

Table 17 Mean factor scores for all engineering (Engr.) discipline subgroups

Engineering discipline group	*n*	%	Mean scores *Teamwork*	*Collaboration*	*Networking*	*Modeling*	*Experimentation*	*Opportunity* *M* (SD)
All	451	100	4.21	3.63	3.16	4.23	4.15	19.38 (4.05)
Aeronautics and Astronautics Engr. students	35	7.8	4.22	3.51	3.34	4.66	3.66	19.39 (4.15)
Agricultural and Biological Engr. students	1	0.2	3.25	3.17	1.50	4.33	4.33	16.58
Biomedical Engr. students	39	8.6	4.62	3.73	3.01	3.91	4.86	20.14 (3.04)
Chemical Engr. students	50	11.1	4.82	3.43	3.07	4.07	4.72	20.10 (3.35)
Civil students	35	7.8	4.14	3.74	3.36	4.56	4.15	20.00 (3.67)
Construction Engineering and Management students	3	0.7	3.08	4.11	3.17	4.11	3.28	17.75 (6.87)
Electrical and Computer students	82	18.2	4.07	3.67	3.36	4.82	4.01	19.93 (3.94)
Engineering Education students	28	6.2	4.64	3.85	2.68	2.34	2.26	15.76 (3.73)
Environmental and Ecological Engr. students	12	2.7	3.69	3.60	3.52	3.89	3.85	18.54 (4.27)
Industrial Engr. students	20	4.4	3.26	3.31	3.11	4.14	3.30	17.13 (4.81)
Materials Engr. students	32	7.1	4.35	4.18	3.45	3.63	5.16	20.76 (3.07)
Mechanical Engr. students	87	19.3	4.02	3.27	2.86	4.47	4.36	18.98 (4.41)
Nuclear Engr. students	8	1.8	4.22	4.00	3.34	4.54	4.15	20.25 (3.11)
Other Engr. major students	19	4.2	4.09	4.27	3.54	4.24	3.99	20.13 (4.50)

Appendix C: Research Experiences Instrument (REI)

The following questions are related to engineering Ph.D. students' perceptions about how often in their Ph.D. research experiences they utilized skills relevant to professional practice. Respondents are asked to take special note that this is NOT an assessment of their research skills; rather, it is an assessment of their perceptions about their research experience and how often they did things related to professional practice in their research experience.

In the survey below, there are 29 questions, and it is estimated it will take approximately 15 minutes to complete. For each of the questions that follow, please indicate how often in your Ph.D. research experience you did the items in question. You should do this by selecting the response that most closely matches from the options provided.

Table 18 Research Experiences Instrument (REI)

How often in your Ph.D. research experience did you:					
1 – Never	2 – Very Rarely	3 – Rarely	4 – Occasionally	5 – Frequently	6 – Very Freq

Teamwork	take on different roles or responsibilities within a research group?
	coordinate research tasks with other graduate students?
	share decision making responsibility with other graduate students?
	develop new skills (e.g., presentation, project management, software, etc.) based on the needs of the research group's goals? (this item was removed at EFA)
	mutually depend on other graduate students to meet the desired outcomes?
Collaboration	present your research results to your sponsors or collaborators who are involved in your research?
	interact at your university with your sponsors or collaborators who are involved in your research?
	correspond (e.g., email, phone, etc.) with your sponsors or collaborators who are involved in your research?
	interact with your sponsors or collaborators at their place of work related to your research?
	co-author journal or conference papers with your sponsors or collaborators?
	co-create a presentation with your sponsors or collaborators?
Networking	develop professional relationships with practicing engineers through research?
	participate in industry or government conferences as part of research?
	participate in professional engineering societies (e.g., Institute of Electrical and Electronics Engineers, Society of Women Engineers, etc.)? (this item was removed at EFA)
	present results of your research to practicing engineers?
	interact with practicing engineers during internships or co-ops?
	interact with support professionals (e.g., project managers, building maintenance, outside vendors, etc.)? (this item was removed at EFA)
Modeling	develop or utilize a mathematical model to help solve a problem?
	specify constraints or assumptions in development of a mathematical model to help solve a problem?
	utilize sophisticated tools to help solve a modeling or simulation problem?
	simulate a system to obtain results?
	iterate on the development of a model or simulation to optimize results?
	verify a model or simulation based on real-world data or actual results?
Experimentation	use testing equipment or instrumentation as an integral part of conducting your research?
	develop plans to use testing equipment or instrumentation?
	ensure testing equipment or instrumentation is appropriately set-up (i.e., calibrated) before use?
	collect data from testing equipment or apparatus using appropriate sensors or instrumentation?
	interpret data gathered from testing equipment or apparatus?
	troubleshoot or modify testing equipment or instrumentation when it does not operate properly?

Note: For the complete final instrument, refer to https://docs.lib.purdue.edu/enepubs/37/ or contact the lead author.

Eric A. Holloway is Courtesy Faculty in the School of Engineering Education, Courtesy Assistant Professor of Mechanical Engineering, and the Senior Director – Industry Research, all in the College of Engineering at Purdue University, 516 Northwestern Avenue, West Lafayette, IN 47907; eahollow@purdue.edu; ORCID: 0000-0002-0343-1709.

Kerrie A. Douglas is an Assistant Professor in the School of Engineering Education at Purdue University, 516 Northwestern Avenue, West Lafayette, IN 47907; douglask@purdue.edu; ORCID: 0000-0002-2693-5272.

David F. Radcliffe is the Kamyar Haghighi Head of Engineering Education and Epistemology Professor Emeritus of Engineering Education at Purdue University, 701 West Stadium Avenue, West Lafayette, IN 47907; dradcliffe@purdue.edu.

William C. Oakes is the 150th Anniversary Professor, Professor in the School of Engineering Education, and the Director of the EPICS Program at Purdue University, 701 West Stadium Avenue, West Lafayette, IN 47907; oakes@purdue.edu; ORCID: 0000-0002-6183-045X.