WIP: Evaluating a Survey Tool to Measure the Longitudinal Development of Student's Perception and Confidence in Using Advanced Data Skills

Ophelia Winslett[1] owinslett3@gatech.edu
Laura Christian[1] laura.christian@gatech.edu
Carmen Thiemann[2] ccortes9@gatech.edu
Todd Fernandez[1] Todd.Fernandez@BME.GATech.edu

[1]Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology and Emory University
Atlanta, GA

[1]School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA

This research category work-in-progress paper addresses the continued need for high-quality tools to quantify learning in engineering education. In particular, learning about artificial intelligence and machine learning (AI/ML) and related data analysis in engineering. We developed a survey tool to track long-term changes in engineering students' perspectives on their confidence, interest, and perception of the career relevance of advanced data skills. This evaluates aspects of the validity and usefulness of the survey using data from two undergraduate engineering courses. The survey data was analyzed using a combination of descriptive statistics and typical tools for quantifying survey validity, including factor analysis. We found that one of the three attitudinal components measured exhibited significant ceiling effects and plan to revise it accordingly. The other two attitudinal components showed positive item-level behavior and appropriate levels of internal coherence via measures of item-item correlation and factor analysis. We identify minor revisions of wording to some instruments and refine the instrument overall. We also identify several approaches to scoring the items that may be useful to others. The instrument is available for other interested parties interested in participating in further development.

Keywords— machine learning, survey development, curricular change, psychometrics

I. Introduction

This work-in-progress research paper addresses the initial development and assessment of a quantitative instrument to measure engineering students' advanced data skills in the face of rapidly advancing artificial intelligence/machine learning (AI & ML). The use of AI & ML in engineering is increasing rapidly [1], [2]. This creates a need to integrate these topics, and those of data handling in general, into engineering education [2], [3]. Unsurprisingly, that adoption in engineering and education is occurring at a rapid pace [1], [2], [4], [5]. How institutions, programs, and courses have or are planning to do so, has varied widely – with consistent patterns still emerging.

To meet the need to develop students' knowledge and skills related to AI and ML tools, our department launched a curriculum-wide initiative to embed data skills in all required undergraduate courses. Each course is developing one or more interventions that involve data, ML, AI, and related concepts. The interventions enhance rather than replace existing course content by linking it to modern engineering practices. Curriculum wide, our goal is to provide broad exposure to how modern engineering uses data and modern analysis tools to drive decisions in different contexts. For example, in a physiology course, students train and test neural networks to diagnose Atrial Fibrillation [6]. The activity improves students' knowledge of cardiac physiology while also introducing them to tools that leverage that knowledge to engineer clinical decision-making. Similarly, an exercise in a design course tasks students with using ChatGPT to generate FDA design inputs and then analyze how the scope and quality of the output compares to design inputs they wrote unassisted. Though interventions vary extensively, the shared core is that students learn more by reshaping what we already teach to use data, AI, and ML. We see a similar variety in work across the field to introduce AI and ML-related activities into engineering education [1]. Links to the survey and all course activities described in this paper are provided in Appendix A (Due to FIE systems find appendices here: https://shorturl.at/I4BRh).

A. Need for assessment and an instrument

Inherently, curricular changes create a need for effective assessment [7]. One challenge in doing so for this project is the lack of consistent and precise definitions of what we hope to achieve when we teach data skills or AI/ML to engineers [2]. This challenge is not abnormal, nor are calls for more high-quality instruments that can measure the complex concepts that are fundamental to engineering students' growth. However, what is unique to this push is the variety of approaches being taken and the variety of goals being sought.

We are seeking to measure a diverse set of activities, which causes several measurement challenges. What is being taught varies, as does what is learned and the context it is taught in. Those features require us to identify measurable connections rather than surface-level construct alignment. Further, because interventions occur across the curriculum, it is important to have the range to track longitudinal growth while also having the precision to evaluate the impact of specific interventions. Finally, tools of the AI & ML trade are still rapidly evolving, meaning the content being taught may change as well and evoke new challenges [8], [9]. In parallel, little work has documented transferable or generalizable measures for tracking the effect of introducing AI & ML into curricula generally.

To address these challenges, we divide the assessment of AI/ML activities into two parts. One is assessing learning at the intervention level, drawing data from how students' complete tasks in each intervention. We plan future work to develop rubrics and tools to score learning on a set of generalized data skills. The second important part of assessment is tracking the longitudinal change of attitudinal outcomes. We see this as an important aspect of assessing the implementation of AI & ML into our curriculum. Specifically, we measure students' interest and their perception of relevance of the interventions to their pursuit of biomedical engineering careers. We identified a generalized quantitative self-report instrument that is consistent across interventions and curricula as the most appropriate assessment method. A first draft of the instrument is described

2025 IEEE Frontiers in Education Conference (FIE) | 979-8-3315-0105-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/FIE63693.2025.11328291

and evaluated in the remainder of our paper. The instrument is in ongoing development, and others interested in using or assisting in the development should contact the senior author.

B. Development and definition of measured construct

As described above, the instrument measures the attitudinal effects of the implementation of AI & ML activities in undergraduate engineering courses. The specific construct we intend the instrument to measure occurs at the intersection of students' attitudes and our interventions.

Our department developed a list of *data skills* to articulate things that we believe students should achieve through our undergraduate curriculum (see Appendix **Error! Reference source not found.**). As with ABET programmatic outcomes, the skills are high-level. The list of data skills drew on our faculty's expertise in AI/ML research fields and a review of literature on teaching AI & ML. An initial list was refined through a process of internal and external review. Notably, that revision process was responsible for the move from a list of *AI & ML skills* (early draft) to *data skills* (final draft). Our position is that emerging AI & ML techniques are tools for doing data-driven engineering better. That situation makes the effective use of AI & ML tools inseparable from the fact that an application of any specific tool or algorithm is inseparable from the data it is trained on or used on. This makes attention to the analytic pipeline (i.e., *data skills*) important to ensuring rigor in students' learning.

We used a similar process to identify attitudinal aspects of our construct. A review of instruments in engineering education identified confidence as a measure frequently paired with measurements of skill. Methodologies such as confidence-weighted scoring add insights to correct and incorrect responses [10], [11]. However, we see confidence as incomplete given how AI & ML technologies affect professional identity [12]–[14]. Studies show some students questioning the shift in nature and conception of work that results from workplace AI & ML. Therefore, we added two secondary components to fully capture the potential attitudinal effects of implementing AI & ML into our curriculum: (1) Students' sense of the applicability of these data skills to their work as biomedical engineers as well as (2) students' interest in careers that would require them to use these skills. We refer to confidence, personal interest, and applicability as the attitudinal *context* of the items.

In sum, our instrument measures attitudes towards advanced AI & ML data skills. That construct manifests as a combination of two components: Technical + Attitudinal. For example, students' <u>confidence in their ability</u> *(attitudinal component)* to <u>use data to solve engineering problems in biomedical engineering and medicine</u> *(data skill 4)*.

C. Purpose of this paper

This paper reports an initial assessment of the structure and usefulness of our newly developed instrument and its pool of items. To do this, we use psychometric techniques including descriptive statistics and exploratory factor analysis. We specifically address the following research questions:

RQ1: Do the distributional properties of individual items support their use as part of a latent measurement scale?

RQ2: Do the items show interrelatedness with each other and latent factors in a way that supports interpreting the instrument as a coherent construct(s)?

RQ3: What modifications to the items and instrument are necessary to support effective distribution and scoring to measure change across multiple curricular interventions?

Because this paper focuses on psychometric analyses, it is important to note that such analyses are a valuable but incomplete component of the validity of measurements. That is why we present this paper as a work-in-progress. The research questions do not support a generalized claim that the inferences from the instrument are valid or useful. Rather, they present evidence of structural validity that is helpful for broad use.

II. METHODS

A. Survey and survey development

Survey development used a multistage process like other instruments in engineering education research. The initial development focuses on two goals: (1) Ensuring sufficient measurement range for a four-year curriculum and (2) covering the broad mix of construct-relevant skills and activities in specific classes. Those challenges drove item and instrument development, including the decision to maintain a large item pool and be less sensitive to item-level ceiling and floor effects in our analysis.

The items were drafted from the list of data skills described earlier (Appendix B). The faculty authors drafted items using a typical brainstorming-style ideation process that was informed in part by drafts of data skills activities being developed for classes in our program. About 50 draft items were developed at this stage. The draft items were then organized by data skill(s). The groups of items were then refined by adding and combining items to balance relatedness with broad coverage, in line with the construct definition above. We did not attempt to minimize or limit items at this stage. Finally, we edited the wording of each draft item individually by repeatedly reading it silently and aloud to consider potential alternative interpretations. We plan to engage in think-aloud testing of the revised instrument in future work.

The final instrument contains 26 items, each with a four-point ordered categorical response scale. An example item, including the text, response categories, and data skill it aligns with, appears in figure 1. Due to concerns about survey length, we decided to limit the number of items each participant responds to. We identified 12 'anchor' items that represented broad representations of the data skills and were presented to all participants. Each participant is then presented with 3 of the remaining 14 items randomly. The 26 items were the same for each item context (e.g., confidence or personal interest) with different instructions.

Item #: 12
Data Skill: (4) Use data to solve engineering problems in biology and medicine
Text: Draw conclusions from the results of data analysis
Response categories:
Confidence: None, a little, some, a great deal
Personal interest: None, a little, some, a great deal
Applicability: None, a little, some, a great deal

Figure 1: Example item, linked data skill, and response categories.

B. Population, Data Collection, and Sample Size

Data was collected from two junior-level undergraduate physiology courses in the Georgia Tech Biomedical Engineering Department. Both courses implemented data skills activities as part of our broader change effort. The first course covers systems physiology (e.g., nerve and renal systems) and the activity related to cardiac measurement and diagnostics [6]. The second course covers cell level physiology and implemented an activity where students predict protein folding using AlphaFold, which is presented at this conference. Surveys were administered in a pre-post design using an electronic survey tool [15]. Students who completed both the pre- and post-survey received extra credit in the course, regardless of consenting to study participation. Students had the option to write a written reflection about what they learned in the data skills activity for the same amount of credit.

1) Sample Size

After data cleaning (see Appendix C), we had a total of 720 survey responses with 333 matched pre-post responses. In Fall 2024, we collected data from 1 course with 124 students and included 119 pre responses, 102 post responses, and 98 matched pre-post responses. In Spring 2025, we collected data from 2 courses with 271 students and received 264 pre, 235 post, and 135 matched responses. Due to the timing of students taking courses, students in the second course in the Spring may have taken the first course in the Fall. We expect to explore longitudinal aspects in future research.

C. Analytic methods

We employed a combination of descriptive statistics, correlation analysis, and exploratory factor analysis (EFA) to address our research questions. After descriptive statistics, we prioritized combining data from all item-contexts because they use the same item text, each analysis step is repeated with the item contexts separated in the appendices. Throughout analysis, we used techniques tolerant of missing data, because most items are seen only by a portion of participants. Correlation analyses used pair-wise calculation to address missingness. EFA used multiple imputation, which is described further below.

1) Descriptive Statistics

Descriptive statistics assess the location, variance, and distributional properties of responses at the item and item-context (i.e., construct) level. Our primary interest was ensuring each item and item-context had a meaningful measurement range and rational behavior. These metrics ensure that our data was appropriate for more complex analysis [16]–[18].

For each item, we calculated kurtosis, skewness, median, variance, and mean. We did the same for each item-context (e.g., confidence, applicability, and personal interest). We considered nonnormality severe if skew was greater than 2 and kurtosis greater than 7, matching guidelines in literature [19]. We considered items as having abnormal center location if the mean and/or median response was in the top or bottom response category. We note, and discuss in the results, that analysis of the ‘applicability’ item context stopped at descriptive statistics due to measurement ceiling effects.

2) Correlation Analysis

We used correlation analysis to analyze the level of interrelatedness of items with each other. Evaluating interrelatedness enables us to assess whether the items measure distinct skills while remaining sufficiently related to provide construct coverage [16]. Item-item relationships for the personal interest and confidence item-contexts were analyzed separately and in aggregate. Because some items are randomly presented, the sample size of each correlation item-pair varies. Pairwise Pearson correlation coefficients were computed among all item pairs [16], [18]. Pearson correlations are sufficient for correlation analysis of well-behaved ordinal scales in survey development [16]. We also calculated the minimum, maximum, and mean correlation from the item-item correlations in each item-context.

3) Exploratory Factor Analysis

EFA assesses the relatedness of items to an emergent latent construct. It is useful for assessing how items are related as well as refinement of an instrument. Specifically, we employed multiple imputation EFA using R and the mifa package [20]–[23]. Using multiple imputation was necessary to address the missing data that resulted from not showing all items to all participants. The mifa package generates a combined covariance matrix from the multiple imputations. We performed factor analysis on both the combined matrix and the individual imputations to evaluate the stability of results. We determined the optimal number of factors by performing parallel analysis on the combined matrix.

Our factor analysis used oblique promax rotation and polychoric correlations to account for the ordinal nature of the item responses. We evaluated the proportion of variance, cumulative variance, and average loading to evaluate overall model fit. We also used individual item loading with .4 as the threshold of concern for low loading and cross-loading.

III. Results

A. Descriptive Statistics

Table 1 displays a descriptive summary for each item-context. The confidence and personal interest item-contexts perform acceptably. Notably, the applicability item-context shows multiple points of concern – including mean, median, skew, and kurtosis. A review of items showed 87% of all responses were 3 or 4 (See Figure 2). Due to the ceiling effects, we dropped applicability from further statistical analysis.

Table 1: Applicability, confidence, and personal interest average kurtosis, skewness, median, and mean for all pre- and post-data.

Statistic	Applicability	Confidence	Personal Interest
Mean	**3.41**	2.68	2.94
	2.99-3.71	*1.99-3.27*	*2.53-3.42*
Median[1]	**3.63**	2.79	**3.01**
	3-4	*2-3*	*2.5-**4***
Skew	-0.99	-0.1	-0.47
	***-2.34** - -0.39*	*-.56-.65*	*-1.08-.01*
Kurtosis	3.27	2.55	2.49
	*2.07-**8.56***	*2.03-3.38*	*1.9-3.49*

Note: italicized values represent range of statistics for each parameter and item context. Values of concern are bolded.
[1]We report the interpolated median.

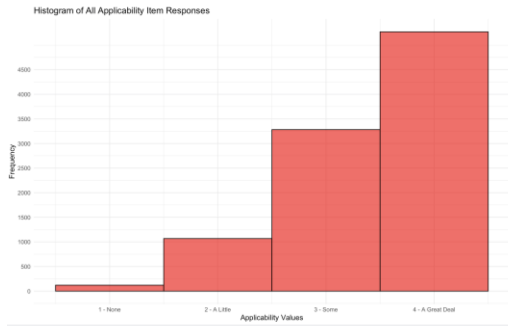

Figure 22: Histogram of all applicability responses by response bin.

B. Correlation

Figure 3 is a correlation plot of the item-item relationships for the combined confidence and personal interest item-contexts. Correlation results (Table 2) show reasonable levels of correlation. The average correlation coefficient (.48) is reasonable for proceeding with factor analysis, as is the range (.17 - .87). Specifically, we see the absence of any negative correlations as an important indicator of relatedness. Similarly, the lack of any item-item correlations above .9 indicates low concern about item redundancy. The confidence context had the broadest range of correlation values between items. Separating personal interest and confidence contexts did not significantly alter the correlation results (see Appendix D).

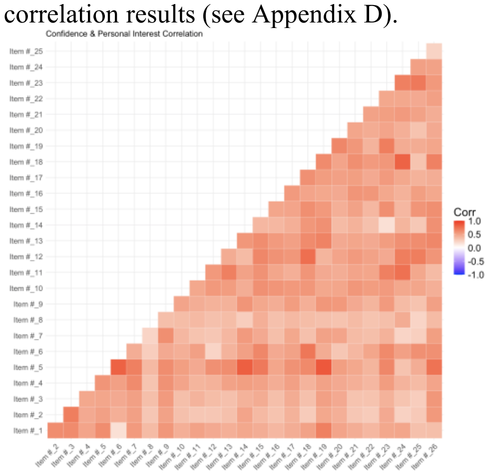

Figure 33: Correlation plot for all personal interest and confidence pre- and post-responses. X and Y categories are item number.

Table 2: The minimum, maximum, and mean for correlation plots of confidence, personal interest, and combined item-contexts.

Correlation	Min.	Max	Mean
Confidence	-.23	.89	.44
Personal Interest	.04	.87	.48
Combined	.17	.86	.48

C. Parallel analysis

A scree plot illustrating the parallel analysis results appears in Figure 4. The combined imputation covariance matrix from the mifa package suggested 2 factors. When separated, one imputation suggested 1 factor, and four suggested 2 factors. Based on that, we decided to compare one, two, and three-factor solutions. Such comparison of factorization to fully evaluate model fit is a best practice in EFA [19]. Parallel analysis of the confidence and personal interest contexts separately gave similar results (see Appendix E).

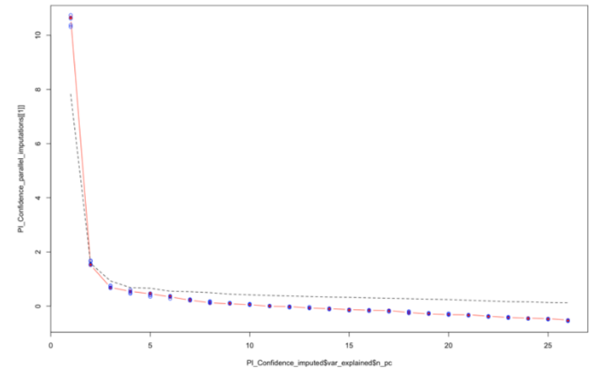

Figure 44: Scree plot of combined item-contexts. Black small dash line shows random data. Red dashed line shows results from combined imputation. Blue points show imputation by imputation variation.

D. Exploratory Factor analysis

In testing, we confirmed that a two-factor model provided the best balance of model fit and interpretability as predicted by parallel analysis. When model-context was combined, the two-factor model explains 47% of variance (Table 3). The two factors are well balanced in terms of variance explained and number of items. They also have higher average loading. The first factor consisted solely of items written for data skill 4. This includes items across the process of analytics from data identification to presentation. The second factor primarily contained items from skills 1, 2, and 3, related to implementing tools and analysis algorithms, especially in software code. We see these groupings as rational and construct-relevant. We note that two items that loaded well in the one-factor solution did not load above .4 on either of factors in the two-factor. Adding a third factor did not meaningfully increase variance explained, induces crossloading, and decreases average item loading for all factors. During analysis, we removed one item and identified several changes that are described in the discussion.

Table 3: Personal interest and confidence combined EFA analysis

Factor	1:1	2:1	2:2	3:1	3:2	3:3
Avg. Loading	.62	.66	.67	.62	.56	.54
Loaded Items	26	14	10	11	7	6
Item concerns	0	3 low		1 low, 2 crossload		
Prop. Var.	40%	26%	21%	24%	15%	11%
Cum. Var	40%	26%	47%	24%	40%	50%
df	299	274		250		
χ^2/df ratio	4.62	2.89		2.28		

A full table showing the item loading from the one and two-factor solution appears in Appendix F. Loading of most items was within the range of reasonable measurement, especially for a pilot instrument. When treated as 1 factor, item loading ranged from .44 to .7 with an average of .63, which is reasonable during instrument development [19], [24]. Given that we measured a small subset of the curriculum across which we expect growth, we see this loading as sufficient and allowing for greater measurement range. We also see the lack of items loading above .9 as an indicator that we do not have redundant items – which was a concern for instrument length.

IV. Discussion and Future Work

The National Academies have noted the workforce implications of AI tools and state "workers [must] navigate the changing landscape of job opportunities and demand shifts for different skills." [1]. Work to support the engineering education required by that shift comes with a need for assessment tools. In this paper, we present initial results from one such work-in-progress instrument that is relatively independent of course or curricular interventions. Our results support continued development of the instrument which we intend to pursue.

A. *Instrument performance*

The factors observed support the overall relationship between our construct definition and item behavior. The instrument has reasonable psychometric behavior as a single factor (evidence of overall construct alignment). Further, psychometric properties improve, and alignment with our construct map (i.e., data skills, Appendix B) improves when the number of factors is increased. Those factors establish a logical breakdown of specific data skills where items align to latent factors in two coherent groups (i.e., (1) *data analysis as a process* and (2) *application of analysis tools*).

B. *Item modifications*

Based on our results, we made three adjustments to the instrument. First, we dropped the 'applicability' item-context. Our results showed significant ceiling effects for this context. We suspect most students generally view AI/ML as so relevant to engineering practice that measuring level within that is difficult. Second, we removed one item (*differentiate between statistical and machine learning approaches to data analysis)*. That item had hire randomness than other items and did not factor well. Rereading the item, we see it referring to a differentiation that is unlikely to have a clear distinction with students or experts, making it less useful for measurement. Third, we reduced the number and changed the assignment of 'anchor' items based on our results. We originally had 12 anchor items that every student answered, and the rest were randomized. However, after the EFA analysis, we modified the anchor items to include only two items with the highest loading from each factor. This gives more randomized measurements while keeping survey length reasonable – increasing the measurement coverage across a course sample.

C. *Future work*

With these results showing the structure of the instrument functions as expected, we are planning future research to expand the argument for validity and usability of the instrument across broader context, especially modeling longitudinal change. As our interventions expand, they will include testing across a broader range of student levels and course types. We expect to confirm the properties established here as well as to refine methods of scoring to maximize the usefulness to faculty and researchers. We are interested in partnering in this work with others performing similar curricular projects related to data skills, AI, or ML. Anyone interested in using the instrument and participating in such testing and development should contact the senior author.

V. References

[1] Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective *et al.*, *Data Science for Undergraduates*. Washington, D.C.: National Academies Press, 2018.

[2] B. Allen, A. S. McGough, and M. Devlin, "Toward a framework for teaching Artificial Intelligence to a Higher Education audience," *ACM Trans. Comput. Educ.*, vol. 22, no. 2, pp. 1–29, Jun. 2022.

[3] Y. Fan *et al.*, "Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance," *Br. J. Educ. Technol.*, Dec. 2024.

[4] H. J. Warraich, T. Tazbaz, and R. M. Califf, "FDA perspective on the regulation of artificial intelligence in health care and biomedicine," *JAMA*, vol. 333, no. 3, pp. 241–247, Jan. 2025.

[5] L. M. F. Bertoline, A. N. Lima, J. E. Krieger, and S. K. Teixeira, "Before and after AlphaFold2: An overview of protein structure prediction," *Front. Bioinform.*, vol. 3, p. 1120370, Feb. 2023.

[6] Laura Christian Ophelia Winslett Alpa Gautam Todd Fernandez, "Developing an AI/ML activity for a BME physiology course," in *Proceedings of the 2025 ASEE Annual Conference and Exposition*, Montreal, Canada.

[7] K. A. Douglas and Ş. Purzer, "Validity: Meaning and relevancy in assessment for engineering education research," *J. Eng. Educ.*, vol. 104, no. 2, pp. 108–118, Apr. 2015.

[8] J. Angwin, J. Larson, L. Kirchner, and S. Mattu, "Machine Bias," *#creator*, 23-May-2016. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [Accessed: 21-Feb-2025].

[9] N. Sambasivan, S. Kapania, H. Highfill, D. Akrong, P. Paritosh, and L. M. Aroyo, "'Everyone wants to do the model work, not the data work': Data Cascades in High-Stakes AI," in *2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan, 2021.

[10] R. Ebel, Confidence weighting test reliability. *Journal of Educational Measurement*, vol. 2, pp. 49–57, Jun. 1965.

[11] S. S. Jacobs, "Confidence-Weighting as a Scoring Technique," 1970.

[12] J. Faulconbridge, A. Sarwar, and M. Spring, "How professionals adapt to artificial intelligence: The role of intertwined boundary work," *J. Manag. Stud.*, vol. 62, no. 5, pp. 1991–2024, Jul. 2025.

[13] L. Tredinnick, "Artificial intelligence and professional roles," *Bus. Inf. Rev.*, vol. 34, no. 1, pp. 37–41, Mar. 2017.

[14] L. Lombi and E. Rossero, "How artificial intelligence is reshaping the autonomy and boundary work of radiologists. A qualitative study," *Sociol. Health Illn.*, vol. 46, no. 2, pp. 200–218, Feb. 2024.

[15] "Qualtrics." [Online]. Available: https://www.qualtrics.com/.

[16] F. P. Holgado-Tello, S. Chacón-Moscoso, I. Barbero-García, and E. Vila-Abad, "Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables," *Qual. Quant.*, vol. 44, no. 1, pp. 153–166, Jan. 2010.

[17] J. Baglin, "Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR," *Practical Assessment, Research and Evaluation*, vol. 19, no. 5, pp. 1–14, May 2014.

[18] A. Gadermann, M. Guhn, and B. Zumbo, "Estimating ordinal reliability for likert-type and ordinal item response data: A conceptual, empirical, and practical guide," *Practical Assessment, Research and Evaluation*, vol. 17, no. 3, pp. 1–13, 2012.

[19] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, and E. J. Strahan, "Evaluating the use of exploratory factor analysis in psychological research," *Psychol. Methods*, vol. 4, no. 3, pp. 272–299, Sep. 1999.

[20] D. Shi, T. Lee, A. J. Fairchild, and A. Maydeu-Olivares, "Fitting ordinal factor analysis models with missing data: A comparison between pairwise deletion and multiple imputation," *Educ. Psychol. Meas.*, vol. 80, no. 1, pp. 41–66, Feb. 2020.

[21] V. Nassiri, A. Lovik, G. Molenberghs, and G. Verbeke, "On using multiple imputation for exploratory factor analysis of incomplete data," *Behav. Res. Methods*, vol. 50, no. 2, pp. 501–517, Apr. 2018.

[22] U. Lorenzo-Seva and V. G. J. R., "Multiple imputation of missing values in exploratory factor analysis of multidimensional scales: estimating latent trait scores," *Anales de Psicología/Annals of Psychology*, vol. 32, no. 2, pp. 596–608, 2016.

[23] "MIFA - multiple imputation for factor analysis." [Online]. Available: https://cran.r-project.org/web/packages/mifa/readme/README.html.

[24] M. Shevlin and J. N. V. Miles, "Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis," *Pers. Individ. Dif.*, vol. 25, no. 1, pp. 85–90, Jul. 1998.

[25] K. A. Douglas, T. Fernandez, M. Fosmire, A. S. Van Epps, and S. Purzer, "Self‐directed information literacy scale: A comprehensive validation study," *Journal of Engineering Education*, vol. 109, no. 4, pp. 685–703, Oct. 2020.

[26] K. A. Douglas, T. Fernandez, Ş. Purzer, M. Fosmire, and A. Van Epps, "A self-assessment instrument to assess engineering students' self-directedness in information literacy," in *2015 IEEE Frontiers in Education Conference (FIE)*, 2015.