ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #37265

Exploring the Exploratory Factor Analysis: Comparisons and Insights from Applying Five Procedures to Determining EFA Item Retention

Joseph Mirabelli

Joseph Mirabelli is an Educational Psychology graduate student at the University of Illinois Urbana-Champaign with a focus in Engineering Education. His work focuses on mentorship, mental health, and retention for STEM students and faculty. He was awarded the 2020 NAGAP Gold Award for Graduate Education Research to study engineering faculty perceptions of graduate student well-being and attrition. Before studying education at UIUC, Joseph earned an MS degree in Physics from Indiana University in Bloomington and a BS in Engineering Physics at UIUC.

Karin Jensen

Karin Jensen, Ph.D. is a Teaching Associate Professor in bioengineering at the University of Illinois Urbana-Champaign. Her research interests include student mental health and wellness, engineering student career pathways, and engagement of engineering faculty in engineering education research. She was awarded a CAREER award from the National Science Foundation for her research on undergraduate mental health in engineering programs. Before joining UIUC she completed a post-doctoral fellowship at Sanofi Oncology in Cambridge, MA. She earned a bachelor's degree in biological engineering from Cornell University and a Ph.D. in biomedical engineering from the University of Virginia.

Sara Rose Vohra

Sara Vohra is an undergraduate studying Bioengineering with a minor in Chemistry at the University of Illinois at Urbana-Champaign. Her interests lie in education as well as medicine with a future career goal as a physician.

Eileen Johnson (Research Associate)

Eileen Johnson received her bachelor's and MS in bioengineering from the University of Illinois at Urbana-Champaign. She previously worked in tissue engineering and genetic engineering throughout her education. During her undergraduate career, she worked with Dr. Brendan Harley developing biomaterial implants for craniomaxillofacial defects and injuries. In graduate school, she worked with Dr. Pablo Perez-Pinera working on new genetic engineering tools. There, she became interested in engineering education after helping develop and teach an online only laboratory class. She currently works as a research associate under Dr. Karin Jensen with a focus on engineering student mental health, retention, and development of resources.

Exploring the Exploratory Factor Analysis: Comparisons and Insights from Applying Five Procedures to Determining EFA Item Retention

Introduction

This theory paper considers standards in the use of Exploratory Factor Analysis (EFA) in engineering education research. EFA is a commonly used method across many social sciences disciplines, including education, political science, psychology, and marketing [1]. The goal of the technique is to reduce an amount of data, such as a list of survey items, to a more parsimonious form, such as a small number of factors which the survey items describe in bulk [1], [2]. These factors which summarize a larger number of items are called latent factors. In engineering education research, the technique is frequently and powerfully applied to the development and validation of novel quantitative scales, with some recent examples from the field including measures of students' responses to instruction [3], quality of experiences with internships [4], and reported degrees of engineering identity [5]. Despite the use of factor analysis techniques for over a century in social sciences research [1], consistent sets of standards between and within disciplines often differ [6], [7], and disagreement between researchers regarding analysis decisions in EFA has long affected the method's use [8].

Qualitatively, the mathematical procedure of an EFA can be imagined by a geometrical approach [9] which treats each individual item as a basis of vectors. Thus, a survey of N items forms an abstract N-dimensional basis, starting with an assumption that each item is independent. An EFA calculation performs a rotation and transformation of these vectors to find the optimal projection of that basis to a lower-dimensional basis, whose (eigen)vectors are the latent factors [9]. This projection accomplishes a means by which the same data can be described in a more parsimonious way.

Performing an EFA requires many decisions, starting with determining that the EFA method itself is an appropriate method of study for the data and research questions at hand (e.g., versus a similar option such as principal components analysis [9]). Decision points during the implementation of an EFA include a choice of procedures to determine whether to retain or remove individual survey items, a choice of procedures to determine the finalized number of latent factors, and a choice of which items are associated with the factors, e.g., when to end the analysis [1], [9]. In this paper, the authors explore decision-making during the item retention process; however, we will briefly review the entire process of an EFA to elaborate on the number of decisions required during the method and justify the decisions made in our methods below.

For measure development, a procedure for conducting an EFA on a novel set of survey items follows these steps: first, the researcher must determine the type of factor extraction method to be used, i.e., which estimation method (such as a maximum likelihood or least squares technique) should be followed to produce the factors [2], [9]. Next, the researcher must determine the type of rotational method (e.g., oblique, orthogonal) that is most appropriate for the data [2], [9]. In general, interpretation of the type of data used – including missingness, normality, and expectations about the characteristics of factors (e.g., should the factors be independent or correlated?) – will guide these first two decisions, as will the choice of software

being used for analysis [1], [2]. Next, the researcher must determine how many factors to include in the analysis, requiring a choice of which method (e.g., Kaiser's rule, scree test) is appropriate to make that determination [10]. Then the researcher can use their chosen software package to conduct the analysis and view the results. The EFA software will generally produce a list of the measured items' correlations, also called loadings, onto each latent factor. Items must then be decided to be retained in the scale or removed. For example, an item whose loadings (correlations) with all of the factors are low would be likely to be removed. We will refer to the choice of removing or keeping an item based on its loadings as item retention.

Item retention decisions require thought about how many items should be retained and at what values items should be considered to be of high quality. For example, a poor-quality item may "under-load" if its highest loading on any single factor is less than a certain cutoff value, or it may "cross-load" if its highest loadings on any two factors are higher than a certain cutoff value. However, there is not consensus on which cutoff values should be used [6]. Once items are chosen to remain based on cutoffs, the EFA becomes an iterative process between conducting an individual EFA calculation, determining how many factors to include and which items are retained, and modifying the item and factor numbers, all to conduct an EFA calculation again and repeat this process until all item loadings are decided to be satisfactory. Between each step, the researcher must decide if any items which load poorly on factors should be retained regardless of loading (e.g., an item with high face validity to one of the factors might load near a cutoff threshold), and thus deciding when to stop the process is also a significant choice [9].

As much of our team are novices to both the craft and the science of EFA, we negotiated a process by which we would conduct EFA procedures on our data multiple times and determine what seemed to be the most valid and trustworthy result between these procedures. Among the EFA decision points described above, we felt that decisions regarding *the order of which items to retain* would most greatly affect our factor structure, the number of factors and assignment of items to each factor. Thus, we performed simultaneous EFA procedures with different rules for determining item retention and compared their different results. We present the results of these different methods and our reflections upon how our analysis decisions modified the results.

To the knowledge of the authors, there is not a set of consistent guidelines for approaches to EFA available for engineering education researchers. Such a set of guidelines may be instrumental for new researchers, or to establish baselines for trustworthiness in the field. While we acknowledge that the exercise of judgement by researchers is important and breaking "rules" such as established cutoff values in appropriate situations is valid, a common language around which these judgements are made and presented in research (e.g., to choose to retain an item despite it being below a threshold) could support new researchers' understanding and use of the EFA method, promote trustworthiness and validity in researchers' approaches to EFA and measurement design, and provide review criteria for engineering education referees and authors. As a team whose members are relatively inexperienced social sciences researchers, we believe that we would have been personally benefited by such guidelines.

Research Questions

The research questions guiding this study represent an ongoing process experienced by the authors while navigating the methods of a commonly used technique in engineering education research and are open questions within and outside of the field:

RQ1: How do different approaches to item retention algorithms affect the results of EFA procedures for measure development?

RQ2: Are different algorithmic approaches to conducting EFA procedures during measure development more or less valid?

Researcher Positionality

Our research methods were strongly influenced by our prior experiences and our perspectives as both researchers and learners during this project were important to our decision-making process, as our learning is itself an exploration of EFA as a method. Considering the amount of decision points and interpretations involved in EFA research, the team agrees with recent calls in the field for engineering education scholars to include positionality statements in quantitative work (e.g., see [11-13]). The project team is composed of one undergraduate student in an engineering program, one research scientist who recently earned a Master's degree in an engineering program and is an incoming engineering education research PhD student, a doctoral student in educational psychology with an engineering background, and a faculty member in an engineering program whose research is in engineering education. Two of the four members of the research team were conducting an EFA procedure for the first time during the study, and a third member had limited classroom experience with the method. All four members were trained in engineering before transitioning to engineering education research.

The team was consistently encouraged to carefully and methodically explore the EFA method. The faculty member, graduate student, and undergraduate student designed and administered a pilot of the survey measure, leveraging their past experiences as engineering students and instructors during the drafting of the measure's items. The survey measure was also influenced by results from prior literature and interviews with undergraduate students [14], [15]. All members of the project team contributed to the analysis and interpretation of the data, including the EFA analysis presented here, as well as a descriptive analysis and analysis of the latent factors by various demographic groups. Given the relatively large number of items in the scale used for this research and the possibility for multiple valid procedures to produce different results, the team expressed concern across multiple discussions for finding a result which was "the most valid." The faculty member leading the project encouraged the team to explore the data thoroughly by adopting different approaches and consulted the other team members multiple times during and after the analysis. The graduate student designed the five method conditions. The first two methods were performed by the undergraduate student and research scientist, and the last three methods were performed by the graduate student. In multiple separate meetings, the project team compared the results of the first four methods, designed the democratic selection process as described below, and interpreted the results of the final method.

Methods

To address the research questions, our team performed five similar EFA procedures with different item retention strategies while in the process of validating a novel survey related to the culture surrounding stress as experienced by engineering undergraduate students. The research was conducted at a large, Midwestern university with a prominent college of engineering. The research design and draft measure were approved by the site university's Institutional Review Board before data collection began. A sample of $N = 624$ undergraduate, engineering degree-seeking students completed the survey. Descriptions of the sample are published [15].

Descriptions of the development of the 81-item measure of engineering stress culture have also been described previously [14], [15]. We used the R statistical programming language as software for all analyses [16]. An understanding of the survey items or goals of the survey is not a prerequisite for understanding how the different EFA procedures yield different results. However, it may improve a reader's understanding to see the methods in more detail. Appendix A provides a subset of the items. Appendix B names the ten factors produced by the final result of the five EFA procedures used. In future analyses, two of these ten factors were combined [15].

The project advisory board met during the design and early analysis stages of the study, and advice from the advisory board improved the quality of the survey design. Advice on conducting EFA analyses was solicited from advisory board members and the faculty team member multiple times during the research by the student team members, reflecting the number of decision points present in EFA procedures.

While five different item retention strategies were employed for each analysis method presented below, all other EFA decisions were made identically, following the suggestions found in multiple references (e.g., [1], [2], [9]) for each step identified above. For the extraction method, ordinary least squares was determined to be the method for each EFA calculation, as inspection of the skew and kurtosis values of the data suggested severe non-normality. Based on the design of the survey, we anticipated that all latent factors would be part of a larger culture of stress in engineering [14]. Therefore, these factors would potentially be related. Thus, we used an oblique (oblimin) rotation to allow for correlation between factors. To determine the number of factors, a scree plot was generated, and both parallel analysis and optimal coordinates techniques were used. Following guidelines from the literature [7], the optimal coordinates technique, which produces a more conservative number of factors, was trusted over parallel analysis if the two methods suggested different values. The cutoff for under-loading items was also decided from a review of the literature and determined to be .32 [9]. For cross-loading items, an item was considered to correlate too strongly with two factors if its second highest factor loading was greater than .30 [9].

At the beginning of analysis, four procedures for the order of item retention were decided on by the team. Later, a fifth procedure was used.

In the first method, under-loading items were removed first, and then cross-loading items were removed after all under-loading items had been removed. Items were removed one-by-one. Thus, at each step of iteration through the EFA procedure, only the under-loading item with the lowest loading was removed. To reiterate, an item was considered to be under-loading if its strongest correlation with any latent factor was .32 or less. If two items' strongest correlations with any latent factor were under-loading, the one with the overall lowest loading would be removed first. For example, if two items under-loaded, with one item's strongest correlation with any latent factor having a strength of .31 and another item's highest correlation being .29, the second item with a weaker correlation would be removed.

To determine the retention of cross-loading items, the second-highest loadings of items were compared. Items were considered to cross-load if they loaded onto two factors with a correlational strength of .30 or higher on each. The cross-loading item with the highest factor loading among all cross-loading items' second-highest factor loading was removed. For example, a cross-loading item might load strongly onto two factors, say with correlation values of .51 on one factor and .33 on another. A second cross-loading item might have loadings of .46

and .38. The second item would be removed because its second-highest factor loading (.38 vs .33) was highest. If removing a cross-loading item was to cause a new item to under-load, the under-loading items were removed before any additional cross-loading items were removed.

At times, identical loadings were found in cases where two items had the same lowest under-loading value or second-highest cross-loading value. In these cases, the researcher attempted two branching paths in which each item was deleted. Generally, it was found that if two items tied for the lowest loading, either branching path yielded both items being deleted in sequence.

In a second method, cross-loading items were removed first, and then under-loading items were removed, i.e., using criteria identical to the first method and iterating by removing one item at a time, testing for the number of factors, and repeating the EFA calculation.

In the third method, five under-loading items were iteratively removed, then one cross-loading item was removed. This process repeated until fewer than five under-loading items or no cross-loading items were present; in which case, the method followed the rules of the first method. A five to one ratio was chosen based on the approximate ratio of under- and cross-loading items.

In all methods, after each item was deleted, a scree plot analysis was once again performed to determine if the total number of factors changed. Thus, a number of factors was determined, an EFA was performed, an item was deleted, and this algorithmic procedure repeated iteratively until there were no cross-loading or under-loading items remaining.

In a fourth method, the third method was repeated with the number of factors fixed to 10. The team decided to implement this analysis method after the first three analyses were concluded. This fourth method followed the iterative rules of the third method but did not recreate scree plots at each step. The decision to conduct this method was made upon reviewing the prior three methods and determining that their number of factors averaged to ten and that the first two methods overlapped in terms of results with the third, e.g., the third was qualitatively and methodologically “between” the first two methods.

For the final method, once the four methods were complete, a democratically selected item deletion method was performed, where items that were not retained in at least three of the four methods were removed from the 81-item survey, and a new EFA procedure was conducted with the reduced item list and used to generate a final result. To the knowledge of the authors, this process has not previously been recommended in EFA literature, however this we determined that this method was both intuitive and a way to conservatively correct for an individual method’s erroneous removal of an item.

After items were removed, the fifth and final EFA method was conducted on the items that had not been removed due to the democratic selection process, using the rules from the third method: iteration between five under-loading items and one cross-loading item. The third method was chosen for this final procedure because the third method produced results which overlapped with both the first and second method and was implemented in the fourth method as well. The final results were tabulated, and identities were given to the final latent factors based on the questions

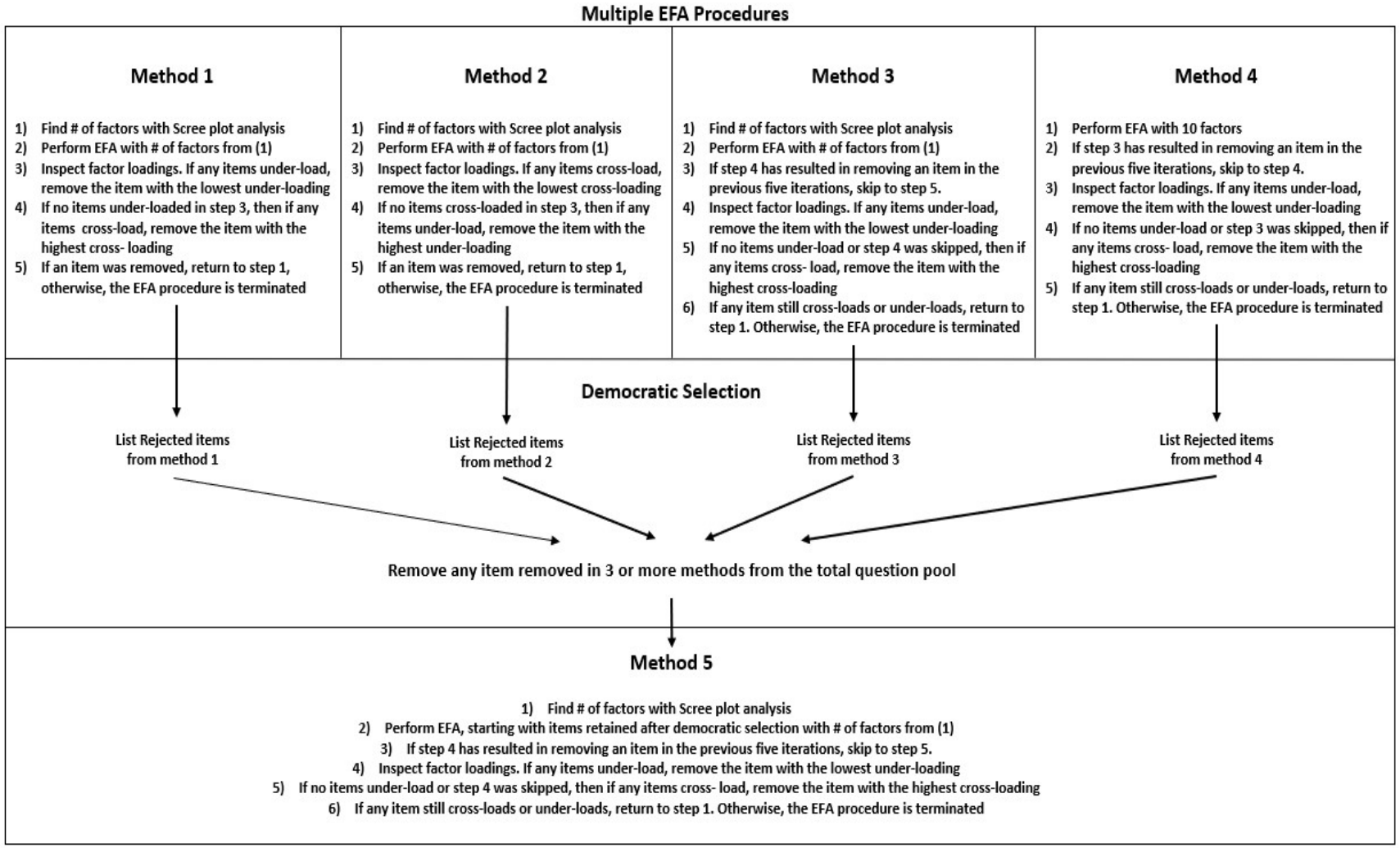

Figure 1. Visual Interpretation of Multiple EFA Procedures.

within their structure. However, the interpretation of the latent factors in the context of engineering stress culture is beyond the scope of this paper and will be reported by the authors in a future report. Figure 1 shows the different steps and order of steps performed to summarize these multiple EFA procedures.

Results

Here we summarize the five EFA methods implemented in this analysis. Table 1 provides a list of the methods used, the total number of items retained in each, and total factors present in each. Table 2 lists the 81 individual items and if they were retained (accepted) or deleted (rejected) during the democratic selection process. Finally, Table 3 shows the results of the EFA conducted during the final democratically selected set of items, including the 10 factors and 51 loadings.

Method 1: Under-loading item deletion first. In this method, all under-loading items were deleted, and then cross-loading items were deleted. A total number of 44 items were retained across 10 factors. While conducting this method, the researchers noted that there were many instances of ties between items and attempted several branching paths, which resulted in no change to the overall list of items deleted. The total number of factors changed slowly during the first 10-20 iterations, and then accelerated, changing with every three to five iterations near the end of the analysis. Usually, the count of factors, which started at 14, decreased, however the count occasionally increased when a cross-loading item was removed.

Method 2: Cross-loading item deletion first. In this method, all cross-loading items were deleted first, and then under-loading items were deleted. A total number of 50 items were retained across 11 factors. The researchers noted that because of the small number of cross-loading items, this method generally became methodologically equivalent to the first method, however, the number of factors to include decreased quicker than in Method 1, which caused the factor structure to be "shaken up" after a few iterations, at which point the results of Method 2 swiftly diverged from Method 1. As a result, in addition to a different number of factors between the two methods, a small number of items from Method 1's strongest factors were coupled with a different set of items in Method 2's factor structure.

Method 3: Under-loading and cross-loading iterations. In this method, five under-loading items were removed, one cross-loading item was removed, and then that procedure was repeated iteratively. In total, 49 items were retained across 10 factors. We noticed that while implementing this method, the removal of one cross-loading item often reduced the total number of factors, presumably by weakening the factor structure of two latent factors from the previous iteration and causing the weaker of the two latent factors to be removed from the sample.

Method 4: Fixed factors. In this method, the number of factors was fixed to 10, and otherwise the rules from Method 3 were followed. A total of 52 items were retained across the 10 factors. Because the purpose of this paper is to discuss the iterative order of determining item retention, this method is not the focus of this paper, but contributed to the democratic selection process. However, it is notable that this method produced different results from Method 3 for 23 of the 81 items and from Method 1 for only 11 of the 81 items, despite following the rules of Method 3.

Method 5: Democratically selected items; iteration method repeated. This method followed the rules of Method 3 after removing any items removed by at least three of the first four methods. The final number of items retained was 51, in a structure containing 10 factors.

Table 1. Summary of Methods

Method Number	1	2	3	4	5
Number of Factors	10	11	10	10	10
Number of Items	44	50	49	52	51

Table 2. Results of the Democratic Selection Process for Item Retention Methods

Item*	Method 1	Method 2	Method 3	Method 4	Item	Method 1	Method 2	Method 3	Method 4
1	Accept	Accept	Accept	Accept	42	Reject	Reject	Reject	Reject
2	Accept	Reject	Accept	Accept	43	Reject	Accept	Reject	Reject
3	Accept	Accept	Accept	Accept	44	Reject	Accept	Accept	Accept
4	Reject	Accept	Reject	Accept	45	Accept	Accept	Accept	Accept
5	Reject	Reject	Accept	Reject	46	Reject	Reject	Reject	Accept
6	Accept	Accept	Accept	Accept	47	Reject	Reject	Reject	Reject
7	Reject	Reject	Accept	Reject	48	Reject	Reject	Accept	Accept
8	Reject	Reject	Accept	Reject	49	Accept	Accept	Accept	Accept
9	Accept	Accept	Accept	Accept	50	Accept	Accept	Accept	Accept
10	Reject	Reject	Reject	Reject	51	Accept	Accept	Accept	Accept
11	Reject	Accept	Reject	Reject	52	Accept	Accept	Accept	Accept
12	Reject	Accept	Accept	Reject	53	Accept	Accept	Accept	Accept
13	Accept	Accept	Accept	Accept	54	Reject	Reject	Reject	Reject
14	Accept	Accept	Reject	Accept	55	Reject	Reject	Reject	Reject
15	Reject	Reject	Reject	Accept	56	Accept	Accept	Accept	Accept
16	Reject	Reject	Accept	Reject	57	Reject	Reject	Accept	Accept
17	Accept	Accept	Accept	Accept	58	Accept	Reject	Accept	Accept
18	Reject	Reject	Reject	Reject	59	Accept	Accept	Accept	Accept
19	Accept	Accept	Accept	Accept	60	Reject	Accept	Accept	Reject
20	Reject	Accept	Reject	Reject	61	Accept	Accept	Accept	Accept
21	Accept	Accept	Accept	Accept	62	Accept	Accept	Reject	Accept
22	Accept	Accept	Accept	Accept	63	Reject	Reject	Reject	Accept
23	Reject	Accept	Accept	Reject	64	Accept	Accept	Accept	Accept
24	Reject	Reject	Reject	Reject	65	Reject	Reject	Reject	Reject
25	Accept	Accept	Accept	Accept	66	Reject	Accept	Accept	Reject
26	Reject	Reject	Accept	Accept	67	Accept	Accept	Accept	Accept
27	Accept	Accept	Accept	Accept	68	Reject	Reject	Reject	Reject
28	Reject	Accept	Reject	Reject	69	Reject	Reject	Reject	Reject
29	Accept	Accept	Accept	Accept	70	Reject	Accept	Accept	Reject
30	Accept	Accept	Reject	Accept	71	Accept	Reject	Accept	Accept
31	Accept	Accept	Accept	Accept	72	Reject	Accept	Reject	Reject
32	Accept	Accept	Accept	Accept	73	Accept	Accept	Accept	Accept
33	Reject	Reject	Reject	Reject	74	Accept	Accept	Accept	Accept
34	Reject	Reject	Reject	Reject	75	Accept	Reject	Reject	Accept
35	Accept	Accept	Reject	Accept	76	Accept	Reject	Reject	Accept
36	Accept	Reject	Reject	Accept	77	Accept	Accept	Accept	Accept
37	Accept	Accept	Reject	Accept	78	Reject	Reject	Reject	Reject
38	Accept	Accept	Accept	Accept	79	Accept	Reject	Accept	Accept
39	Accept	Accept	Accept	Accept	80	Reject	Reject	Reject	Reject
40	Accept	Accept	Accept	Reject	81	Reject	Accept	Reject	Accept
41	Reject	Accept	Accept	Accept	-	-	-	-	-

* Shaded items (i.e., 5, 7, 8, etc.) were rejected by the democratic process, thus the items were not retained.

Table 3. Final EFA Results (Method 5)

Factor		1		2		3		4		5	
ID*	Loading	31	.822	56	.732	3	.711	1	.578	19	.748
ID	Loading	32	.844	58	.388	4	.412	6	.578	25	.800
ID	Loading	38	.363	59	.348	44	.354	9	.636	27	.519
ID	Loading	39	.425	60	.534	50	.509	13	.483		
ID	Loading	41	.417	61	.605	51	.658	29	.484		
ID	Loading	48	.613	64	.802	57	.413	40	.515		
ID	Loading	49	.396	67	.590	71	.541				
Factor		6		7		8		9		10	
ID	Loading	45	.836	26	.323	21	.695	14	.375	2	.364
ID	Loading	52	.762	73	.916	75	.509	30	.338	17	.439
ID	Loading	53	-.363	74	.720	76	.351	35	.460	22	.484
ID	Loading	62	.432	79	.351	77	.721	36	.626		
ID	Loading					81	.339	37	.519		

* ID represents the item number; Loading is that item's correlation to the factor it appears with

The democratic selection process omitted 26 items; thus, the fifth method began with 55 items, and iterations of that method only removed four items. It can be observed that for the democratic selection process, items were unanimously removed or accepted in only 44 out of 81 items. For the remaining 37 items, one or two methods demonstrated discrepancies in determining item retention. Of note, for 12 items, two of the four methods rejected the item and two kept the items, suggesting these items may have been more "borderline" items than other more accepted items. For example, item number 70 had two votes to be rejected and was kept by democratic selection, but was among the four items deleted during the EFA iterations on the democratically selected items in Method 5.

Among the items receiving only one "vote" to be rejected, item 35 had a correlation strength of .460 in the final democratic selection result, item 37 had a strength of .519, item 40 had strength .515, and item 41, .417. These moderately high correlations present in the final result may have been rejected if only one method would have been used.

Discrepancies across the methods. As noted in Method 4, there was a large degree of variation across the methods. Method 1 and Method 2 failed to agree to either reject or accept 21 items. Method 1 and Method 3 disagreed 22 times. Method 1 and Method 4 failed 11 times. Method 2 and Method 3 failed 23 times; and Method 2 and Method 4 failed 24 times, the largest discrepancy between any two methods. Finally, Method 3 and Method 4 failed to retain the same items 23 times. Each discrepancy total suggests the number of questions the two methods would not share in common. For example, if Methods 1 and 2 were to produce a finalized set of results by two independent researchers, despite originating from the same set of 81 items, 21 items would not be shared between the two methods, constituting a discrepancy of 25.9% of the entire set of questions. On average, the four methods had a discrepancy for 20.67 (25.5% of) items. Table 4 summarizes the discrepancy count, or the number of items not in common between two methods and the percentage of questions with a discrepancy. An additional discrepancy between

Method 2 and the other methods was the total number of factors; Method 2 produced 11 latent factors while the other methods had 10.

Table 4. Discrepancy Counts (and Percent Discrepancy) Between Methods

Method	1	2	3	4
1	–	*	*	*
2	21 (25.9%)	–	*	*
3	22 (27.2%)	23 (28.4%)	–	*
4	11 (13.6%)	24 (29.6%)	23 (28.4%)	–

*Items have the same scores across the diagonal

Discussion

RQ1: How do different approaches to item retention algorithms affect the results of EFA procedures for measure development?

In sum, we noted that deleting cross-loading items at different points of the analysis changed the factor structure throughout an iterative EFA process, suggesting that the inclusion or deletion of cross-loading items and the order of selecting items to be retained affects the results of an EFA procedure. From the results in Table 2 and the discrepancies between methods, it is apparent that for this data, iterating different orders of rules for item retention determines which questions remain in the sample and has major consequences on the composition of items in the final result, generally changing the result by 25% of questions for this survey. This has implications for the validity of EFA work – if our team had instead performed an analysis using only one of the first three methods, the results would differ greatly. For example, two researchers who used Method 1 and Method 2 not only would have a total of 21 items not in common, but would report a different number of latent factors, suggesting different structures to the results. In fact, the first three methods each suggest a different total number of factors, and each method results in a similar number of discrepancies between questions.

One might argue that the importance of conducting such an EFA process is to produce reliable subscales, and therefore any of the methods that achieve that are equally useful. However, having no firm metric for determining the better of the first three approaches to an analysis which are arguably all equally valid seems problematic. So, which is the better method?

RQ2: Are different algorithmic approaches to conducting EFA procedures during measure development more or less valid?

In some ways, our team feels that the democratic selection process presented here poses a more valid result than any other individual method, so long as the identity of a “proper” procedure is not formally established. The democratic process resulted in the average number of factors and more items retained than most of the other procedures, suggesting that by first democratically removing items which were more than likely to be rejected, the procedure was able to better fit the remaining stronger items to the data. It is likely that the items rejected by all of the methods were poor items (or at least not easily fit into the factor structure) and that the items accepted by all of the methods were strong items. Based on the final correlational strength of the items rejected by only one method, it’s likely that those items were largely worth saving and thus inversely the items with three votes to reject were also worth removing.

For example, as stated in the Results, items 35 and 37 both loaded onto factor 9 and had final correlations of .460 and .519 with that factor and was rejected only once in Method 3. In that five-item factor, four of the items were only rejected in Method 3. The text of those four items reads: “My engineering professors assign homework and projects that benefit my learning”, “Engineering professors are NOT approachable”, “Engineering professors promote student well-being”, and “My engineering professors offer welcoming and helpful office hours”. Item 36 was rejected by two methods but retained in the final method, its text reads “Engineering professors know my name and acknowledge me.” At face value, these five questions seem obviously related based on student perceptions of interactions with engineering faculty members. But it is possible that by removing one of these items too early due to cross-loading effects (item 36 was rejected by Methods 2 and 3), a factor with otherwise clear value might be mistakenly eliminated.

However, we feel some discomfort still in the retention or rejection of items based on a democratic procedure – a researcher with a strong bias towards a particular number of latent factors or the inclusion of a certain item or identity of a certain latent factor could manipulate the voting process, consciously or subconsciously, to achieve a desired result. Further, adding more decision points to the EFA process complicates a method already requiring researchers to make many judgements. Should it be expected that multiple EFA paths are tried for every approach at designing a measure? And can the validity of approaches like these even be measured?

Having answers to questions and observations like those above would benefit our team, but there are many considerations to be had for and against making decisions about standards for EFA methods. As one of the more widely used and foundational methods applied in our field’s quantitative research, there is merit for a focus on what practices may be preferred within the community. New researchers would be greatly benefited by having access to more resources which establish standards within the field, both in terms of conducting EFA and more broadly across methods. As we have suggested, learning the best practices of EFA methods from handbooks or papers involves drawing from the standards of other fields. Reviewers of journals in the engineering education would also be benefited by adoption of standards for analyses. However, given the large number of decisions in EFA methods, too stringent standards may limit the freedom of researchers to apply the proper techniques for fear of diverging from approved methods.

In this paper, the authors explored decision-making during the item retention process. However, other decisions, such as determining the initial number of factors, also would require standardization, making the process a major undertaking. Standards may also inhibit researchers from exploring research techniques in the way our team has in this report.

While the adoption of reliable standards in the field may make both the conduct and review of research easier, any standards adopted would require far less qualitative testing for validity than what is present in this discussion and would require carefully crafted evidence. Past efforts to improve factor analysis procedures have included large-scale simulation of data analyses using factor analysis. For example, Costello and Osborne’s simulations of factor retention yielded evidence that Kaiser’s rule too liberally includes factors and should not be used for factor selection [9]. Other Monte Carlo simulations of EFA decisions could provide best practices or trends for different analysis decisions. One limitation of simulated data is the lack of “art” in simulated EFA techniques, as latent factors should appear to measure a sensible phenomenon by having relationships between the topics of the questions, e.g., achieving face validity.

In the opinion of the authors, the issues above do not have clear answers, and more established researchers are likely to have strong, and likely differing, stances based on their experiences with EFA methods. Regardless, we are in support of further research and discussion on the use of common statistical techniques in our field and believe that by adopting more clear standards as a field, we grow in terms of both our trustworthiness and identity as a field.

Finally, we note that there are limitations to this type of analysis: most significantly, this analysis involved a single set of data whose mathematical idiosyncrasies may not be generalizable. The dataset used for this analysis was also large, involving 80+ items; many EFA-validated surveys are smaller, and may be less mathematically ambiguous. For groups considering this type of democratic selection process, the analysis requires a significant investment of time, as multiple EFA procedures, and documentation and analysis of those procedures must occur. It may be impractical or impossible for three researchers to work independently on EFA iterations. A further danger in producing and comparing multiple techniques is the potential for the introduction of unconscious bias, such as by researchers evaluating which results from performing several methods align best with their beliefs. Finally, we note that the choice of a democratic selection process represents a departure from existing literature, however it was performed in response to a deficit which is well-recorded in our review of EFA decision-making processes.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 1943541. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors thank our project's advisory board members Jennifer Cromley, Allison Godwin, and Nicola Sochacka for their support and advice on this project's analysis. The authors would also like to thank the reviewers of this submission for beginning the thoughtful conversation about EFA methods in our field. Finally, the authors thank the survey participants for their insights and contributions to our research.

References

[1] L. R. Fabrigar, and D. T. Wegener, "Exploratory factor analysis," *Oxford University Press*, 2012.

[2] A. G. Yong, and S. Pearce, "A beginner's guide to factor analysis: Focusing on exploratory factor analysis," *Tutorials in Quantitative Methods for Psychology*, vol. 9, pp. 79-94, 2013.

[3] M. DeMonbrun, C. J. Finelli, M. Prince, M. Borrego, P. Shekhar, C. Henderson, and C. Waters, "Creating an instrument to measure student response to instructional practices," *Journal of Engineering Education*, vol. 106, pp. 273-298, 2017.

[4] L. Y. Y. Luk and C. K. Y. Chan, "Adaptation and validation of the Work Experience Questionnaire for investigating engineering students' internship experience," *Journal of Engineering Education*, vol. 109, pp. 801– 820, 2020.

[5] A. Godwin, "The development of a measure of engineering identity," *in proceedings of the 2016 ASEE Annual Conference & Exposition,* New Orleans, Louisiana, USA, Jun. 2016.

[6] M. C. Howard. "A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve?" *International Journal of Human-Computer Interaction,* vol 32(1), pp. 51-62, 2016.

[7] M. W. Watkins. "Exploratory factor analysis: A guide to best practice," *Journal of Black Psychology,* vol 44(3), pp. 219-246, 2018.

[8] L. L. Thurstone, "Multiple-factor analysis; A development and expansion of The Vectors of Mind," *University of Chicago Press*, 1947.

[9] A. B. Costello, and J. Osborne, "Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis," *Practical Assessment, Research, and Evaluation*, vol. 10, ed. 7, 2005.

[10] W. F. Velicer and D. N. Jackson, "Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure," *Multivariate Behavioral Research*, vol. 25, ed. 1, pp. 1–28, 1990.

[11] A. Godwin, "Sitting in the tensions: Challenging whiteness in quantitative research," *Studies in Engineering Education*, vol. 1, ed. 1, pp. 78–82, 2020.

[12] D. S. Ozkan, D. P. Reeping, C. Hampton, and C. Edwards, "A critique of quantitative methodologies to yield critical quantitative methods in engineering education research (EER)," *presented at the 2021 Research in Engineering Education Symposium (REES AAEE 2021)*, Perth, WA, Aus, December 2021.

[13] C. Hampton, D. Reeping, and D. S. Ozkan, "Positionality statements in engineering education research: A look at the hand that guides the methodological tools, *Studies in Engineering Education*, vol. 1, ed. 2, pp. 126–141.

[14] K. Jensen, S. R. Vohra, J. F. Mirabelli, A. J. Kunze, I. Miller, and T. E. Romanchek, "CAREER: Supporting undergraduate mental health by building a culture of wellness in engineering," *presented at the 2021 ASEE Virtual Annual Conference & Exposition,* June 2021.

[15] K. Jensen, E. Johnson, J. F. Mirabelli, and S. R. Vohra, "CAREER: Characterizing undergraduate engineering students' experiences with mental health in engineering culture," *presented at the 2022 ASEE Annual Conference & Exposition*, Minneapolis, Minnesota, USA, June 2022.

[16] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing,* Vienna, Austria, 2021.

Appendix A

Sample subset of the first twenty survey items.

Prompt: To what extent do you agree with the following statements?

Strongly Disagree *Disagree* *Somewhat Disagree* *Somewhat Agree* *Agree* *Strongly Agree* *No basis for judgement*

1	Engineering courses are designed such that students are set up to get low grades
2	Engineering students in my college/department are relaxed
3	Engineering students can succeed while having a positive work-life balance
4	Engineering students will NOT succeed unless they are almost always working
5	My engineering college/department expects too much of its students
6	My engineering college/department wants me to fail or drop out
7	Engineering students in my department are overworked
8	My engineering professors are welcoming to students of all genders and races
9	My engineering professors are trying to weed out the "weak students"
10	My engineering college/department has a diverse population of students (e.g., in terms of gender, race, income, etc.)
11	The rank and prestige of my engineering college/department places pressure on me
12	My engineering college/department sets reasonable expectations for its students
13	My engineering college/department only cares about job percentages and retention rates
14	My engineering professors assign homework and projects that benefit my learning
15	My engineering professors encourage teamwork among peers
16	I have to compete against other engineering students (e.g., for grades, scholarships, recognition for my academic work)
17	I have felt pressure to stay up late in order to complete engineering work
18	Most engineering students lack social/professional skills (e.g., are "socially awkward")
19	Most engineering students learn concepts faster than me
20	I worry what my peers will think of me if I fail

Appendix B

Names of ten factors and sample items.

Factor	1: Identity-Related Engineering Interactions	2: Programmatic Mental Health Communication	3: Impact of Engineering Work on Wellness and Time Management	4: Expectations of Academic Stress	5: Feeling of Falling Behind Peers
Sample Items	Q41: My engineering classes have respectful environments to students from diverse backgrounds	Q59: My engineering program has negatively impacted my mental health	Q3: Engineering students can succeed while having a positive work-life balance	Q6: My engineering college/department wants me to fail or drop out	Q19: Most engineering students learn concepts faster than me
	Q49: I have experienced an engineering student being intimidating and/or dismissive	Q61: Mental health issues are NOT often discussed among engineering faculty or students	Q71: I have time for activities outside of classwork that support my professional development	Q29: Engineering professors expect students to compete against each other	Q25: Most engineering students are ahead of me in terms of STEM skills
Factor	**6: Pressures and Prestige of Engineering Majors**	**7: Preparation for Engineering Careers**	**8: Challenges of and Competition for Starting Careers**	**9: Experiences with Engineering Professors**	**10: Impact of Engineering Work on Relaxation and Rest**
Sample Items	Q45: Engineering students experience higher levels of stress compared to other majors on campus	Q26: I believe that other students in my engineering program can succeed	Q21: I worry about competing against my peers in the job market or for internships	Q30: Engineering professors are NOT approachable	Q2: Engineering students in my college/department are relaxed
	Q52: Engineering students do much more work compared to other majors at my institution	Q73: Students in my engineering college/department are able to find internships	Q81: I am not sure what life will be like once I enter the workforce	Q35: Engineering professors promote student well-being	Q22: I RARELY have to stay up late in order to complete engineering coursework