Exploratory Factor Analysis for a Measure of Fundamental Instructor Behaviors for Inclusion in Undergraduate Math Classes

Abby Sine
Penn State University

Nathanial Brown
Penn State University

Neil Hatfield
Penn State University

Katherine Lewis
Washington University

Lindsay Palmer
UMass Chan Medical School

Deja Workman
Penn State University

Caroline Ho
Penn State University

Opemipo Esan
Penn State University

A recent movement has developed to emphasize the importance of inclusive teaching practices in undergraduate STEM education. Mathematics courses have one of the largest impacts on students' persistence and success in STEM majors, making efforts to improve diversity and inclusion in these classes all the more important. In this paper, we discuss our initial steps to develop a measure of students' perceptions of instructor behaviors that promote inclusivity in their classrooms. This paper briefly discusses the initial development of items for our measure and focuses on our exploratory analysis of the latent factor structure and our process for defining these latent factors in the context of math instructors' actions.

Keywords: Mathematics, instructor behaviors, factor analysis, instrument development, inclusion

Introduction

Efforts to diversify STEM fields, both in academia and industry, will benefit from efforts to cultivate inclusive introductory STEM courses. Mathematics courses in particular are notorious for contributing to disparities as they are often required in order for STEM undergraduates to graduate (Ellis et al., 2016; Leyva et al., 2021, Battey et al., 2022). Multiple efforts to promote inclusive teaching practices in higher STEM education have tackled aspects such as curricula, classroom interpersonal relationships, and pedagogy. (Dewsbury & Brame, 2019; Hogan & Sathy, 2022; Inclusive STEM Teaching Project, 2023; Johnson, 2019; Kachani et al., 2020; Tanner, 2013). While these works highlight the importance of inclusion in STEM classrooms, we find little work done to identify and functionally measure aspects of inclusivity in post-secondary STEM classrooms. This report details our team's effort to develop an instrument that would capture students' perceptions of how their instructor's behaviors support or diminish inclusivity in undergraduate mathematics classrooms. For our work, we conceptualize inclusiveness as students' feeling that they belong in and can achieve success in the math class. Our current findings are part of a long-term instrument development project (a more detailed description is available in Brown et. al. (2023)). The aim of this project is to create an instrument that can be used primarily for research purposes in investigating relationships between students' feelings of inclusion and other student-level variables (sense of belonging, mathematical confidence, etc.) or for instructors' personal use in improving their teaching practice. Our instrument should not be used for institutional assessment of instructors as such an application is likely to introduce racial and gender bias, among other concerns. In this report, we present an abbreviated description of the exploratory process for developing the 28-item Fundamental Instructor Behaviors for Inclusion (FIBI) instrument followed by the details of the exploratory factor structure, our process of naming the latent factors, and concludes with a discussion of future work.

Methods and Data Collection

Instrument

Development of the FIBI instrument began with the creation of 79 items that started with the common stem "My mathematics instructor…". Each item required a response on a 5-point Likert scale (1-Strongly Disagree, 2-Disagree, 3-Neither agree nor disagree, 4-Agree, 5-Strongly Agree). We created these items based on students' responses during focus groups discussions of the following questions:

- What has your instructor done or said that conveyed a sense you can succeed in math?
- What have they done or said that conveyed a sense you can't succeed in math?
- What have they done or said that conveyed a sense you belong in math?
- What have they done or said that conveyed a sense you don't belong in math?

(Focus groups are more fully reported in Brown et. al. (2023).) In addition to these items, we included nine items from two published instruments for measuring teacher behaviors: the Teacher Communication Behavior Questionnaire (TCBQ; She & Fisher, 2002) and the Questionnaire on Teacher Interaction (QTI; Wubbels et al., 1991). We included those items which covered behaviors not specifically mentioned by our focus groups, but that we thought could also impact students' feelings of inclusion in their classroom. These items were modified to fit the sentence structure of our previous 79 items before being added to the instrument, bringing the total number of items to 88.

Data Collection

In the Fall 2021, we collected data using the 88 items as well as demographic questions at a large, predominantly white, research-intensive university in the Northeastern United States. We intentionally recruited students from introductory chemistry and biology courses but asked them to provide responses based on their current mathematics instructor. Our goal for this strategy was to reduce the students' potential fears/concerns that their mathematics instructor would have access to their responses and to encourage students to provide honest responses to the items. Participants in the study received extra credit (equivalent to 0.5% of final grade) in their biology or chemistry course for completing the study. A total of 1,276 students participated in the study.

Due to the predominantly white makeup of this university and our desire to center the voices of students underrepresented in STEM classrooms, namely women and persons excluded because of their ethnicity or race (PEERs; Asai, 2020), we implemented a secondary strategy to specifically recruit PEER participants from other universities via emails to university chapters of student organizations such as the National Society of Black Engineers (NSBE) and the Society for the Advancement of Chicanos/Hispanics and Native Americans in Science (SACNAS) as well as colleagues at other institutions. Participants in this wave of recruitment were incentivized with a raffle for Amazon gift cards (20 $50 prizes, five $100 prizes, and two $250 prizes). This recruitment led to an additional 149 responses, bringing the total number of participants to 1,425.

Following the data collection, we cleaned the data by removing participants who (a) did not consent to participate (n = 161), (b) did not complete all FIBI items (n = 72), (c) were not enrolled in a mathematics course in Fall 2021 (n = 260), and (d) any responses identified as duplicates via student emails (n = 62). Thus, 870 participants remained in our sample. While the majority of our sample identified as female (n = 551), we still had a lack of participation from students who identify as PEERs (21%; n = 185). Approximately 71% of participants (n = 614) identified with at least one of these two groups.

Methods

To conduct our exploratory factor analysis (EFA), we used two sets of software: SPSS (version 28) and R (version 4.2.0). Analysis of the data was completed using multiple factor extraction methods (principal components analysis and principal axes) as well as using both Pearson and polychoric correlations (Holgado–Tello et al., 2010) to investigate the impacts of these choices on our analyses and to understand the FIBI instrument's structure in more detail. Across these different design choices, we consistently used an oblique rotation method (promax) to allow for correlations between factors which we suspected from our prior work with student focus groups (Brown et. al., 2023). In addition to considering multiple methodological choices, we were also interested in any potential differences in the factor structure for the subsets of women, PEERs, and the combined subset of women and PEERs and thus, we repeated these analyses for each subset as well as the whole sample. Due to the limited sample size, analysis of the PEER subgroup failed to converge. However, the other three analyses produced generally similar results (similar factor structure, number of factors, primary loading of items, and factor correlations). Thus, only the results for the whole sample are reported here.

Our EFA relied on several heuristics in tandem to determine the appropriate number of latent factors to retain including (a) the Kaiser-Gutman criterion (eigenvalues greater than one), (b) parallel analysis (Horn, 1965), and (c) analysis of Scree plots (R. A. Johnson & Wichern, 2007). After using these heuristics to limit the number of factors to consider, we then eliminated any items which did not have a factor loading with a magnitude greater than 0.4 to any of the remaining factors (Stevens, 1992). Following these removals via benchmarks, we moved into more detailed item-to-item analysis where we investigated the Pearson and polychoric correlations between items with primary loadings on the same factor. In doing so, we identified clusters of items containing correlations with magnitude of 0.6 or higher and then selected representative items from each cluster to remain in the instrument while removing the others. (See Brown et. al., 2023) for our guiding principles for examining item clusters.) Our process of exploratory analysis and item reduction was iterative to ensure the stability of the measure. Thus, after each wave of item removal by item benchmarks and item-to-item analysis, EFA was rerun on the subset of items to ensure similar performance (factor structure—item assignment to factors—and cumulative variance explained) after the removal of those items.

Results

Exploratory Factor Analysis and Item Reduction

For our initial EFA containing all 88 items, the Kaiser-Gutman criterion suggested a 12-factor solution which explained approximately 63% of the total variance in responses while parallel analysis suggested a slightly less complex, 10-factor model which explained 56% of the total variance. Unfortunately, with such a large number of factors, our initial benchmark for removal (magnitude of the factor loading greater than 0.4) only flagged 9 items for removal. Due to our desire for a more parsimonious model, we considered the Scree plot of our factors, which suggested a model containing four to six factors, and thus we conservatively chose to include six factors. With this smaller subset of factors, our benchmark for factor loadings identified a total of 23 items for removal while still explaining approximately 55% of the total variance. After the removal of these items, we began analysis of item-to-item correlations and using our research team's guiding principles we identified another 23 items for removal, leaving 42 items left after our first wave of EFA and item reduction.

After completing the first wave of analysis, EFA was re-run to confirm that items remained in similar clusters and the number of factors was as we expected after removing those items. Scree plot analysis now suggested a seven-factor model. However, the seventh factor contained only a single item while the other six factors remained thematically intact from our previous analysis. Thus, the item was removed, and we retained the six-factor structure which still explained approximately 55% of the total variance. Again, employing the factor loading benchmark of 0.4, we identified 7 additional items for removal which no longer loaded significantly onto any of the factors. Item-to-item analysis was used again to identify six final items for elimination, resulting in a 28-item instrument.

Once more, to ensure our factor structure was stable, we conducted EFA on the subset of 28 items. Again, we arrived at the same six-factor model which explained 55% of the variation in responses. All items remained in their initial factors from the previous wave, except for the item, "My mathematics instructor talks too fast." Additionally, the item "My mathematics instructor gives exams which are hard to finish in the allotted time," now had a factor loading of 0.339, which did not meet the original benchmark of 0.4. However, our research team believed that the item provided student perceptions that were not equally addressed by any other items in the instrument and thus chose to retain the item. The details of the factor structure for the final 28-item instrument are discussed in the following section.

Resulting Exploratory Factor Structure

Following our iterative EFA process followed by item reduction, we arrived at a six-factor solution containing 28 items and explaining approximately 55% of the total variation in responses. A summary of the variation explained by each factor can be seen in Table 1. Factor labels (Factor 1, 2, …) were assigned in order based on the proportion of variance explained by the factor in the 28-item EFA. After reducing the instrument to 28 items, we see that Factor 1 explains the most variation in responses (12.7%) with Factor 2 explaining a similar amount of variation (12.6%). We note that Factor 1 and Factor 2 contain strictly negatively worded or positively worded items respectively, but both emphasize the irritation/willingness that instructors may exhibit when it comes to engaging with students and their questions which may have resulted in both factors explaining essentially identical amounts of variation.

Table 1: Summary of 6-Factor Solution on 28 Item Instrument

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
SS Loadings	3.55	3.52	2.74	1.95	1.92	1.59
Proportion of Variance Explained	0.127	0.126	0.098	0.070	0.069	0.057
Cumulative Variance Explained	0.127	0.253	0.351	0.421	0.490	0.547

Factor are labeled (Factor 1, 2, …) in order based on the proportion of variance explained in the 28 item EFA.

Unlike the other factors, Factor 3, which explained 9.8% of the variation in responses, was the only factor to contain both positively and negatively worded items. Item assignments as well as their factor loadings are provided in Table 2. For Factor 3, we can see that both negatively worded items associated with this factor, Q20 and Q21, have negative factor loadings while all other items have positive loadings, which is consistent with our expectation that negatively

Table 2: Primary Factor Loadings 6-Factor Solution

Question	Factor Loading	Question	Factor Loading
Factor 1: My math instructor is disrespectful to us sometimes.		Factor 4: My math instructor models the math-person archetype.	
Q28: Blames us when we don't understand something.	0.816	Q22: Gives overly complex or technical explanations.	0.680
Q5: Puts us down.	0.694	Q1: Teaches too fast.	0.601
Q6: Criticizes our answers to their questions.	0.761	Q12: Is too advanced to be teaching this course.	0.535
Q25: Smirks or laughs at our questions.	0.781	Q18: Gives exams which are hard to finish in the allotted time	0.339
Q26: Says the material is simple, easy, or obvious.	0.458	Factor 5: My math instructor doesn't prioritize teaching us.	
Q27: Seems annoyed, frustrated, or exasperated by our questions.	0.766	Q14: Doesn't manage class time well.	0.611
Factor 2: My math instructor is open and adaptable to us.		Q11: Makes mistakes which confuse students.	0.469
Q15: Makes sure all questions get answered before moving on.	0.831	Q13: Pays more attention to some students than others.	0.495
Q2: Explains things in different ways when we ask.	0.675	Q17: Takes too long to grade and/or return homework, quizzes and/or exams	0.534
Q3: Pauses to let us absorb material or formulate questions.	0.522	Factor 6: My math instructor does more than just lecture at us.	
Q9: Teaches to all students, not just those who've already seen this material.	0.562	Q19: Checks in on us when we work in groups.	0.758
Q10: Is helpful during office hours.	0.586	Q8: Encourages us to work together.	0.745
Q16: Makes us feel comfortable asking questions.	0.592	Q23: Provides supplementary materials like handouts or worksheets.	0.508
Factor 3: My math instructor isn't apathetic toward us.			
Q4: Jokes with the class.	0.858		
Q7: Talks enthusiastically about math.	0.699		
Q20: Has little or no enthusiasm for teaching.	-0.728		
Q21: Has little or no interest in students' success.	-0.410		
Q24: Engages in small talk with the class.	0.661		

Item labels (Q1, Q2,…) are the item order in the survey. The first item listed for each factor has the largest factor loading for the factor; subsequent items are listed in order by item label.

worded items would have a negative relationship with an overall positive factor. When naming the factors, we devised a positively framed name for this factor, incorporating the antithesis of Q20 and Q21 as themes within the factor. For example, we considered "an interest in student success" as a theme defining Factor 3 rather than the lack of interest.

Based on the item-factor associations presented in Table 2, our team took a collaborative approach to develop names to define our latent factors. To begin this process, each team member created their own name/phrase to describe each factor. Then, as a team, we discussed the proposed names/phrases, identifying consistencies between them and any potential weaknesses. Our aim was to create a unified name that could unite all items under a common theme rather than naming based on a few of the higher loading items. Thus, after creating a group name for a factor, each item was compared to the factor name to ensure that the name represented the item well. As we continued this process for multiple factors, our team decided to implement a common structure for each factor name to emphasize that these factors represent students' perceptions of their instructor. Thus, each factor name is written using the same structure as the FIBI items: "My mathematics instructor …" This commonality is meant to provide FIBI users with six collective statements, from the students' perspective, about their instructor's behaviors that may be increasing/decreasing the students' feeling of inclusion in their classroom. In this respect, we phrased the titles to not be a definitive judgement of the instructor, but rather a current understanding that the students have developed with respect to their own class. For example, for Factor 5, a few of our initial phrases defining the factor were: "poor time management" and "inattentive to teaching." As we worked to incorporate both the issues of timing in Q14 and Q17 and attention to details in Q11 and Q13, our discussion turned to the idea of having priorities above providing quality instruction for students. While the factor could have been given a definitive title such as, "My mathematics instructor doesn't know how to manage their time well," we wanted to avoid the declaration that an instructor "doesn't know" or "lacks" some ability or effort. Instead, we chose the title, "My math instructor doesn't prioritize teaching us," because it addresses the students' current perception of their instructor's lack of attention and time management skills through the lens of "having priorities other than teaching" which an instructor can then address without feeling as if they have been labeled as "inattentive".

Table 3: Factor Correlations for 6-Factor Solution

	Factor 1 (–)	Factor 2 (+)	Factor 3 (+)	Factor 4 (–)	Factor 5 (–)	Factor 6 (+)
Factor 1 (–)	1	-0.61	-0.44	0.51	0.62	-0.33
Factor 2 (+)		1	0.71	-0.59	-0.58	0.65
Factor 3 (+)			1	-0.35	-0.31	0.6
Factor 4 (–)				1	0.51	-0.46
Factor 5 (–)					1	-0.43
Factor 6 (+)						1

(+)/(–) after each factor denotes whether the factor emphasizes positive/ negative impacts on inclusivity.

After providing names for each of our factors, we have three positively framed and three negatively framed factors. Based on previous work, we expected the factors of inclusion to be dependent on each other. Thus, we also investigated the correlations between our factors which are presented in Table 3. Of our six factors, three are focused on positive behaviors to support inclusion and the other three address negative behaviors that may inhibit feelings of inclusion in a mathematics classroom. So, we expected that factors addressing negative behaviors will have an inverse relationship with positively focused factors. Moreover, two factors that are both positively (negatively) focused were expected to have a positive association between the two factors. This relationship did indeed present itself in our data. Additionally, we note that the magnitude of the associations between factors ranged from 0.31 to 0.71, suggesting that all of our factors are moderately dependent on one another and thus the different aspects of students' perceptions of inclusive instructor behaviors are linked to each other.

Discussion

This paper addresses the beginning stages of development for the FIBI instrument, a measure of students' perceptions of inclusive instructor behaviors in undergraduate mathematics classrooms. Development of this instrument began with an emphasis on the inclusion of women and PEER students in defining inclusive practices through targeted focus groups which became the basis for items included in the FIBI instrument. After developing these items, we proceeded through an iterative process of exploratory factor analysis and item reduction to arrive at our current 28-item instrument which contains six factors, of which three describe instructor behaviors that promote inclusion and three describe behaviors that inhibit feelings of inclusion.

One limitation of our work is the lack of diversity present in our sample. Although we were able to obtain a majority of our responses from women, our sample did not contain enough students who identified as PEERs to assess the factor structure for this subgroup alone. Additionally, other underrepresented identities such as Indigenous, Native Hawai'ian, Pacific Islander, non-binary gender identities, are not well represented in our sample and we did not investigate sexual identity as an identifier for underrepresented students in mathematics class. It may be possible that students from these groups would identify alternative factors which we did not see in this study. However, for our future confirmatory work, we plan to make a more targeted effort to include more PEER participants and to understand any possible differences in the factor structure. Another limitation of our study comes from our focus on deriving items from the responses of students in our focus groups. Because many students have only experienced traditional, lecture-based classrooms, it is doubtful that they have experienced any teaching strategies that may promote feelings of inclusion. Hence, we have aimed our instrument at defining only "fundamental" behaviors that may be present in a wide variety of teaching styles.

While this paper has presented a basic factor structure for the FIBI instrument, future analysis is needed to confirm this factor structure. Thus, our future steps for instrument development include a secondary data collection to use for confirmatory factor analyses (CFA) to validate our factor structure. Additionally, to continue centering women and PEERs in defining inclusive practices, we plan to run a multiple groups CFA model to analytically compare the factor structures and factor loadings based on gender and race/ethnicity. Finally, the FIBI instrument must be compared to other measures of similar constructs to assess construct validity and we plan to investigate the relationship between FIBI factor scores and other constructs such as math confidence, belonging, as well as course grades. These next steps will be crucial for establishing the validity of the FIBI so it can then be used by instructors aiming to assess and improve feelings of inclusion in their math classrooms.

Acknowledgments

This work was supported by the National Science Foundation, grant number 1937617.

References

Asai, D. J. (2020). Race matters. Cell, 181(4), 754-757

Battey, D., Amman, K., Leyva, L. A., Hyland, N., & McMichael, E. W. (2022). Racialized and gendered labor in students' responses to precalculus and calculus instruction. *Journal for Research in Mathematics Education*, 53(2), 94–113. https://doi.org/10.5951/jresematheduc-2020-0170

Brown, N., Hatfield, N., Lewis, K., Sine, A., Palmer, L., Workman, D., & Esan, O. (2023). Development of an instrument for fundamental instructor behaviors for inclusion in undergraduate math classes. Manuscript in preparation.

Dewsbury, B., & Brame, C. J. (2019). Evidence based teaching guide: Inclusive teaching. *CBE Life Science Education*. https://lse.ascb.org/evidence-based-teaching-guides/inclusive-teaching/

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLOS ONE*, 11(7), e0157447. https://doi.org/10.1371/journal.pone.0157447

Hogan, K. A., & Sathy, V. (2022). *Inclusive teaching: Strategies for promoting equity in the college classroom* (First edition). West Virginia University Press.

Holgado–Tello, F. P., Chacón–Moscoso, S., Barbero–García, I., & Vila–Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153–166. https://doi.org/10.1007/s11135-008-9190-y

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. https://doi.org/10.1007/BF02289447

Inclusive STEM teaching project. (2023). Inclusive STEM Teaching Project. https://www.inclusivestemteaching.org/

Johnson, K. M. S. (2019). Implementing inclusive practices in an active learning STEM classroom. *Advances in Physiology Education*, 43(2), 207–210. https://doi.org/10.1152/advan.00045.2019

Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed). Pearson Prentice Hall.

Kachani, S., Ross, C., & Irvin, A. (2020). 5 principles as pathways to inclusive teaching. *Inside Higher Ed*. https://www.insidehighered.com/advice/2020/02/19/practical-steps-toward-more-inclusive-teaching-opinion

Leyva, L. A., Quea, R., Weber, K., Battey, D., & López, D. (2021). Detailing racialized and gendered mechanisms of undergraduate precalculus and calculus classroom instruction. *Cognition and Instruction*, 39(1), 1–34. https://doi.org/10.1080/07370008.2020.1849218

She, H.-C., & Fisher, D. (2002). Teacher communication behavior and its association with students' cognitive and attitudinal outcomes in science in Taiwan. *Journal of Research in Science Teaching*, 39(1), 63–78. https://doi.org/10.1002/tea.10009

Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2. ed). Erlbaum.

Tanner, K. D. (2013). Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity. *CBE—Life Sciences Education*, 12(3), 322–331. https://doi.org/10.1187/cbe.13-06-0115

Wubbels, T., Brekelmans, M., & Hooymayers, H. (1991). Interpersonal teacher behavior in the classroom. In *Educational environments: Evaluation, Antecedents, and Consequences*. (pp. 141–160). Pergamon Press.