Received: 25 February 2016 | Revised: 13 February 2020 | Accepted: 12 May 2020

DOI: 10.1002/jee.20355

JEE Journal of Engineering Education

ORIGINAL ARTICLE

Self-directed information literacy scale: A comprehensive validation study

Kerrie A. Douglas[1] | **Todd Fernandez**[2] | **Michael Fosmire**[3] | **Amy S. Van Epps**[4] | **Senay Purzer**[1]

[1]School of Engineering Education, Purdue University, West Lafayette, Indiana

[2]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia

[3]Libraries and School of Information Studies at Purdue University, West Lafayette, Indiana

[4]Harvard Science Center, Cabot Science Library, Cambridge, Massachusetts

Correspondence
Kerrie A. Douglas, School of Engineering Education, Purdue University, 516 Northwestern Avenue, West Lafayette, IN 47906.
Email: douglask@purdue.edu

Funding information
National Science Foundation TUES, Grant/Award Number: 1245998

Abstract

Background: Engineering and technology students need to acquire, evaluate, apply, and document information to solve complex ill-defined problems. However, there are few assessment tools to evaluate how these students approach information.

Purpose: The authors integrated information literacy standards with self-directed learning theory to create the self-directed information literacy (SIL) scale as an assessment of students' self-directedness using high-quality information skills in engineering projects. The purpose of this research is to examine the measurement properties of SIL to inform the use and interpretation of SIL's scores among diverse learners.

Method: The authors administered SIL to first-year engineering and technology students ($n = 1{,}603$). To test three hypotheses about SIL scores related to validity, reliability, and fairness, we conducted a series of psychometric analyses, including confirmatory factor analysis (CFA) and measurement invariance between groups.

Results: The results of the CFA were acceptable for the proposed higher-order factor structure of SIL and a separate *Beginner Behavior* factor. Measurement models for male and female scores were found invariant; however, there were measurement differences between groups of students based on their experience with instruction in the English language.

Conclusions: SIL can be scored to assess engineering and technology students' specific SIL subfactors (*Recognize*, *Seek*, *Evaluate*, *Apply*, *Document*, and *Reflect*) or scored as an overall broader measure of self-directedness with information. A separate factor, *Beginner Behavior*, can be used to moderate ceiling effects. SIL scores can be used for gender comparisons but should be carefully evaluated when the sample includes students who are new to curriculum and instruction in the English language.

KEYWORDS

assessment, confirmatory factor analysis, engineering design, engineering students, information literacy, measurement invariance, self-directed learning theory

1 | INTRODUCTION

Engineers continuously engage with new information. The critical role of gathering information and integrating new, useful information to inform design decisions is well established in the literature (Atman et al., 2007; Atman, Kilgore, & McKenna, 2008; Mosborg et al., 2005). Whether eliciting more depth of information from a design brief, finding the best material for a component, synthesizing ideas to create new technologies, or communicating evidence for their decisions, engineers continually need to learn and integrate new information into their engineering practice.

Information search behavior distinguishes experts from novices, where experts seek more varied types of information and use a diverse array of search strategies (Johnson, 1988). In many situations, engineers work with internal information in their context (Cervero, Miller, & Dimmock, 1986). Those with advanced degrees or exposure to information literacy coursework are more likely to consult external information sources such as journals, standards, and patents (Holland & Powell, 1995; Kwasitsu, 2004). Similarly, studies of engineering students' final design reports have shown that senior students use more resources than first-year students (Atman et al., 2007; Yu, Sullivan, & Woodall, 2006). Atman and colleagues (1999) suggested one explanation for the differences in these behaviors is that first-year engineering students may not be aware of the need to gather information before developing a solution. More recent research has found that first-year students' information literacy behavior also depends upon their personal investment or motivation in the project (Lamont et al., 2020). The researchers noted that when students thought the project was meaningful, students engaged in more information literacy behaviors.

Researchers agree on the necessity to develop information skills especially concerning the use of diverse resources, appropriate documentation and citation, strategies used when searching for information, and selection of which database to search (Denick, Bhatt, & Layton, 2010; Hensel, Brown, & Strife, 2012; MacAlpine & Uddin, 2009; Welker, McCarthy, Komlos, & Fry, 2012; Wertz, Purzer, Fosmire, & Cardella, 2013). ABET's recent changes emphasize that engineering students need to develop the “ability to acquire and apply new knowledge as needed, using appropriate learning strategies” in its approved accreditation standards for the 2020–2021 academic year (ABET, 2020, p. 6).

While many disciplinary faculty and librarians agree on the need for integrating information literacy into the design process (Fosmire & Radcliffe, 2013; Williams et al., 2004), students may not see information literacy as something they need to learn. The Attaining Information Literacy study (Gross & Latham, 2007) and a follow-up study (Gross & Latham, 2011) both found that students who performed poorly on an information literacy skills assessment also believed they were proficient in finding and using information. The Dunning–Kruger effect (Dunning, 2011) refers to such situations when one is not aware of his or her shortcomings. Engineering and technology students as a group may be particularly susceptible to the Dunning–Kruger effect. For example, first-year engineering students tend to rate their information skills rather high (Douglas, Wertz, Fosmire, Purzer, & Van Epps, 2014) despite research suggesting first-year students struggle with correctly applying and documenting information in design tasks (Wertz et al., 2013). Thus, to equip engineering and technology students with information literacy skills, instructors need to cultivate students' intrapersonal beliefs and habits regarding the role of information in design.

Whether learners' self-reports are entirely accurate or not, how learners perceive themselves does influence what they do (Paulhus & Vazire, 2007). For example, much literature supports that context-specific self-efficacy is a mediator in various forms of academic achievement (Schunk, Meece, & Pintrich, 2012). Additionally, Litzinger, Wise, and Lee (2005) found correlations between measures of learners' self-directed learning and their grade point average. Theory-driven self-report surveys are an essential part of educational research and can be used to effectively to study perceptions that influence student learning when validated for that purpose.

High-quality assessments have evidence of reliability, validity, and fairness (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education Association, & National Council on Measurement in Education [NCME], 2014). Fairness is particularly important in engineering education because undergraduate engineering programs in the United States are approximately 78% male, 62% White, and 90% domestic (Roy, 2019). Thus, for most validation studies on engineering classroom assessments, measurement models for examining aspects of validity and reliability will be largely influenced by one group of learner, that is, White domestic males whose native language is English. The extent to which diverse students read and interpret the assessment questions in the same way or understand the context of the questions in the same way is unknown without further studies. Fairness in educational assessment includes many aspects, including being free of measurement bias, equal opportunity to learn, recognition of individual differences, free of stereotype threat, accommodations, and so forth (AERA, APA, & NCME, 2014), and a crucial step is to examine measurement invariance between groups to empirically evaluate the potential presence of bias toward or against a group of learners.

2 | RESEARCH PURPOSE

While researchers have developed several self-directed learning scales (e.g., Fisher, King, & Tague, 2001; Guglielmino, 1977; Knowles, 1975; Williamson, 2007), we were unable to locate an instrument developed specifically for engineering and technology students in the context of information literacy. To begin research focused on how to educate engineering and technology learners to value and actively develop information literacy skills, there is a need to assess their self-directedness with information literacy. The purpose of this research article is to describe a validation study of the self-directed information literacy (SIL) scale, designed as a measure of students' reports of self-directedness with information when engaged in engineering projects. We approached the development of SIL as a construct-centered design process, following recommendations for scale development (Netemeyer, Bearden, & Sharma, 2003).

After we generated the scope and items for SIL, we examined evidence of validity and reliability following an argumentation approach to validation (Kane, 2016, 1992). Assessment validation is the process of scientific argumentation, where one tests the plausibility that resulting scores can be interpreted as intended (Kane, 1992). Jorion et al. (2017) also follow this approach in evaluating the claims of concept inventories. Based on the Argument-Based Approach for validation framework (Kane, 1992), we first articulated how we intend SIL scores to be interpreted and used, then identified the relevant information to evaluate the appropriateness of those interpretations and uses. The Standards for Educational and Psychological Testing state that the cornerstones of quality assessment reside in evidence of reliability, validity, and fairness (AERA, APA, & NCME, 2014). Thus, we made the following conjectures: If SIL measures self-directed information literacy, then (a) items written for each factor will be internally consistent (internal reliability); (b) the factor analysis model results will be aligned to the theoretical design (structural aspect of validity); and (c) there will be measurement invariance between demographic groups of students (an aspect of fairness).

Specifically, we asked the following research questions: To what extent are the SIL scale items written to represent a factor internally consistent? To what extent does SIL's factor structure align with the theoretical framework? To what extent are there measurement invariances between gender and language groups?

3 | PHASE 1: SIL SCALE DEVELOPMENT

SIL was designed based on integration of Information Literacy Standards (Association of College and Research Librarians, 2000) and self-directed learning theory (Knowles, 1975). In addition, we followed Netemeyer et al.'s (2003) guidelines for scale development to create the SIL scale. These guidelines include four steps: (a) construct definition; (b) item generation and evaluation; (c) designing and conducting studies to develop and refine the scale; and (d) finalizing the scale. This process was iterative as item refinements occurred with the collection of new evidence at each step. Furthermore, our development and validation procedures are grounded in modern approaches to educational assessment validity where validity is not a property of the assessment but rather an evaluation of the resulting scores for a particular purpose (Messick, 1987). The design and evaluation of an assessment instrument is based on a chain of evidentiary reasoning needed for a particular use (Mislevy, 2007).

3.1 | Construct definition

An essential first step in developing a chain of reasoning to base a validity claim is for the researchers to specify a theory and define the construct(s) to be measured (Cronbach & Meehl, 1955). Thus, we began by defining what SIL is in the context of undergraduate engineering classrooms. We define SIL based on the incorporation of definitions from three sources (see Figure 1): (a) the information literacy standards developed by the Association of College and Research Libraries (ACRL, 2000), (b) Self-directed learning theory (Knowles, 1975), and (c) characteristics of information literacy behavior of engineering students identified through qualitative interviews (Douglas, Wertz, et al., 2014; Douglas, Van Epps, Mihalec-Adkins, Fosmire, & Purzer, 2015).

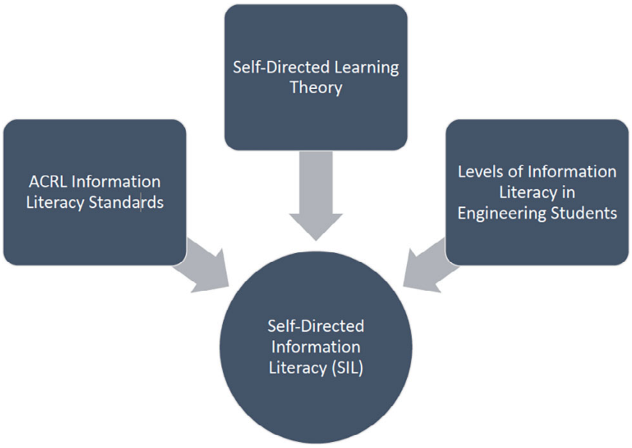

FIGURE 1 The construct definition model of self-directed information literacy. ACRL, Association of College and Research Libraries [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Dimensions of self-directed information literacy for engineering students

Dimension	Self-directed information literate student behaviors
Recognize	Begin projects by analyzing the problem for information needs. As part of the problem scoping and task definition stages of design, they think critically beyond the information provided to identify elements of underlying intent, potential ambiguities, gaps in the provided information, and gaps in their knowledge. They ask or form questions to identify what information they need and develop a plan to obtain that information.
Seek	Develop a contextualized information search strategy based on their current level of knowledge and gaps in their understanding. They use an intentional and structured process to gather formal and informal information. They are able to identify the appropriate resources (e.g., databases, search engines, forums, colleagues) to find specific information (e.g., patents, industry standards). They are also able to efficiently navigate and access information.
Evaluate	Probe and determine the credibility of their information sources through use of evaluative criteria and heuristics (e.g., relevancy, dates of publication, purpose of source, intended audience, scholarly agreement). They are able to identify trustworthy and appropriate sources of information for their intended use and articulate why they are relevant.
Use	Incorporate found and personal information into their ideation, analysis, and sense-making process. They use information to inform all aspects of their project from problem scoping to concept generation and testing. They resolve rather than just reject information that may contradict their ideas or lead to a change in a design solution.
Document	Organize, document, and appropriately cite their information so that others may obtain access to their sources. When writing reports or presenting orally, they reference the source of the information used in making decisions in accordance with professional norms (e.g., professional society standards, policies).
Reflect	Give careful thought to how they handled information in the course of their project: How they recognized their information needs, gathered, evaluated, used, and documented information. They reflect on what they did, what information they still do not know, what strategies worked well, and what could be done differently in future projects.

There is much overlap between self-directed learning behaviors and information literacy behaviors. Self-directed learning theory (Knowles, 1975) postulates five critical behaviors of the self-directed learner: diagnosing learning needs, formulating learning needs, identifying human/material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes. Similarly, according to ACRL (2000), information literacy involves recognizing when information is needed and the extent needed; locating the needed information effectively and efficiently; evaluating information and its sources critically; and using this information effectively, ethically and legally. Table 1 describes our definition of SIL synthesized from these sources.

3.2 | Item generation and evaluation

Over the course of 2 years, the research team wrote, tested, and refined items to assess facets of student SIL skills (Douglas, Fernandez, Purzer, Fosmire, & Van Epps, 2015; Wertz, Ross, Fosmire, Cardella, & Purzer, 2011). During item generation and evaluation, we wrote Likert-scale items on a scale of 1 to 5, Strongly Disagree to Strongly Agree, for each subarea of SIL shown in Table 1. In addition, to control for students' tendency to rate self-report measures rather positively (e.g., Mowbray, Boyle, & Jacobs, 2015), we also wrote items to describe less effective SIL behaviors such as, "Once I think of a solution, I try to stick with it." Through this process, a total of 74 potential SIL items were created. Next, to obtain evidence of face validity, we requested 10 engineering and librarian faculty members map each item to a construct definition. We asked these experts for feedback on the construct definitions and wording of items. Items were deleted when experts did not have a clear consensus in mapping them to the intended constructs. In addition, revisions were made to clarify the construct definitions.

Next, we conducted a think-aloud study with 10 engineering students from diverse backgrounds to ascertain how students read and interpret the items (Douglas, Wertz, et al., 2014). Think-alouds are helpful in gathering evidence on substantive aspects of validity (i.e., the extent to which learners read and interpret survey items as the writers intended) as well as identifying items that need to be revised (Blair, Czaja, & Blair, 2013). We told students that there were no right or wrong answers and that the purpose was to inform the development of the assessment instrument. When students re-read items or paused to think, we noted this and further probed the students to talk about their thoughts. We incorporated observations of students' behavior and their struggles with the items into the next round of revisions. After this item evaluation phase, SIL had a total of 69 items.

3.3 | Pilot study and revision

To examine how well SIL items functioned in terms of means, item relationships, and factor structure, SIL was piloted with first-year engineering and technology students ($n = 366$) (Douglas, Fernandez, et al., 2015). We used an exploratory factor analysis (EFA), following the method of Fabrigar, Wegener, MacCallum, and Strahan (1999) where items with insignificant interitem correlations within a proposed factor and poor factor loadings were discarded. This process resulted in 38 items spanning six factors; *Apply*, *Document*, *Evaluate*, *Recognize*, *Reflect*, and *Seek*. The resulting SIL items had a high internal consistency ($\alpha = 0.92$), but construct representation was a concern after the removal of so many items. Furthermore, item means were negatively skewed ($M = 3.83$ out of 5), indicating students still tended to rate themselves quite highly.

Items that functioned well in the pilot data were maintained, yet several additions and revisions were made to ensure construct representativeness and further attempt to limit acquiescence. First, new items were written to cover the full aspect of each proposed factor. The new items were written to be more sensitive to the variances of student behavior found in student interviews (Douglas, Van Epps, et al., 2015) and tap into more advanced skills (e.g., *I acknowledged sources of information when formally communicating about the project* was revised to *I was transparent about the role of similar products in informing my project*). One suggestion for dealing with acquiescence is to present survey respondents with both positive and negative behavior (e.g., Spector, 1992). For data analysis and interpretation purposes, the negative behavior items are reverse scored so that they are on the same scale as the positive behavior items. From this recommendation, we added items for each of the six dimensions that described beginner (and less effective) information strategies. This resulted in our creating a seventh factor, *Beginner Behavior*. Examples include *trusting that information from .org sites is always unbiased, waiting until the project is at the end to find the required number of citations*, and *using general search engines to find technical information*.

We created additional instructions for the students before they answered the Likert-type scale items. We asked them to concretely think about how they used information when engaged in a specific design project. The instructions directed students to first respond to three prompts: (a) *Describe a recent engineering project*; (b) *What new information did you need to learn to complete the project?*; and (c) *How did you go about acquiring the information needed?* These constructed-response items were used exclusively to create a context for the students to focus on before answering the Likert-type scale items. As such, they were not analyzed in the current study. To allow for more variance in responses, the scale was also changed from a 5-point scale to a 7-point scale (1—*Not at All True of Me* to 7—*Very True of Me*), based on the recommendations of Preston and Colman (2000); see the appendix for a list of the SIL items. Once again, we sent out the item pool and proposed factors to a panel of experts. We requested feedback from both members of the ASEE Engineering Libraries Division through the ELDNET-L listserv, as well as from engineering faculty ($n = 39$). The expert panel was provided with SIL's proposed factors and definitions and then asked: *Do you agree with these subcomponents?* and *What concerns do you have with the conceptualizations of these categories?* Next, the panel was asked to classify each item according to the given proposed factors of SIL. Finally, the panel was asked, *In your opinion, are there any other aspects of information literacy related to engineering and* *technology students that these questions and categories do not cover?* The panel's classification of items and feedback on SIL's subareas were used to make further revision to the instrument. In total, at this stage, SIL had 49 Likert-scale items and five open-ended questions, plus one item to check whether respondents were engaged in the survey, *If you are reading this*, *select the number* 4.

4 | PHASE 2: VALIDATION STUDIES

The research questions driving the validation study are (a) To what extent are items written to represent a factor internally consistent? (b) To what extent does SIL's factor structure align with the theoretical framework? and (c) To what extent are there measurement invariances between genders and language groups?

4.1 | Methods

4.1.1 | Setting and participants

We administered SIL to students in two first-year introductory engineering design courses, one in a college of engineering and the other in a college of technology, during the fall semester of 2015. Students completed SIL using an online survey tool as part of course modules on information literacy in engineering design.

Overall, 1830 students completed the SIL instrument. The survey presented demographic questions for respondents to select their gender (Male, Female, Other, Prefer Not to Answer) and report their experience with instruction in the English language. Based on data from the registrar office, the student population was 85% White, 8% Asian, 4% Latinx, and 2% African American or Black, and 1% Other; also, 24% were international students. From the demographic breakdown, we understood that we did not have a sample sufficient to conduct measurement tests between racial or ethnicity groups. We cleaned the data based on response quality, following recommendations from Meade and Craig (2012) where respondents' data were deemed careless if there were no variation in their responses to items (e.g., selecting all "I strongly agree" for all items); time spent responding to the survey was insufficient for reading the survey statements (e.g., finished in under a minute); or answered the filter statement incorrectly, "If you are reading this statement, please select the number 4." This process resulted in the removal of 227 responses deemed careless and a final sample of 1,603 responses was used for analysis. The majority of respondents (88%) were from the first-year engineering course, a much larger course enrollment than the first-year technology course. The breakdown of participant by demographic variables is presented in Table 2.

To determine students' experience with the English language, the survey asked students to report the language of instruction at their high school (the students were in their first-year at university). For analytic purposes, these data were categorized into two groups: experienced with English language (EEL) and English language learner (ELL). EEL were students who indicated having had previous instruction in the English Language at their prior educational institution. ELL was the grouping of students who indicated they did not have instruction in the English language at their previous educational institution. In addition, items written as *Beginner Behavior* were reverse scored to place them on the same scale interpretation as items in the other factors, as suggested by Spector (1992). So, for example, answers of a 1 were scored as a 7; 2 became 6; and vice versa.

TABLE 2 The demographic distribution of participants (n = 1,603)

Group	Technology	Engineering	Total
Male	147 (78%)	1,064 (75%)	1,211
Female	41 (22%)	343 (24%)	384
Other	2 (<1%)	7 (<1%)	9
EEL	173 (92%)	1,265 (89%)	1,438
ELL	16 (8%)	149 (11%)	165
Total	189 (12%)	1,414 (88%)	1,603

Abbreviations: EEL, students experienced with English language instruction; ELL, English language learner.

4.1.2 | Data preprocessing

We evaluated the appropriateness of analyzing the data using confirmatory factor analysis (CFA) following common guidelines (Matsunaga, 2010; Thompson & Daniel, 1996). For each item, the mean, variance, skew, and kurtosis were calculated. Additionally, the bivariate correlations between all pairs of items within a proposed factor were calculated. Finally, we calculated the coefficient alpha for the overall scale and with each item removed. Four items did not satisfy the criteria for inclusion in CFA and were discarded prior to further analysis.

4.1.3 | CFA procedures

Both EFA and CFA are based on the common factor model, where CFA assesses the fit of data to a presupposed measurement model through the use of goodness-of-fit indices (Jöreskog, 1967). We had previously conducted EFA and had a well-defined theory (Douglas, Fernandez, et al., 2015), so we proceeded to measure the fit of the hypothesized relationship of the items. We used the Lavaan package within the R statistical programming language (Rosseel, 2012) to test the hypothesized relationship of scores. In addition, the hypothesized model was compared to three other alternative models holistically, based on common recommendations (T. A. Brown, 2015; Kline, 2005).

There are a number of goodness-of-fit indices available to more holistically evaluate CFA models. The indices we examined cover all three categories of fit indices for CFA: comparative (also called incremental), absolute, and parsimony-corrected (also called predictive) (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Hu & Bentler, 1999; Kass & Raftery, 1995). Comparative indices (e.g., comparative fit index [CFI] and the Tucker–Lewis index [TLI]) evaluate the proportional improvement of fit for a hypothesized model in comparison to the baseline model (Hu & Bentler, 1999). Absolute indices (e.g., root mean square error of approximation [RMSEA]) evaluate the extent to which the model fits the data with no comparison to a baseline model (Byrne, 2012). Parsimony-corrected indices of fit (e.g., Bayesian information criterion [BIC]) compare non-nested models to evaluate the model fit as well as correct for the complexity of the model.

Following the example of Byrne (2012), the overall measurement model was evaluated based on five goodness-of-fit indices: (a) ratio of chi-square to degrees of freedom, (b) RMSEA, (c) CFI, (d) TLI, and (e) BIC. The ratio between χ^2 and degrees of freedom (χ^2/df), is interpreted as lower ratios indicating better fitting models (Byrne, 2012). Next, we examined RMSEA, which compares the difference of the variance–covariance matrix of the theorized model to the data. For RMSEA, lower values indicate better model fit, with values below 0.08 considered acceptable and values below 0.05 considered good (Hu & Bentler, 1999). CFI and TLI compare the hypothesized model fit to the baseline model. CFI is normed on a scale of zero to one, where higher values are indicative of better fit. Values close to 0.90 are considered acceptable and values close to 0.95 are good (Hu & Bentler, 1999). TLI is non-normed, but it is interpreted similar to CFI. TLI also penalizes overly complex models. BIC is a parsimony index, which compares non-nested models, where smaller values represent the best fit, given the complexity of the model (Schwarz, 1978). After evaluation of alternate models, we then refined the best fitting model by calculating modification indices and then evaluating suggested modifications based on theory and rationale (Steiger, 1990).

4.1.4 | Measurement invariance testing

Prerequisite to the reporting of group mean comparisons, such as mean differences or correlation coefficients, is that the assessment scores have the same meaning across groups of learners (Chen, Sousa, & West, 2005). This is a central concern for ensuring fair assessment of minority learners in engineering. Measurement invariance is broadly understood as scores for the same measure having the same meaning under different conditions (Meade & Lautenschlager, 2004). Differing conditions can include populations, time, and methods of administration.

Equal group sizes were created by randomly sampling the same number of male responses (384) to match the total of female responses and randomly sampling from EEL to match the total number of ELL (165). Equal group sizes are necessary for comparing group differences without generating bias toward the larger group. For male and female comparison, each group had 384 students (total $n = 768$). For the language comparison, each group had 165 students (total $n = 330$).

We conducted several tests of measurement invariance from a hierarchical approach, where each model imposes additional constraints and compares the model fit to the previously analyzed model. Following recommendations from T. A. Brown (2015) and Byrne (2012), measurement invariance was examined using the Lavaan R package (Rosseel, 2012) through testing the following models: Model 1: configural invariance, which constrains the groups to have the same factor pattern (Chen et al., 2005); Model 2: metric invariance, which constrains the groups to have equal factor loading; Model 3: scalar invariance, which constrains the intercepts as equal across groups; Model 4: error invariance, which tests whether the error variance for each item is equal between groups; Model 5: means invariance, which constrains the latent means to be equivalent between group as a test of population heterogeneity. We followed Byrne's (2012) recommendation that the following be used to evaluate measurement invariance: CFI, RMSEA, the changes in chi-square and degrees of freedom comparing models with increasing constraints, and p-values of whether changes in chi-square are significant.

Next, we compared group mean scores using t-tests for significance and Cohen's d to assess effect sizes. Students' scores on each factor were calculated from the final model by averaging their responses to each item from that scale. For each group comparison, a two-tailed t-test was conducted assuming unequal variances to test for mean differences in the scores of students of each group. When the results were found to be significantly different (at $p < .05$), effect size using Cohen's d was also calculated.

5 | RESULTS

5.1 | Descriptive statistics

Item level means ($M = 4.87$), standard deviation ($M = 1.33$), skew ($M = -0.43$), and kurtosis ($M = 0.13$), were found acceptable for factor analysis (Fabrigar et al., 1999; Kline, 2005). Bivariate correlations between each item and all other items from the same conceptual factor were calculated to identify items that would not be appropriate for CFA. Four items did not significantly correlate to any other item in their theoretical factor and were therefore discarded, leaving 46 items. Correlations between each conceptualized factor were also calculated as a check that items within a factor correlated higher than items conceptualized as another factor. Table 3 shows the average interitem correlations per factor and bivariate correlations across factors after the removal of insignificant items. The highest average correlation among items of the same subscale was in the *Reflect* scale ($r = 0.51$), while the lowest was on the *Recognize* scale ($r = 0.23$). Furthermore, all factors were positively correlated, suggesting a possible higher-order factor of SIL. A higher-order factor is when factors significantly correlate, have shared variation, and are conceptually understood to capture a larger construct rather than just unrelated independent factors. The overall instrument based on coefficient alpha was high ($\alpha = 0.92$), further evidencing that the items were appropriate for CFA.

5.2 | Results of CFA

We conducted a CFA with the 46 items determined to be acceptable for CFA. To evaluate the model fit based on the theoretical framework, we not only tested the hypothesized model in a CFA but also alternative models as suggested by Kline (2005). The first model (Table 4—Original) loaded items onto the six factors only with no higher-order factor. The model produced less than acceptable fit indices with a TLI of 0.74 and a CFI of 0.76. Two other models were then analyzed: a higher-order factor (Table 4—Higher-Order), where SIL was modeled as the compilation of the subfactors

TABLE 3 Average bivariate correlation coefficients for items with each factor

Subscale	Recognize	Seek	Evaluate	Apply	Document	Reflect
Recognize	0.23					
Seek	0.15	0.27				
Evaluate	0.14	0.19	0.27			
Apply	0.15	0.16	0.19	0.33		
Document	0.15	0.20	0.22	0.18	0.31	
Reflect	0.20	0.21	0.23	0.20	0.24	0.51

TABLE 4 Comparison of CFA fit indices

Model	X^2	*df*	X^2/df	RMSEA	CFI	TLI	BIC
Null	21,861*	820	26.7	0.13	**	**	**
Unstructured reference							
Original	7,212*	983	7.34	0.06	0.76	0.74	236,653
Six theoretical factors, no higher order, *Beginner Behavior* items load with other factors as conceptually written							
Higher-order	5,946*	1,035	5.74	0.06	0.81	0.80	234,760
All seven factors load on a higher-order *InfoLit* factor							
Two-component	3,462*	759	4.34	0.05	0.87	0.86	196,683
Remove *Beginner Behavior* factor from higher-order factor Creates two score components							

Abbreviations: BIC, Bayesian information criterion; CFA, confirmatory factor analysis; CFI, comparative fit index; RMSEA, root mean square error approximation fit index; TLI, Tucker–Lewis fit index.
$^*p < .001$; **TLI and BIC are comparative measures and are not meaningful for the null model.

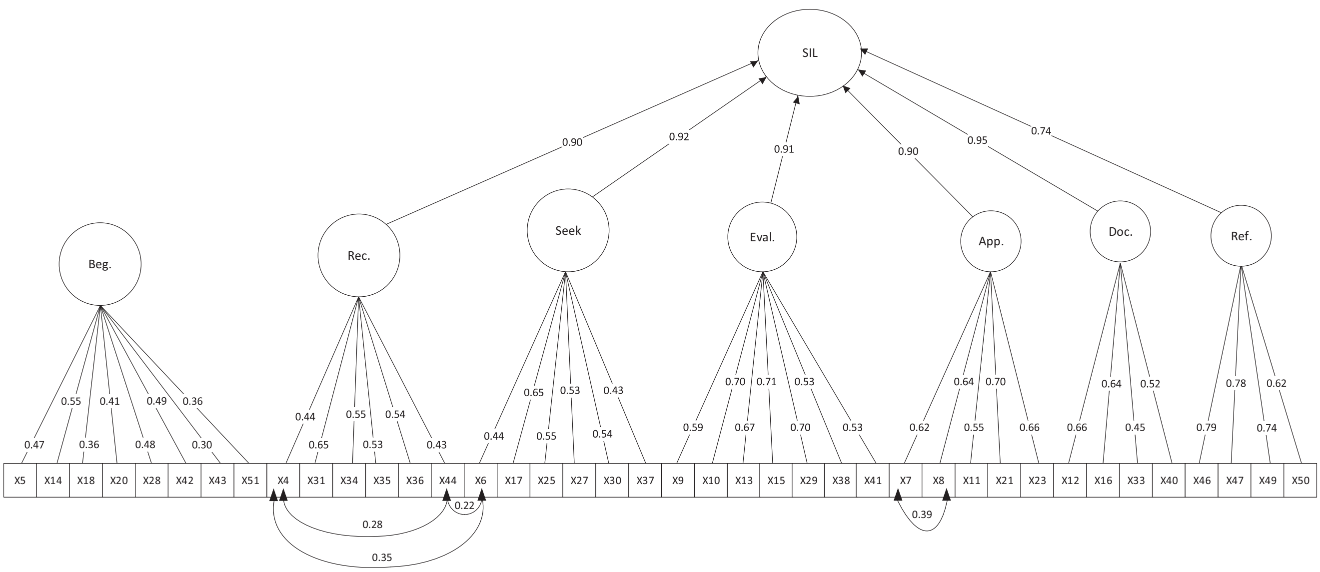

FIGURE 2 Two-component CFA model with standardized loadings. App., Apply; Beg., Beginner Behavior; CFA, confirmatory factor analysis; Doc., Document; Eval., Evaluate; Rec. Recognize; Ref., Reflect; SIL, self-directed information literacy

Recognize, *Seek*, *Evaluate*, *Use*, *Document*, *Reflect*, and *Beginner Behavior*, and a two-component model with a higher-order SIL (*Recognize*, *Seek*, *Evaluate*, *Use*, *Document*, *Reflect*) and an entirely separate *Beginner Behavior* factor.

Based on the overall evaluation of the models and in alignment with theory, we selected the two-component model as the final model. SIL is understood as a higher-order factor with subcomponents of *Recognize*, *Seek*, *Evaluate*, *Use*, *Document*, and *Reflect*. Extracting higher-order factors whenever factors are correlated is commonly recommended to provide more clarity into the structure underlying the data (Thompson, 2004). Because the model fit was less than acceptable when *Beginner Behavior* was included as a subfactor along with the other six factors, we conclude that it should be scored separately from the other factors. Additionally, other researchers have noted that negatively scored items tend to not factor well with positively scored items even if they are intended to measure the same construct (Tomas & Oliver, 1999). By scoring *Beginner Behavior* separately, instructors or researchers can readily identify the behaviors students may not understand as being ineffective or inefficient.

After selecting the model, we then deleted five items based on item level functioning at this stage. Modification indices indicated that three items (32, 39, and 48) loaded with more than one of the theoretical subscales. Upon

TABLE 5 Summary of factor loadings in the final model

Factors	# of items	Standardized pattern coefficient	Range of standardized item pattern coefficients
Recognize	6	0.90	0.43–0.71
Seek	6	0.92	0.42–0.65
Evaluate	7	0.91	0.53–0.71
Apply	5	0.90	0.55–0.70
Document	4	0.95	0.45–0.66
Reflect	4	0.74	0.62–0.79
Beginner Behavior	8	N/A	0.30–0.55

Note: The reverse-scored *Beginner Behavior* is a factor separate from the higher-order information literacy factor as described in the text.
Abbreviation: N/A, not applicable.

TABLE 6 mCFA measurement invariance across gender groups

Model	*df*	x^2	CFI	RMSEA	$\Delta\chi^2$	Δdf	*p*
Model 1. Configural							
Equal factor pattern	1,458	3,087*	0.84	0.05			
Model 2. Metric							
Equal loadings	1,496	3,122**	0.84	0.05	32	35	.07
Model 3. Scalar							
Equal intercepts	1,528	3,199**	0.83	0.05	73	77	<.001
Model 4. Error							
Equal error variance	1,568	3,278**	0.83	0.05	62	79	<.001
Model 5. Means							
Equal latent means	1,576	3,287**	0.83	0.05	12	4	.40

Note: $n = 764$ (382 female and 382 male respondents).
Abbreviations: CFI, comparative fit index; mCFA, multigroup confirmatory factor analysis; RMSEA, root mean square error approximation; $\Delta\chi^2$, nested chi-square difference.
*$p < .01$; **$p < .001$. The last *p* column is the testing of changes in the fit from the previous model to the current model, with additional constraints.

reviewing the items, we decided to remove these items as they did not clearly reflect only one construct. Two items (19 and 24) were highly redundant with other items in their factor, both in wording and in their contributions to fit. In addition, modification indices pointed out items with similarities, and, thus, we allowed their error residuals to correlate (4, 6, and 44; 7 and 8). Figure 2, shows the final CFA model along with all loadings. Table 5 lists the standardized pattern coefficient for each factor with SIL and the range of standardized pattern coefficients within each factor. Standardized pattern coefficients, sometimes referred to as standardized loadings, reflect the amount of variance represented in the correlation matrix being analyzed where the factors are extracted (Thompson, 2004). They are useful for evaluating how well each item is contributing to the factor and how well the first-order factors belong in the second-order factor. The factors of *Recognize*, *Seek*, *Evaluate*, *Apply*, *Document*, and *Reflect* strongly load onto a higher-order factor of *SIL* as shown in Figure 2 and Table 5, with coefficients ranging from 0.74 to 0.95. The item pattern coefficients ranged from 0.30 to 0.79, indicating all items adequately load on their conceptual factor and contribute to the measurement of SIL.

5.3 | Results of measurement invariance

While there are numerous goodness-of-fit indices used to interpret measurement invariance tests, according to Byrne (2012), the most relevant are degrees of freedom, chi-square, CFI, root mean square error (RMSEA), the comparison of changes in chi-square and degrees of freedom in models with increasing constraints, and *p*-values of whether changes in chi-square are significant.

TABLE 7 mCFA measurement invariance test results by language groups

Model	df	x^2	CFI	RMSEA	$\Delta\chi^2$	Δdf	p
Configural							
Equal factor pattern	1,458	2,691*	0.76	0.05			
Metric							
Equal loadings	1,496	2,725**	0.76	0.05	34	38	.01
Scalar							
Equal intercepts	1,528	2,786**	0.76	0.05	61	32	.01
Error							
Equal error variance	1,568	2,923**	0.74	0.05	137	40	<.001
Means							
Equal latent means	1,576	3,005**	0.72	0.05	87	8	<.001

Note: n = 330 (165 English language learners and 165 randomly selected experienced with English language).
Abbreviations: CFI, comparative fit index; mCFA, multigroup confirmatory factor analysis; RMSEA, root mean square error approximation; $\Delta\chi^2$, nested chi-square difference.
*$p < .01$; **$p < .001$. The last p column is the testing of changes in the fit from the previous model to the current model, with additional constraints.

5.3.1 | Invariance across gender

Table 6 displays the results of the multigroup confirmatory factor analysis (mCFA) measurement invariance testing across gender groups. The configural model is the least restrictive (Model 1) as it tests whether the overall factor pattern is the same between groups. The configural model had a chi-square value of 3,087 with 1,458° of freedom. The metric (Model 2) forced the groups to have the same factor pattern and factor loadings. The chi-square value changed to 3,122 with 1,496° of freedom, increasing the chi-square value by 32 and degrees of freedom by 38. The change is insignificant, $p > .05$. However, additional constraints of the scalar (i.e., equal intercept) ($\Delta\chi^2[32] = 73$, $p < .001$) and error (i.e., equal error variance) models ($\Delta\chi^2[40] = 62$, $p = .003$) did significantly change model fit. However, the effect as shown by Cramer's V (Cohen, 1987) is small ($V_{\text{scalar}} = 0.21$, $V_{\text{error}} = 0.20$). Requiring the factor means to be equal across groups did not significantly change the model. Taken together, the results did not find meaningful measurement differences between male and female scores.

5.3.2 | Invariance across language groups

As shown in Table 7, the results of mCFA testing measurement invariance between groups based on their experience with English language instruction show statistically significant changes with the addition of each parameter constraint. The configural model (Model 1) has a less than ideal fit, with a CFI of 0.76 and an RMSEA of 0.05, indicating that item functioning may differ across groups based on language experience. Furthermore, each additionally constrained model is significantly different from the previous model, and the most constrained model has a CFI value of 0.72 and RMSEA of 0.05. From these results, caution should be taken when comparing the mean abilities between the language groups.

5.4 | Comparison of scores

To aid in interpretation of factors where there are differing numbers of items, factors were scored based on the factor mean response. In other words, item responses were summed according to factor and then divided by the number of items in that factor to compare factor scores. Table 8 presents the means, standard deviations for the entire sample, subgroups of male and female, and the results of the t-test. Based on an alpha threshold of 0.05, the t-test results indicate that on average, females scored significantly higher on *Document* scores, $M = 5.03$, $SD = 0.86$, than males, $M = 4.90$, $SD = 0.87$; of small effect $d = 0.15$. There were no other significant differences.

TABLE 8 Mean comparison of factor scores between gender groups

Factor	*M* (*SD*)	M_{male} (*SD*)	M_{female} (*SD*)	*t*	*p*
Recognize	5.30 (0.81)	5.27 (0.80)	5.33 (0.82)	−1.16	.17
Seek	4.77 (0.86)	4.77 (0.85)	4.77 (0.86)	0.02	.49
Evaluate	5.07 (0.86)	5.06 (0.87)	5.08 (0.86)	−0.32	.39
Apply	5.59 (0.74)	5.55 (0.73)	5.63 (0.75)	−1.78	.09
Document	4.97 (0.87)	4.90 (0.87)	5.03 (0.86)	−2.43	.02
Reflect	5.10 (1.03)	5.05 (1.04)	5.14 (1.02)	−1.22	.11
Beginner Behavior	4.21 (0.83)	4.25 (0.85)	4.17 (0.81)	1.64	.09
All items	4.95 (0.57)	4.93 (0.58)	4.96 (0.57)	−1.10	.27

TABLE 9 Mean comparison of factor scores between language learner groups

Factor	*M* (*SD*)	M_{EEL} (*SD*)	M_{ELL} (*SD*)	*t*	*p*	*d*
Recognize	5.35 (0.82)	5.35 (0.80)	5.36 (0.84)	−0.09	.47	
Seek	4.89 (0.85)	4.72 (0.88)	5.07 (0.79)	−4.48	<.01*	0.49
Evaluate	5.11(0.83)	5.10 (0.86)	5.13 (0.79)	−0.30	.40	
Apply	5.56 (0.78)	5.62 (0.74)	5.49 (0.82)	1.96	.06	
Document	4.99 (0.90)	4.94 (0.94)	5.03 (0.85)	−0.93	.20	
Reflect	5.18 (1.04)	5.08 (1.10)	5.28 (0.97)	−1.69	.04*	0.19
Beginner Behavior	3.94 (0.91)	4.14 (0.95)	3.73 (0.82)	4.66	<.01*	0.51
All items	4.93 (0.58)	4.94 (0.59)	4.92 (0.56)	0.35	.42	

Note: The effect size, shown as Cohen's *d*, was calculated for significant differences.
Abbreviations: EEL, experienced with English language; ELL, English language learners.
*Statistically significant at the .05 level.

Table 9 shows the results from the comparison based on experience with English language instruction. There are three factors with significant differences between groups: *Seek*, *Reflect*, and *Beginner Behavior*. The effect sizes were calculated to better understand the magnitude of the differences. *Reflect* had a small effect size of 0.19, whereas *Seek* differences had a moderate effect size of 0.49 and *Beginner Behavior* differences had an effect size of 0.51 (Cohen, 1987). Specifically, 69% of the EELs scored above the mean of ELL scores in the *Beginner Behavior* factor, which translates to a mean difference of 0.41 points on a 7-point scale.

5.5 | Discussion of Hypotheses 2 and 3

The results of the mCFA showed measurement invariance between the gender groups, and their mean scores were also not significantly different. These findings demonstrate evidence for fair and valid interpretation of SIL with males and females in first-year engineering courses.

We sought evidence of whether the factor structure of SIL remained the same when used to assess students who were new to instruction in the English language and whether these students rated their behavior similarly to those more experienced with instruction in the English language. As additional constraints were added in the mCFA, the goodness-of-fit indices decreased, indicating significant measurement differences in the measurement between groups. Based on these mCFA results, it is understood that some items may function differently for different groups of learners. Thus, caution should be applied when comparing scores between students who are more experienced with English and those who are newer to having instruction in the English language. Future research should consider the source of the differences in measurement models and whether students who are newer to the English language read and interpret the items in the same manner as students who are native to the language.

Similar to other researchers who have examined differences between groups based on language, we found the significant mean score differences between language groups in the *Seek*, *Reflect*, and *Beginner Behavior* factors (Lindwall et al., 2012; Taylor, Bagby, & Parker, 2003). However, differences in overall scores on SIL between language groups were not significant, indicating that the measurement differences found did not result in a lower score for either language group. Future research should consider the cultural differences in how learners have been educated in precollege settings to seek information and reflect on the information found, their habits of finding information, and the prevalence of using less effective search strategies.

6 | DISCUSSION

In this study, we provide validity evidence for the use of SIL to assess first-year engineering and technology students' self-directedness with information literacy by evaluating the structure of student responses and measurement invariance across groups of learners. The SIL scale was designed based on prior research in assessment of engineering and technology students' information literacy skills and behavior and their perceptions of self-directedness with information when engaged in engineering projects. In the development of SIL, we successfully controlled for students' tendency toward inflated self-report through three approaches: developing items that target specific information literacy strategies students report they use (Douglas, Van Epps, et al., 2015); moving to a 7-point scale to provide more granularity of responses; and including reverse-scored items that represent beginning undergraduate engineering students' information literacy behavior.

A higher-order *SIL* factor was supported through CFA, with evidence that *Recognize*, *Seek*, *Evaluate*, *Apply*, *Document*, and *Reflect* are distinct factors in students' self-directedness with information literacy applied to engineering projects. The *Beginner Behavior* items were not determined to be part of the *SIL* factor. This result was not unexpected, as reverse-scored items frequently do not load onto their conceptual factor (see, e.g., Tomas & Oliver, 1999). While *Beginner Behavior* items are administered as part of the SIL instrument, these findings support that *Beginner Behavior* items must be scored separately from the other SIL factors.

As students begin to engage in more advanced information literacy behaviors in their design projects, we would anticipate scores on the *Beginner Behavior* to begin to decrease. While students may self-report higher levels of SIL, the *Beginner Behavior* factor enables an additional piece of evidence regarding what level of novice strategies students report employing.

The factor structure of scores between males and females were found to have measurement invariance and, thus, support the use of SIL to compare SIL scores between these groups. While mean scores between language groups was not significant in this study sample, measurement variance was found, indicating there may be some differences in how students newer to the English language read and understood the items. Thus, future research should examine the source of that variance.

In conclusion, SIL can be scored based on each subfactor to assess engineering and technology students' specific SIL skills (*Recognize*, *Seek*, *Evaluate*, *Apply*, *Document*, and *Reflect*) or scored as an overall broader measure of self-directedness with information. A separate factor, *Beginner Behavior*, can be used to moderate students' tendency to report themselves as highly skilled. To calculate SIL, one could sum the scores for the six factors of SIL and then subtract the Beginner Behavior score; $\text{SDIL} = \text{SIL} - \text{BB}$. Based on scores, instructors and researchers could design interventions to increase students' information literacy.

7 | LIMITATIONS

Although SIL's factor structure and interpretation of means were tested for generalizability across gender and English language instruction groups, one limitation to the current study is that all data were collected from a single university. However, given the large sample size, high communalities of the items, high standardized pattern coefficients, and relative measurement invariance across genders, there is strong evidence to suggest SIL's factor structure will remain stable when used at other institutions with first-year students (see Comrey 1978 for a discussion of threats to generalizability of factor structures). Students in the sample were also predominantly White; however, sample restrictions did not allow for comparison based on race and/or ethnicity. Taking into account Pawley's (2017) admonishment that anyone calibrating assessment instruments on primarily White, male students should provide explanation for this choice, we regret that it was not possible to examine statistical differences based on race or ethnicity due to the low numbers. However, considering whether the measurement model fit for women and students without previous instruction in the English language is a step toward collecting evidence of the fairness of scores for assessing diverse groups' SIL. In future research, we hope to obtain evidence to determine whether all learners read and interpret SIL in a similar manner. Since we found experience with English language instruction affected measurement invariance, future research should also examine whether students of diverse ethnic, racial, and socioeconomic status understand SIL questions in a similar manner.

8 | CONCLUSIONS

In this study, evidence is provided to justify the intended use of SIL as an assessment of engineering and technology students' SIL for research, evaluation, and classroom purposes but in low-stakes situations with no consequences, such as grading. SIL is an aid for researchers, evaluators, and instructors to identify areas of information literacy where engineering students feel strong and areas where they feel they may need to be bolstered through curriculum interventions. Instructors can use SIL in conjunction with performance assessments of information literacy to provide concrete feedback to students regarding areas they perform well in compared to how well they rated themselves. Likewise, researchers can triangulate SIL in conjunction with other information literacy assessments to provide a more holistic understanding of students' learning needs and whether there is consistency between how they view their information behavior and what they actually do when information is needed.

We tested specific hypotheses to inform the intended use of SIL, but additional validity evidence could be examined for other uses. While we explored the invariance of the measurement model for students of different English language instructional experience and gender, future research may explore invariance based on major, race, and ethnicity. Additional future areas of research with this instrument could determine whether it could be used, for example, to determine baseline and post-intervention effects or pre/post changes in student perceptions about specific information literacy factors. As this research did not examine pre/post changes, it is unknown how sensitive SIL would be in capturing changes from an intervention. It could also be tested in tandem with performance-based assessments of student skill levels to examine statistical relationships between the two measures. In any case, SIL provides complementary evidence to performance-based measures of students' SIL skills and behavior.

Very few studies in engineering education consider potential bias or measurement differences between minority groups and dominant student groups (Douglas, Purzer, Rynearson, & Strobel, 2016). Prerequisite to interpreting mean group differences is testing the assumption that the assessment measures groups of learners in an equitable way. The findings here add to the scant research that questions whether all students in engineering are fairly assessed. Statistical validation of the engineering student population is, unfortunately, rather dominated by one large group. Thus, engineering education has much evidence about how well published assessments function for one group. What is less clear is the consequences to underrepresented engineering students when we blindly assume all learners read and understand the assessment in the same way. We argue that all assessments administered to engineering students should be examined for measurement bias. Both the *Standards for Educational and Psychological Testing* (AERA/APA/NCME, 2014) and NCME *Classroom Standards* (2015) are clear: quality assessment instruments show evidence of reliability, validity, and fairness. To realize a truly fair assessment of all engineering students, it is imperative that more studies consider if and under what conditions measurement differences between groups are found and how to create fair assessments for all students. Additionally, more methods must be determined for examining fairness when sample sizes of underrepresented minority students are too small for statistical models.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation TUES grant No. 1245998. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

ORCID

Kerrie A. Douglas https://orcid.org/0000-0002-2693-5272
Senay Purzer https://orcid.org/0000-0003-0784-6079

REFERENCES

ABET. (2020). *Criteria for accrediting engineering programs, 2020–2021*. Retrieved from https://www.abet.org/wp-content/uploads/2020/03/E001-20-21-EAC-Criteria-Mark-Up-11-24-19-Updated.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*, Washington, DC: American Educational Research Association.

Association of College and Research Librarians. (2000). *Information literacy competency standards for higher education*. Retrieved from https://alair.ala.org/bitstream/handle/11213/7668/ACRL%20Information%20Literacy%20Competency%20Standards%20for%20Higher%20Education.pdf?sequence=1&isAllowed=y

Atman, C. J., Adams, R. S., Cardella, M. E., Turns, J., Mosborg, S., & Saleem, J. (2007). Engineering design processes: A comparison of students and expert practitioners. *Journal of Engineering Education*, *96*(4), 359–379. http://doi.org/10.1002/j.2168-9830.2007.tb00945.x

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design Studies*, *20*(2), 131–152. https://doi.org/10.1016/S0142-694X(98)00031-3

Atman, C. J., Kilgore, D., & Mckenna, A. (2008). Characterizing design learning through the use of language: A mixed-methods study of engineering designers. *Journal of Engineering Education*, *97*(3), 309–326. http://doi.org/10.1002/j.2168-9830.2008.tb00918.x

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

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

Byrne, B. M. (2012). *Structural equation modeling with Mplus*. New York, NY: Routledge.

Cervero, R. M., Miller, J. D., & Dimmock, K. H. (1986). The formal and informal learning activities of practicing engineers. *Journal of Engineering Education*, *77*(2), 112–114.

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

Cohen, J. (1987). *Statistical power analysis for the behavioral sciences* (pp. 24–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology*, *46*(4), 648–659. http://dx.doi.org/10.1037/0022-006x.46.4.648

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281. http://doi.org/10.1016/b978-1-4832-0087-3.50007-3

Denick, D., Bhatt, J., & Layton, B. (2010). Citation analysis of engineering design reports for information literacy assessment. *Proceedings of the ASEE Annual Conference and Exposition*. Retrieved from https://peer.asee.org/16508

Douglas, K. A., Fernandez, T., Purzer, S., Fosmire, M., & Van Epps, A. (2015). A self-assessment instrument to assess engineering students' self-directedness with information literacy. *Proceedings of the IEEE Frontiers in Education Conference*. Retrieved from https://ieeexplore.ieee.org/abstract/document/7344250

Douglas, K. A., Purzer, S., Rynearson, A., & Strobel, J. (2016). Reliability, validity, and fairness: A content analysis of assessment instrument publications in major engineering education journals. *International Journal of Engineering Education*, *32*(5), 1960–1971.

Douglas, K. A., Rohan, C., Fosmire, M., Smith, C., Van Epps, A., & Purzer, S. (2014). 'I Just Google It': A qualitative study of information strategies in problem solving used by upper and lower level engineering students. *Proceedings of Frontiers in Education Conference*. https://doi.org/10.1109/fie.2014.7044298

Douglas, K. A., Van Epps, A. S., Mihalec-Adkins, B., Fosmire, M., & Purzer, S. (2015). A comparison of beginning and advanced engineering students' description of information skills. *Evidenced-Based Library and Information Practice*, *10*(2), 127–143. https://doi.org/10.18438/B8TK5Z

Douglas, K. A., Wertz, R. E. H., Fosmire, M., Purzer, S., & Van Epps, A. S. (2014). First-year and junior engineering students′ self-assessment of information literacy skills. *Proceedings of the ASEE Annual Conference & Exposition Conference*. Retrieved from https://peer.asee.org/20498

Dunning, D. (2011). Chapter 5: The Dunning–Kruger effect: On being ignorant of one's own ignorance. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). New York, NY: Academic Press.

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*, *79*(3), 272–299. http://doi.org/10.1037/1082-989X.4.3.272

Fisher, M., King, J., & Tague, G. (2001). Development of a self-directed learning readiness scale for nursing education. *Nurse Education Today*, *21*(7), 516–525. http://doi.org/10.1054/nedt.2001.0589

Fosmire, M., & Radcliffe, D. F. (2013). *Integrating information into the engineering design process*. West Lafayette, IN: Purdue University Press.

Gross, M., & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library & Information Science Research*, *29*(3), 332–353. http://doi.org/10.1016/j.lisr.2007.04.012

Gross, M., & Latham, D. (2011). What's skill got to do with it?: Information literacy skills and self-views of ability among first-year college students. *Journal of the American Society for Information Science and Technology*, *63*(3), 574–583. http://doi.org/10.1002/asi.21681

Guglielmino, L. M. (1977). Development of the self-directed learning readiness scale (Unpublished doctoral dissertation). University of Georgia, Athens, GA.

Hensel, R. A. M., Brown, O., & Strife, M. L. (2012). Engineering and information literacy program for first year engineering students. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/21292

Holland, M. P., & Powell, C. K. (1995). A longitudinal survey of the information seeking and use habits of some engineers. *College & Research Libraries*, *56*(1), 7–15. http://doi.org/10.5860/crl_56_01_7

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

Johnson, E. J. (1988). Expertise and decision under uncertainty: Perfomance and process. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 209–228). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jöreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. *ETS Research Bulletin Series*, *34*(2), 183–202. http://doi.org/10.1007/bf02289343

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

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

Kane, M. T. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198–211. http://dx.doi.org/10.1080/0969594x.2015.1060192

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. http://doi.org/10.1080/01621459.1995.10476572

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York, NY: Guilford Press.

Knowles, M. S. (1975). *Self-directed learning*. New York, NY: Associated Press.

Kwasitsu, L. (2004). Information-seeking behavior of design, process, and manufacturing engineers. *Library & Information Science Research*, *25*(4), 459–476. https://doi.org/10.1016/s0740-8188(03)00054-9

Lamont, G., Figueiredo, R., Mercer, K., Weaver, K., Jonahs, A., Love, H., … Al-Hammoud, R. (2020). *Information-seeking behavior among first-year engineering students and the impacts of pedagogical intervention*. Paper presented at American Society for Engineering Education 127th Annual Conference, Virtual.

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196–204. http://doi.org/10.1080/00223891.2011.645936

Litzinger, T. A., Wise, J. C., & Lee, S. H. (2005). Self-directed learning readiness among engineering undergraduate students. *Journal of Engineering Education*, *94*(2), 215–221.

MacAlpine, B., & Uddin, M. (2009). Integrating information literacy across the engineering design curriculum. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/5433

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts, and how-to's. *International Journal of Psychological Research*, *3*(1), 97–110. Retrieved from http://www.revistas.usb.edu.co/index.php/IJPR/article/view/854

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455.

Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(1), 60–72. http://doi.org/10.1207/S15328007SEM1101_5

Messick, S. (1987). Validity. *ETS Research Report Series*, *1987*(2), 1–208. http://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*(8), 463–469. https://doi.org/10.3102/0013189X07311660

Mosborg, S., Adams, R., Kim, R., Atman, C. J., Turns, J., & Cardella, M. (2005). Conceptions of the engineering design process: An expert study of advanced practicing professionals. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/14999

Mowbray, T., Boyle, C., & Jacobs, K. (2015). Impact of item orientation on the structural validity of self-report measures: An investigation using the TAI-G in an Australian sample. *Journal of Psychoeducational Assessment*, *33*(3), 278–290. http://doi.org/10.1177/0734282914548405

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). Scaling procedures: Issues and applications. *Statistical Medicine*, *23*(15), 2480–2481. http://doi.org/10.1002/sim.1813

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: Guilford Press.

Pawley, A. (2017). Shifting the default: The case for making diversity the expected condition for engineering education and making whiteness and menness visible. *Journal of Engineering Education*, *106*(4), 531–533. https://doi.org/10.1002/jee.20181

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. http://doi.org/10.18637/jss.v048.i02

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

Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2012). *Motivation in education: Theory, research, and applications*. Upper Saddle River, NJ: Pearson Higher Education.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464. http://doi.org/10.1214/aos/1176344136

Spector, P. E. (1992). *Summated rating scale construction*. Newbury Park, CA: SAGE Publications.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. http://doi.org/10.1207/s15327906mbr2502_4

Taylor, G. J., Bagby, R. M., & Parker, J. D. (2003). The 20-Item Toronto Alexithymia scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, *55*(3), 277–283. http://doi.org/10.2466/pr0.101.5.209-220

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, *56*(2), 197–208. http://doi.org/10.1177/0013164496056002001

Tomas, J. M., & Oliver, A. (1999). Rosenberg's self-esteem scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 84–98. http://doi.org/10.1080/10705519909540120

Welker, A. L., McCarthy, L. A., Komlos, J., & Fry, A. A. (2012). Information literacy: A field guide for the path of life-long learning. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/21532

Wertz, R. E. H., Purzer, Ş., Fosmire, M. J., & Cardella, M. E. (2013). Assessing information literacy skills demonstrated in an engineering design task. *Journal of Engineering Education*, *102*(4), 577–602. http://doi.org/10.1002/jee.20024

Wertz, R. E. H., Ross, M. C., Fosmire, M., Cardella, M. E., & Purzer, S. (2011). Do students gather information to inform design decisions? Assessment with an authentic design task in first-year engineering. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/17789

Williams, B., Blowers, P., & Goldberg, J. (2004). Integrating information literacy skills into engineering courses to produce lifelong learners. *Proceedings of the ASEE Annual Conference & Exposition*. Retrieved from https://peer.asee.org/13596

Williamson, S. N. (2007). Development of a self-rating scale of self-directed learning. *Nurse Researcher*, *14*(2), 66–83. http://doi.org/10.7748/nr2007.01.14.2.66.c6022

Yu, F., Sullivan, J., & Woodall, L. (2006). What can students' bibliographies tell us?—Evidence based information skills teaching for engineering students. *Evidence Based Library and Information Practice*, *1*(2), 12–22. http://doi.org/10.18438/b84p4q

AUTHOR BIOGRAPHIES

Kerrie A. Douglas is an Assistant Professor of Engineering Education at Purdue University, 516 Northwestern Avenue, West Lafayette, IN 47906; douglask@purdue.edu.

Todd Fernandez is a Lecturer in the Wallace H. Coulter Department of Biomedical Engineering at the Georgia Institute of Technology, 313 Ferst Drive, Atlanta, GA 30332; todd.fernandez@bme.gatech.edu.

Michael Fosmire is a Professor in the Libraries and School of Information Studies at Purdue University, 504 West State Street, West Lafayette, IN 47907; fosmire@purdue.edu.

Amy S. Van Epps is the Director of Sciences and Engineering Services in the Harvard College Library, Harvard Science Center, Cabot Science Library, 1 Oxford Street, Cambridge, MA 02138; amy_vanepps@harvard.edu.

Senay Purzer is a Professor of Engineering Education at Purdue University, 516 Northwestern Avenue, West Lafayette, IN 47906; purzer@purdue.edu.

How to cite this article: Douglas KA, Fernandez T, Fosmire M, Van Epps AS, Purzer S. Self-directed information literacy scale: A comprehensive validation study. *J Eng Educ*. 2020;109:685–703. https://doi.org/10.1002/jee.20355

APPENDIX

SELF-DIRECTED INFORMATION LITERACY

The purpose of this survey is to learn about your use of information during engineering projects. There are no right or wrong answers. The survey will take approximately 12 min to complete.

Please think about a recent engineering project you have completed (either in or out of class) and answer the following:

1. Briefly describe the recent engineering project you are thinking about (∼2 sentences).
2. What new information did you need to learn to complete the project? (1–2 sentences).
3. How did you go about acquiring the information needed? (1–2 sentences).

Based on the same recent engineering project you described, please read each statement and select your level of agreement on a 1 (not at all true of me) to a 7 (very true of me scale).

1—Not true at all of me	7—Very true of me
Apply	I used information to develop my ideas.
	I used the information I found to make important decisions.
	I used information from existing products in an ethical manner.
	I synthesized information from various sources effectively.
	I modified my plans based on new information I found.
Document	My references in the report clearly communicated how other products or ideas informed my work.
	I had a highly effective method for organizing my information sources.
	I was transparent about the role of similar products in informing my project.
	I knew exactly when to cite information in my report.
Evaluate	I confirmed the results presented in sources by finding additional sources.
	I assessed the accuracy of sources I found.
	I considered the purpose of each source when evaluating how to use the information it contained.
	I evaluated the accuracy of data I generated during the project.
	I evaluated the trustworthiness of the sources I found.
	I checked the publication date of my sources to make sure the information was current.
	I considered the intended audience of a source when evaluating the information it contained.
Recognize	I talked to others to gather a variety of perspectives about the project.
	I thought about what additional information could increase the value of my product.
	I identified what information I most needed to successfully complete the project.
	Early on in the project, I considered how a lack of information could inhibit my project.
	I constantly asked myself questions to clarify my understanding of the problem.
	I checked my understanding of the problem by talking with others.
Reflect	I considered how I could make sense of information more efficiently.
	I generated specific ideas to improve my information gathering process.
	I considered how I could better utilize information in future projects.
	I reflected on whether I had found enough information.
Seek	I consulted with others more knowledgeable than me to gather information.
	I used specialized sources to find technical information.
	I looked at the reference list of sources to locate other potential sources of information.
	I knew where to look for different kinds of information (e.g., patents, standards, or handbooks).
	I used key terms from one source to search for additional sources of information.
	I gathered information from potential end users of my product.

1—Not true at all of me	7—Very true of me
Beginner Behavior	I trusted organization ".org" sites to be unbiased.
	I expected the project instructions to explicitly state all the information necessary to complete the project.
	I only used information that confirmed my thinking about the project.
	I decided on one solution quickly to allow more time to finish the project.
	I started gathering information toward the end of the project.
	I stopped thinking about the project once it was complete.
	I found just the minimum number of required sources for the project.
	I found all the technical information I needed through a general search engine like Google.

Note: Complete final instrument can be found at https://purr.purdue.edu/publications/3050 or at
Fosmire, M., Douglas, K. A., Van Epps, A. S., Purzer, S., Fernandez, T. M. (2018). Self-Directed Information Literacy (SIL) Scale. Purdue University Research Repository. https://doi.org/10.4231/R790221G