Collaboration Through Participation: Rethinking Scale Conceptualization and Development in STEM Education Research

Dr. Cijy Elizabeth Sunny, Baylor University

Dr. Cijy Elizabeth Sunny is a PD Research Associate in the Department of Information Systems and Business Analytics, Hankamer School of Business at Baylor University. She is a research methodologist and psychometrician who has applied her skills in quantitative and mixed methods research methodology in the substantive areas of STEM education research, medical education, and more recently in engineering education. Additionally, she has been an educator and has taught primarily physics and also research methodology on three different continents. In addition to research, she has also conducted workshops on using concept mapping methodology for scale development, mixed methods research methodology for standardized patient educators, and standard-setting for physician educators. Dr. Sunny continues to invest her skills in engineering education research through her collaborations. As part of her new undertaking at Baylor University, she is investing her skills as a research methodologist and data analyst to fight human trafficking through the use of Information Technology working alongside the research team there in collaboration with a diverse group of stakeholders.

Dr. Kathleen Koenig, University of Cincinnati

Dr. Koenig is a Professor of Physics at the University of Cincinnati, with her research in physics education. She has extensive experience in the development and evaluation of pedagogies and curriculum that support student success in college-level science and math courses. Since 2006, she has been funded by NSF and/or NIH to investigate the development and evaluation of curriculum that advances student scientific reasoning abilities. This work also involves the development of reasoning assessments for students in various age groups, including middle school through college. Funding since 2014 has centered on the development of interactive, web-based learning modules to enhance student conceptual understanding and problem solving abilities. She is a Fellow of the American Association of Physics Teachers (AAPT).

Collaboration through Participation: Rethinking Scale Conceptualization and Development in STEM Education Research

Abstract

Traditionally, tools or scales have been developed through a process involving review of the literature on current instruments, expert consultation, and conversations with target groups. While it is an established approach, it does not include the ideas of the STEM community as a whole who are engaged in improving the STEM pipeline. How does one develop a tool that mirrors the diverse perspectives of stakeholders who are uniquely positioned to influence and support students in STEM?

The answer lies with the community-based participatory mixed methods methodology of concept mapping that was developed by William Trochim at Cornell University. Concept mapping methodology innovatively engages the stakeholders through participation from outreach to consultation to collaboration through a shared leadership. This methodology integrates qualitative and quantitative approaches in a multi-step democratic process of community engagement and participation through research.

This traditional research paper presents the application of this five-step concept mapping methodology to develop the Attitude and Persistence Towards STEM (APT-STEM) tool. The primary objective of this paper is to present an innovative methodology to develop scales through the participatory framework by involving the voices and unique experiences of a STEM community of stakeholders (e.g., students, parents, teachers, researchers, educators, and professionals) as part of the research process to conceptualize the constructs using both quantitative and qualitative methods. While not the focus of this paper, we will also briefly present the validation of this tool for use with secondary and pre-secondary students and an updated version for first-year engineering students in a calculus-based introductory physics class using item response theory. In addition to the methodological (concept mapping for scale development) and theoretical implications (participatory framework) of this study, its practical implications include examining the efficacy of STEM education programs in cultivating specific attitudes and persistence traits as conceptualized by the STEM community.

Introduction

As evidenced in the landmark study by Seymour and Hewitt [1], since the mid-1980s, enrollment and retention in science and mathematics related fields have decreased. This study has been a trailblazer in starting a movement that has gained national and international attention and triggered a renaissance in Science, Technology, Engineering, and Mathematics (STEM) education for decades now. In the past, STEM fields have resulted in the U.S. accounting for 40% of all research and development, 35% of science and engineering publishing, and 44% of science and engineering citations, while employing one-third of the world's scientists and engineers [2]. However, this intellectual capital is now the focus of other nations because of the lack of interest of the American youth towards science and engineering [3]. The President's Council of Advisors on Science and Technology (PCAST) report corroborates this and highlights the need for 1 million more college graduates in STEM fields [4]. In response to this need, the

STEM education research community in the U.S. has specified the nature of instructional strategies in retaining students in STEM-related courses, with a focus on an integrated STEM curriculum designed to improve non-cognitive factors, such as interest, while developing positive attitudes towards STEM [5][6][7]. Interests and attitudes in science develop early in a student's life, and it is important to develop these attitudes as they are motivators towards pursuing STEM fields and careers [8] [9]. More recently, the National Academies of Sciences, Engineering and Medicine (NASEM) 2017 report on supporting student's college success has highlighted the importance of intrapersonal and interpersonal competencies and the evolving need for labor market recruits to attain these non-cognitive competencies, which are goals for K-12 and higher education [10].

Given the raised awareness for the importance of these non-cognitive skills, assessments developed to measure these are essential. As per the NASEM report recommendations, the specific skills or constructs need to be clearly conceptualized and must be designed, developed, analyzed, and interpreted based on stakeholder needs [10]. The purpose of this research paper is to introduce a new and innovative methodology to the engineering education research community, named Concept Mapping [11], which has traditionally been used in evaluation and program planning in the health sciences. This methodology will be explained in the context of how it was used in developing the Attitude and Persistence Towards STEM (APT-STEM) tool through the participatory collaborative framework by involving the voices and unique experiences of a STEM community of stakeholders (e.g., students, parents, teachers, researchers, educators, and professionals) as part of the research process to conceptualize the constructs using both quantitative and qualitative methods. While not the focus of this paper, we will also briefly present the validation of this tool for secondary and pre-secondary students as well as an updated version of the APT-STEM with first-year engineering students in a calculus-based introductory physics class using item response theory. The research questions that were investigated included:

1. What are the indicators (ideas) of attitudes and persistence towards STEM as conceptualized by the STEM stakeholders and how were they rated by them?
2. What are the psychometric properties of the developed tool (APT-STEM) as evaluated by item response theory (IRT)?

This paper shares the main highlights from a dissertation (published) study [12], where the stakeholders were mostly part of a local STEM community collaborative. This paper begins with a presentation of the conceptual framework, wherein the constructs attitude and persistence will be defined after a brief presentation of the background. This is followed by a methods section, which will be divided into the concept mapping phase and the instrument validation phase. Results will be presented for each, and the paper ends with a discussion and final conclusions.

Conceptual Framework

Cognitive skills or competencies include strong memory, processing speed, problem-solving, motor, and linguistic skills, or in short, anything that involves thinking and is reasoning related [13] [10]. Non-cognitive skills, on the other hand, refer to a set of personality characteristics that include behaviors, attitudes (beliefs and feelings), and strategies associated with individual success like persistence, self-management, and communication [10]. Additionally, they also include optimism, resilience, adaptability, and conscientiousness amongst others [14]. The majority of the instruments that have been developed to measure non-cognitive skills have primarily involved researchers with limited participation from stakeholders. This holds true in engineering education as well. For example, the student attitudinal success instrument, which was designed to collect data on non-cognitive characteristics for incoming engineering students, was developed by researchers guided by the literature [15] [16]. The national survey to study non-cognitive traits of undergraduate engineering and computing students was developed as part of a three-institution collaboration with six principal investigators, and strategic conversations were handled through emails and conference calls [17]. The Survey of Engineering Ethical Development (SEED) instrument was an exception in that it employed qualitative methods of focus groups and interviews with engineering students and faculty as part of its development. This instrument was designed to measure the curricular and co-curricular events and experiences that would affect the ethical development of undergraduate engineering students [18]. Instruments that have been developed in engineering to measure non-cognitive skills have mostly used the traditional development process of reviewing the literature on existing instruments to guide their own scale development. Here, the conceptual domain is developed through literature reviews, expert consultation, peer review, and conversations with target groups [19]. These are well-established and validated approaches; however, they do not build on the ideas of the various stakeholders that are involved in improving the STEM pipeline. Rather, the ideas developed by the STEM stakeholders need to be structurally conceptualized by using both qualitative and quantitative methods, as these processes can lead to a broader conceptualization of the constructs or domains measured and add to the content validity of the instrument by developing items using the shared experiences of the STEM stakeholders. This is challenging, but a feasible and effective process that is encapsulated in the methodology of concept mapping, which involves five distinct steps as shown in Figure 1.

Concept mapping is a participatory mixed methods approach that integrates qualitative group processes with multivariate statistical analyses. The approach helps a group of stakeholders describe their ideas on a particular topic by representing these ideas in a two-dimensional space [11] [20]. At its core, it employs a participatory framework to aid collaborative description and exploration, while balancing systematic inquiry with respect to the ideas of the stakeholders through a systematic and stepwise process. This framework and approach to research involves a diverse range of participants from educational institutions, communities, industry, etc., and eventually leads to a solution. A mixed methods approach to data collection, analysis, representation, and interpretation in turn adds to the overall value of the inquiry process [21]. This methodology has been used primarily in the context of program evaluation to map the views of health-related stakeholders and help with strategic planning. It has subsequently been used to develop program evaluation instruments in the field of family support programs [22] [23] [19]. Because this paper highlights the use of concept mapping in developing the APT-STEM instrument, it provides readers with an example of this methodology applied in a real context.

APT-STEM was developed to measure two non-cognitive skills, namely attitudes and persistence, chosen based on the needs of a local STEM community collaborative. Attitudes are defined as feelings, beliefs, likes, and dislikes towards various aspects of STEM [7]. Persistence is defined as a passion for persevering through long-term goals [24] [25]. The process of concept mapping that was used in the development of the tool involved a participatory framework, whereby participation was by consultation, including functional and interactive participation [26]. The STEM stakeholders worked collaboratively as part of a community as they developed measures based on their experiences and knowledge from working in the field of STEM and STEM education. Their professional identities and experiences validate their role, while providing representation for their voices on an instrument that was developed in this unique manner.

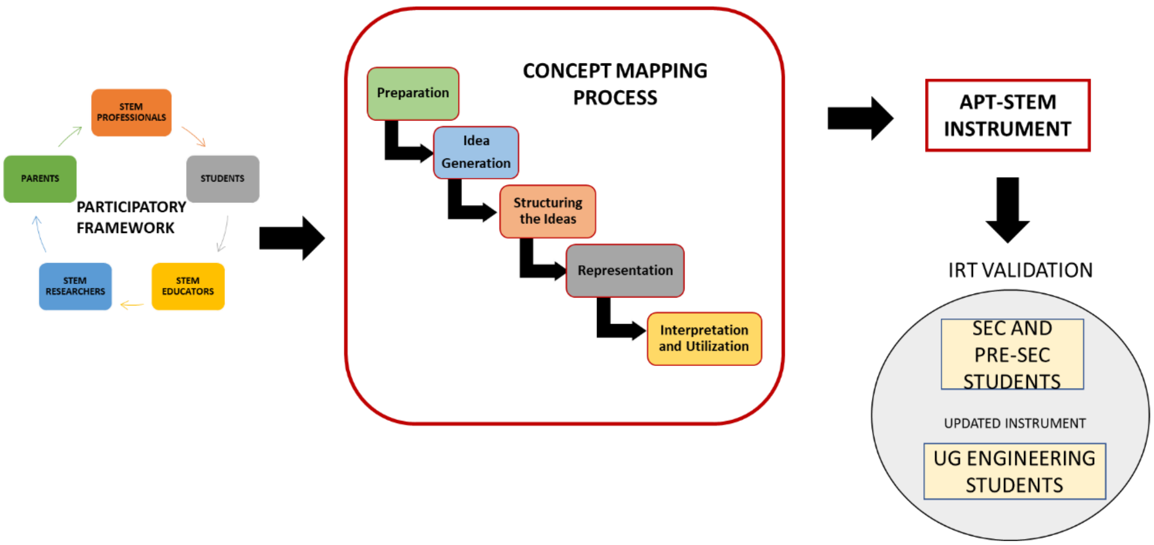

FIGURE 1: Conceptual framework for the APT-STEM instrument

As shown in the conceptual framework in Figure 1, the concept mapping process involves five steps, including preparation (defining community and focus prompt), idea generation (generating responses to the focus prompt), structuring the ideas (sorting and rating), representation (multidimensional scaling and cluster analysis) and finally interpretation and utilization. Each of these steps is presented in the context of the developed APT-STEM instrument in the following methods section.

Methods

Concept Mapping-APT-STEM Instrument Development

Sampling:

The APT-STEM instrument was developed to meet the needs of a local STEM community collaborative. When concept mapping is used for a particular purpose such as this, it involves a non-probabilistic sampling method, which has been done in other studies using this methodology [19] [27]. STEM stakeholders, including educators, researchers, professionals, students, and their parents, generated the content domain of the instrument due to their familiarity with various

STEM programs that were implemented to improve student attitudes and persistence towards STEM. The experiences of these stakeholders also contributed to the development of the conceptual domain for the positive attitudes and persistence characteristics needed to attract and retain students in STEM courses and careers. Additionally, during various stages of the methodology, a mix of nested (subset of the sample of stakeholders that were in the previous stage) and identical (same number of stakeholders) samples were used [28]. Unlike other processes, the goal in concept mapping is to achieve an extensive sampling of ideas rather than a representative sample of persons [20].

For simplicity, the steps involved in the CM process are shown visually in Figure 2.

FIGURE 2: *Different stages in the concept mapping process* [12]

Preparation. In this step of the concept mapping process, the stakeholder group was identified. The first set were adult stakeholders, who were members of the regional, state, or national STEM Collaborative through the Office of Innovations and Community Partnerships (OICP) at the University of Cincinnati, and they were sent the prompts in the Summer of 2016. The second set of stakeholders were secondary and pre-secondary students (collaborating schools/programs with OICP). There were a total of two prompts (incomplete statement that helped generate ideas pertaining to a topic [20] [29]) that were sent to each group of stakeholders (adults and students). The first was to gather ideas on positive attitudes towards STEM (*I believe that some positive attitudes (feelings, beliefs, likes, and dislikes) towards Science, Technology, Engineering, and Mathematics (STEM) that school students' (K-12) could have are…)* and the second was to gather ideas on persistence (*I believe that some of the characteristics of a persistent (desire or passion to stand firm through long term goals) school student (K-12) towards Science, Technology, Engineering, and Mathematics (STEM) coursework and careers are…*). The prompts for the student groups (middle and high school students) had the wording simplified without losing the core idea.

Idea generation. Following IRB approval, the concept mapping idea generation prompts were sent to adult stakeholders in Summer of 2016 and students in Fall of 2016. According to Kane & Trochim [20], typically 40 or fewer people, with a minimum of 10, ensures a variety of ideas. A set of ideas in response to the prompts were generated and these ideas represented the entire conceptual domain. As this process created a vast number of ideas, which would later become APT-STEM assessment items, it was important to impose some constraints for practicality for later stages in the concept mapping process [20]. That is, a final number of 100 ideas or fewer is considered suitable as long as it represents the entire breadth of the ideas generated [11]. Therefore, to reduce the number of ideas to a more manageable number as suggested in the literature, three iterations were used on each set of ideas to check for redundancy using iterative content analysis. Additionally, NVivo [30] was used to ensure the reduced ideas were unique and representative of the original pool. The query feature in NVivo, based on the preferences set by the primary researcher, helped identify the context in which the words or phrases were used in the generated ideas. After this cleaning and refinement process, it was decided that the two sets of ideas be combined as the ideas generated by students overlapped most of the ideas generated by the adult STEM stakeholder group. After combining the ideas, the iteration process continued. In the first iteration, ideas that were not consistent with the focus prompt and single word responses that could be interpreted multiple ways like 'timely', 'practical', etc., were deleted. The second and third iterations were performed to check the primary researcher's reliability with respect to deleting the ideas based on the above criterion for deletion. Some of the ideas were also corrected for grammar and clarity without compromising the meaning of the original idea. As an additional step of validation, the final ideas were uploaded back into NVivo to do a final content analysis as mentioned earlier. The final ideas were then refined further for vocabulary and grammar in consultation with 5 stakeholders from the initial core group who were part of the local STEM collaborative.

Structuring the ideas. This step involved two tasks, namely, sorting and rating as presented below. For the subsequent steps of concept mapping, we used regional stakeholders because of convenience and their partnership with the local collaborative through OICP.

Sorting. A subset of the original sample of stakeholders from the local STEM collaborative participated in the sorting task. Each sorter was presented with all of the final ideas on a set of numbered cards with one idea per card. They then sorted the cards into groups of similar ideas and created a descriptive name for each group. Finally, they recorded their group name and the ideas (numbered cards) were placed in that group. Each idea was only sorted into one group. Once the ideas were sorted by the stakeholders, this raw data was structured into matrices of 0's and 1's using the open source programing language R [31]. This step generated individual and total matrices, which served as an input for the representation step. The sorting step was an onsite activity and was completed in two hours, with each stakeholder sorting the ideas individually. Prior to performing multidimensional scaling (the next stage), the sorted data was entered into EXCEL in a racked format (could be stacked as well). This data comprised of researcher generated sorter identification codes, group labels (names or themes given to groups by the sorters during the sorting task), and idea numbers that formed each group. This was reshaped to get a total matrix (a square matrix that is symmetric) using R.

Rating. While rating is an optional step in concept mapping, it was included in this study to help with item selection on the final instrument. This step addresses each stakeholder's perception of the importance or other value qualifier of an idea [20]. It is usually on a Likert-type rating scale and this study employed a 5-point rating scale, with 1 representing *not very important* and 5 representing *extremely important*. A parallel sample (same as sorting) of 33 adult regional stakeholders performed the rating remotely via Qualtrics for the importance of the attitudes and persistence ideas.

Representation. The two core analyses performed at this stage were multidimensional scaling and cluster analysis. This was performed on the sorted set of ideas.

Multidimensional scaling. Concept maps were computed with a multidimensional scaling (MDS) analysis in R, wherein each idea was placed in a two-dimensional space. This method is used to unravel latent patterns by representing measurements of similarity (or dissimilarity) among pairs as distances between points in an n-dimensional space [32]. In concept mapping (as shown in the results section), MDS analysis creates a map of points, called the point map, which represents the set of brainstormed ideas based on the total similarity matrix developed as a result of the sorting task [11]. Ideas that are closer together on this point map are generally grouped together by the sorters, and more distant ideas are less frequently grouped together. In concept mapping, the ideas are usually mapped into a two-dimensional bivariate plot, and it is easier to visualize and interpret when coupled with cluster analysis [20] [33]. The focus is not dimensionality in concept mapping but the ability of MDS to portray relationships among the generated ideas [20]. This ability is measured using a key diagnostic statistic named stress [33]. Stress measures the degree to which the distances on the point map are different from the input similarity matrix. In other words, it is a badness-of-fit measure such that the higher the stress value, the worse the fit. A meta-analysis study estimated that 95% of the concept mapping projects yield stress values between 0.205 and 0.365. For 100 ideas, the acceptable stress boundary for two dimensions is 0.396 [20] [34]. In other words, for a point map to be a good representation of the data, this value should be below 0.396.

Cluster analysis. This analysis partitions the ideas on the point map into multiple clusters and forms conceptual groupings of the original set of statements (ideas). As concept mapping uses distance-based data, Ward's algorithm is commonly used as it leads to more interpretable and reasonable solutions than others [11] [20]. MDS and cluster analysis was performed using the SMACOF package in R [35]. The decision regarding the number of clusters was made both quantitatively and qualitatively for consistency. For quantitative analysis, the number of clusters was determined using the agglomeration coefficients (elbow plot), majority rule [36], and coefficient selection (Hartigan Coefficient) using the NbClust package in R [37]. For qualitative analysis, an interpretive process was used to describe attitude and persistence ideas with both fewer clusters and sufficient detail [38].

Interpretation and utilization. In the final step, researchers traditionally facilitate the session with the stakeholders to name the final clusters, explore ideas within each cluster, and discuss further insights and actions that should be taken as a community [20] [29]. In this study, as this methodology was used to develop the scales, the interpretation and utilization was done by the primary researcher in collaboration with five stakeholders who were part of the regional STEM collaborative leadership team. During this step, the scales were developed and the names for the clusters were identified using both quantitative and qualitative methods. Quantitatively, the X-Y distance coordinates for all stakeholder-generated labels were calculated using R, and the distance between each label and each cluster center was computed. Later, a qualitative thematic analysis was conducted and labels closest to the center (determined using the above quantitative method) were used as a guide such that the selected names preserved the meaning provided by stakeholders [39].

Item Reduction and Selection. The term “item” is commonly used in the instrument validation literature, so we use that term here to represent any idea that made it to the final APT-STEM instrument. In this paper, it will only be referred to as an item after the concept mapping process. A traditional psychometric process was followed by computing item-total correlation analysis using SPSS to determine which ideas or items related strongly to an overall rating of importance (rating data as mentioned previously). The standard specification that recommends items with item-total correlations below $r = 0.40$ be excluded in the item reduction process was followed [40]. Items were reviewed for clarity and content by a core group of STEM stakeholders ($n = 5$) in order to further eliminate items whose meanings were considered ambiguous for student participants. Some items, which had a low item-total correlation but rated as being important by stakeholders, were retained after consultations with this core group of stakeholders. Five items were selected to be written as negatively worded to assess for response bias.

APT-STEM Validation Using IRT

As part of the dissertation study, the APT-STEM was validated for the secondary and pre-secondary sample using all three psychometric frameworks (classical test theory, item response theory, and Rasch measurement). Given that the results were consistent, this paper will only focus briefly on IRT.

Sampling:

Secondary and pre-secondary sample. A total of 556 middle and high school students from different STEM programs in the U.S. Midwest were remotely (Qualtrics) administered the APT-STEM instrument between May and July 2017 during which the different programs were held. The criteria for selection of programs by the local collaborative was dependent on their capacity to implement, plans for sustainability, innovation, STEM engagement best practices, more high-risk students, schools within the business vicinity, and sponsor priority [12].

Post-secondary sample. In Spring of 2019, an updated APT-STEM was administered to 667 students enrolled in a first semester calculus-based introductory physics course for engineers. This was done for continued validation of the instrument. However, because this was an older group of students, the items were slightly reworded by the primary researcher in collaboration with the course instructor. Also, this updated version had a total of 30-items compared to 24-items from the post-validation phase of the 2017 sample. This resulted because the items were re-worded for the 2019 sample and the negatively worded items (each unique), which were previously deleted after the first validation (2017 sample) due to method effects, were now positively worded and added to the test to replace the negative versions.

Model:
The item responses were analyzed using a graded response model (GRM) [41]. This model is commonly used with ordered polytomous categories on attitude surveys involving Likert scales of strongly disagree, disagree, agree, and strongly agree [42]. The hypothetical model for a two-dimensional GRM with two correlated constructs is shown in Figure 3. This paper will not present the mathematical model but an explanation can be found in the published dissertation. As is traditionally followed in IRT, item fit statistics were obtained. Cut-off criteria for a reasonable fit were SRMR and RMSEA < 0.08, CFI and TLI > 0.90 or 0.95 [43]. Items with |Yen's Q3| > 0.20 (Q3 fit statistic represents the correlation between the residuals for a pair of items) has local dependence and significant item fit values ($p < 0.05$) revealed misfit items [44]. Finally, item and test information functions graphically reflected the reliability (1 - [1 / peak information]) of the items and the test as a whole in estimating the construct over the entire scale range [45].

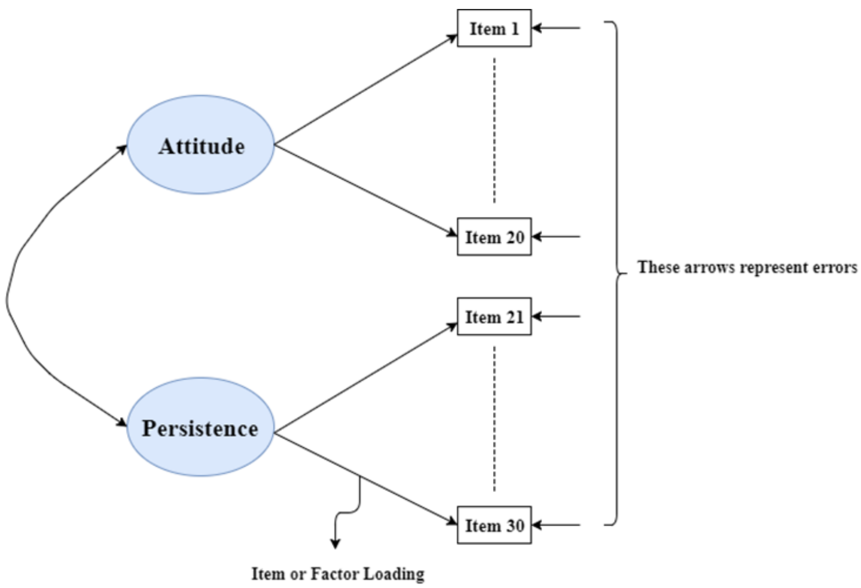

FIGURE 3. Hypothesized 2-D measurement model for the APT-STEM instrument [12]

<u>Results</u>

The results of the concept mapping steps will be presented first, followed by the validation results for each of the two samples (2017 and 2019).

Concept Mapping

Idea Generation:

Ninety-eight adult stakeholders responded to the prompts and conceptualized a total of 350 ideas for attitudes and 348 ideas for persistence. Ninety-two students conceptualized 349 ideas for attitudes and 326 ideas for persistence. The demographic characteristics of this sample of adult and student STEM stakeholders are summarized in Table 1.

Final Ideas. The iterative and NVivo based content analysis generated a final set of 94 ideas that represented attitudes and persistence towards STEM and also represented the entire breadth of representation of all attitude and persistence ideas generated by the stakeholders. This final number of ideas also provided the sorters and the raters in the subsequent stages with a manageable number of ideas. Later, the ideas were taken to the regional STEM collaborative core leadership team (n = 5) and they helped to further revise the ideas.

Structuring the Ideas:

Sorting. There was a total of 32 sorters who participated in the sorting task and their characteristics are shown in Table 1. The majority of the sorters were also parents, but they only identified their primary role as one in the STEM field. Each of the 32 participants sorted the 94 ideas into groups containing similar ideas and gave each group a name or a label. The number of sorted groups varied from 4 to 18.

TABLE 1: *Demographic Characteristics of the STEM Stakeholders (n = 190)* [12]

Characteristic	*n*	%
Adult Stakeholders	98	52
Gender		
Female	62	63
Male	36	37
Age		
<20	4	4
20-39	28	29
40-59	47	48
>60	19	19
Race		
Black	13	13
White	79	81
Other	6	6
Professional Role		
Student	5	4
Teacher	28	22
STEM professional	22	17
Parent	11	9
STEM researcher/educator	32	25
Other	30	23
Student Stakeholders	92	48
Gender		
Female	55	60
Male	37	40
Grade		
Elementary (4)	9	10
Middle (6-8)	73	79
High (9-12)	10	11
Race		
Black	14	15
White	65	71
Other	13	14

Rating. The 94 items were rated for importance by 33 raters, with all but one having participated in the sorting task. This number is higher than the average number of raters reported in a recent meta-analytic study [46]. Seventy three percent (n = 24) of the raters were female and 27% (n = 9) were male. Fifteen percent (n = 5) identified themselves as Black, 82% (n = 27) as White, and 3% (n = 1) as other. There were two distinct groups of raters in the sample. One group (45%, n = 15) was referred to as STEM teachers, which included educators and curriculum specialists. The other group (55%, n = 18) was referred to as STEM professionals and/or researchers, which included engineers, business partners, physicians, STEM researchers, and STEM program evaluators. For all 94 items, the STEM stakeholder ratings for importance were on average in the range of 2.64 to 4.84.

Representation:

Multidimensional Scaling. The MDS of the sorted data located each of the 94 attitude and persistence ideas as a separate point on a two-dimensional map. The point map that was generated for all 94 items is shown in Figure 4. Ideas that are closer together were sorted together more frequently and ideas that are farther apart on this map were sorted less frequently together. The stress value for the two-dimensional MDS analysis was 0.25, which is within the acceptable range of 0.205-0.365. Also, because the stress value is below 0.39 (for 94 ideas), this also confirms that this two-dimensional scaling results are not random and has a structure [34].

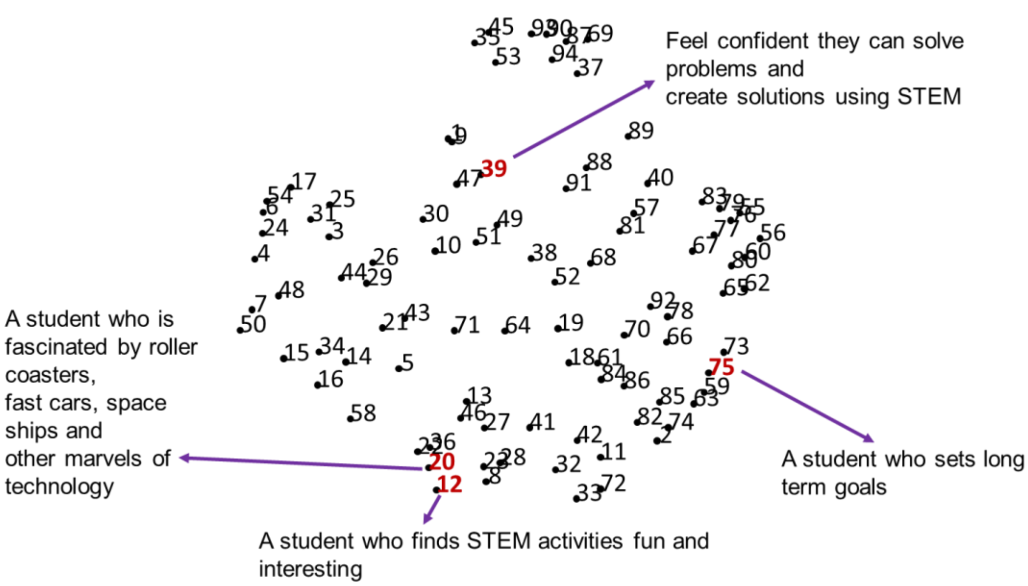

FIGURE 4. A point map of 94 attitude and persistence ideas [12]

Cluster Analysis. A cluster analysis was performed on the two-dimensional X-Y distance data for the ideas using the change in agglomeration coefficients from one cluster to another. It was noted that the change was small for the 5th and subsequent clusters. Based on this method, it was decided that 4 clusters were the optimum cluster solution for this data. This was further validated by plotting a line graph of the agglomeration, with the cutoff approximately around the

90th idea, leading to a four cluster solution (94-90). The elbow plot showed either a 4 or 5 cluster solution. Further, the majority rule criterion in the NbClust package in R [36] and coefficient selection [37] both suggested a four-cluster solution. Qualitatively interpretive analysis also confirmed the four-cluster solution, which was further validated by consultations with the regional collaborative core leadership group (*n* = 5) [12].

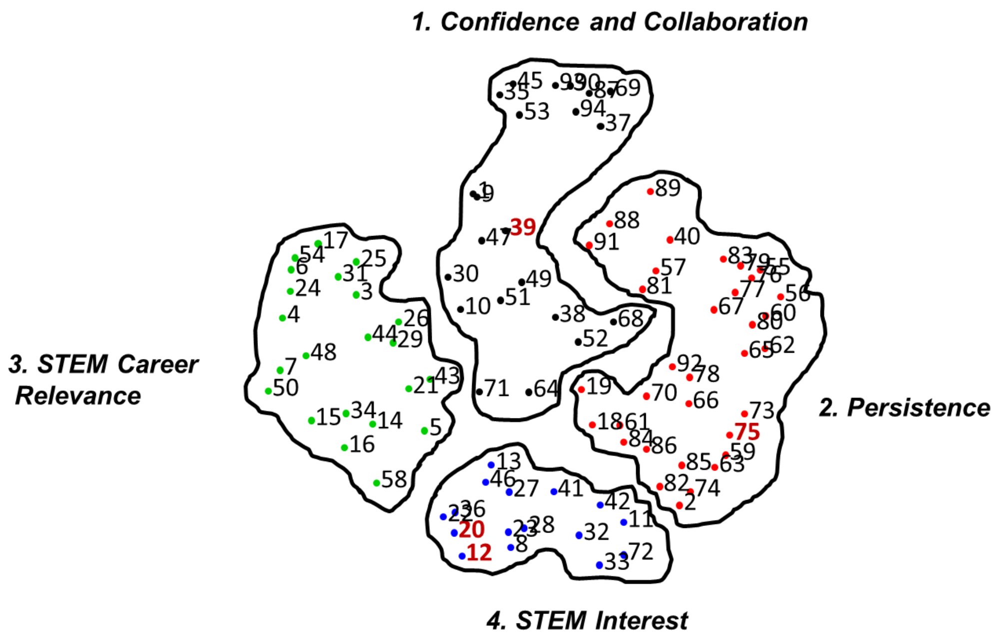

FIGURE 5. Cluster map of the 94 ideas [12]

The labels for the clusters in Figure 5 were grounded in the meaning provided by the stakeholders. For example, idea 5 (*STEM opens a world of wonder and exploration*) and 58 (*Independently looks for information about their future career options*), which are both in the lower middle portion of cluster 3, are close to each other but not near idea 89 (*A student who is ok with constructive "criticism" and "feedback"*), which was located on the opposite side and top of cluster 2. This indicates that the sorters frequently sorted ideas 5 and 58 in the same group, as they thought that these ideas were conceptually similar, but did not sort them with idea 89, as they are conceptually dissimilar. Figure 5 shows the best solution possible for all of the ideas in the two-dimensional space by placing conceptually similar ideas closer together and dissimilar ideas farther apart [12].

Interpretation and Utilization (item reduction and selection).

As mentioned in the methods section, for this study, this step involved the item (idea) reduction and selection using the item-total correlation. A more stringent cutoff of 0.60 was used to select the items from all 94 ideas based on the rating. Most of the stakeholders found all items to be important and a cutoff of 0.60 ensured that items from all clusters were included. If there were items that had a correlation less than 0.60, but were regarded by stakeholders to be important, those items were retained. Qualitatively, items were added and deleted based on this criterion. As a result, the initial item pool was reduced from 94 to 48 items. The primary researcher and another STEM education researcher reviewed the remaining 48 items for clarity as well as content, and 18 additional items were eliminated based on ambiguous wording and concerns over how they may be interpreted by students [19]. Items specific to science and mathematics, rather than only engineering and technology, were also retained as stakeholders reiterated that the knowledge base for engineering and technology stems from science and mathematics. Also, these are terms that middle and high school students are most familiar with.

The final 30- item instrument (APT-STEM) was structured with 25 positive items and 5 negative items for which students responded using a 5-point Likert-type response scale format. Each item was rated by the participants on a range from *strongly disagree* (1) to *strongly agree* (5). Note that the final instrument was arrived at after consultations with the core leadership group of the regional STEM collaborative. This was done to maintain the participatory process and participation of the STEM stakeholders through the entire APT-STEM development process.

IRT Validation

Secondary and pre-secondary sample:

A total of 556 grades 5 to 12 students from the U.S. Midwest participated in the validation phase of this study. The majority of the respondents were from the 2017 Summer STEM programs (88%, n = 487) followed by other programs, including the STEM bicycle club (6%, n = 34), 3D printing club (3%, n = 17) and iSPACE (3%, n = 18). Each club was comprised of students from multiple schools. The APT-STEM survey was administered during the first week of each program, with program durations varying from one to 10 weeks. Of the 556 students included in this study, 55% (n = 306) were female, 39% (n = 217) were male, and 6% (n = 33) chose not to answer the question about gender. Thirteen percent (n = 73) were from fifth grade, 29% (n = 163) from sixth grade, 43% (n = 240) from seventh grade, 7% (n = 38) from eighth grade, and 8% (n = 42) from grades 9-12. Forty one percent (n = 227) identified themselves as White, 32% (n = 179) as Black, 2% (n = 10) as American Indian or Alaskan Native, 4% (n = 24) as Asian, Asian Indian, or Pacific Islander, 6% (n = 35) as Hispanic or Latino, and 15% (n = 81) categorized themselves as "other". All 556 students completed all 30 questions on the instrument.

A two-dimensional GRM model with 24 items (excluding the negative items and item 9 (misfit items)) was found to be the best model after running several models. Compared to the 30-item APT-STEM (AIC = 39082.27; BIC = 39734.70), the final 24-item two-dimensional GRM model was a better model as it had lower AIC (297670.81) and BIC values (30193.63). The relative fit indices, RMSEA = 0.06, CFI = 0.92, TLI = 0.91, and SRMSR = 0.06, were also within the standard cut-offs. For the 24-item two-dimensional GRM model, all items had high slope values (> 1) (discrimination parameters) so they were good at discriminating between people of different attitudes and persistence towards STEM. The reliability of the overall instrument was 0.96. Because this is a two-dimensional instrument, the reliability on the attitude dimension was 0.95 and the reliability on the persistence dimension was 0.89. Two items (item 19, *I like to come prepared to my mathematics and science classes*, and item 29, *I am not discouraged by criticism while working on science or mathematics projects*) were found to have some degree of misfit as determined by the p-value of signed chi-squared test statistic (S_X2) ($p < 0.05$).

Post-secondary sample:

The APT-STEM instrument items (language) were updated to suit post-secondary students. Also, 6 items (5 five negatively worded items and one positively worded item) deleted after the first validation were added back with all negatively worded items now positively worded. This updated version of APT-STEM had 30-items all positively worded on the same 5-point Likert scale. This was administered in the spring of 2019 to primarily first year engineering students in a first semester calculus-based physics course. All 667 students who took the survey completed all questions. There were no significant differences in the response to the survey for students in different sections of this calculus-based course. As a result, all sections were combined for the validation analysis, with 96% (643) between the ages of 18-20 years, 77% (514) identified as male, 22% (148) identified as female, 82% (545) identified as White, 3% (21) as Black or African American, 4% (26) as Hispanic or Latino, and the rest chose "other" for race.

A two-dimensional GRM model with 28 items (excluding items 6 and 20) was found to be the best model for the post-secondary sample of students. Compared to the 30-item APT-STEM (AIC = 29776.72; BIC = 30582.72), the final 28-item two-dimensional GRM model was a better model as it had lower AIC (28457.81) and BIC values (29209.77). The relative fit indices, RMSEA = 0.05, CFI = 0.93, TLI = 0.91, and SRMSR = 0.06 were also within the standard cut-offs. For the 28-item two-dimensional GRM model, all items had high slope values (> 1) (discrimination parameters) so they were good at discriminating between people of different attitudes and persistence towards STEM. The reliability of the overall instrument was 0.98. Because this is a two-dimensional instrument, the reliability on the attitude dimension was 0.97 and the reliability on the persistence dimension was 0.93. Two items (item 6, *While working on group problem solving in science, technology, engineering, or mathematics, it is important to be respectful and be willing to listen to my group members,* and item 20, *I believe studying science, technology, engineering, or mathematics can lead to good careers*) were found to have some degree of misfit, as determined by the p-value of signed chi-squared test statistic (S_X2) ($p < 0.05$), so they were eliminated from the post validation version as shown in the appendix.

<u>Discussion</u>

The purpose of this traditional research paper is to present a unique structured mixed-methods methodology of concept mapping and demonstrate its use to develop scales to measure attitudes and persistence towards STEM. Another purpose was to report on its validation using IRT. The STEM stakeholders for this study generated ideas that covered a wide range of topics based on their experiences in the various STEM fields as an educator, researcher, professional, student, or parent with children in various STEM-related programs. The ideas can be summarized as related to interest in STEM, career relevance, characteristics of a persistent student who does well in STEM courses and careers, confidence in abilities, and being able to collaborate with STEM professionals [12]. Overall, the stakeholders rated all presented ideas as important, with educators giving higher ratings, which could be due, in part, to their interests in developing these skills or competencies in their students given their importance in the Labor market as reported in the National Academies report [10]. A tool that is developed using the participatory and structured concept mapping methodology leads to a broader conceptualization of the attitude and persistence constructs and adds to the content validity of the developed scales. While the labels for the clusters have familiar names, within these clusters there are items that are unique to this tool, details of which are explained in depth in the primary study [12]. For example, liking to better oneself by seeking new opportunities, being responsible for one's learning and experience, and getting better at science and mathematics through practice, are a few measures unique to APT-STEM.

While the APT-STEM versions discussed in this manuscript were slightly different in terms of the number of items and the wording of the items, for both samples of students the IRT analysis evaluated a two-dimensional APT-STEM instrument. Furthermore, the deletion of the negatively worded items after the validation for the secondary and pre-secondary sample due to low item-total correlations and discriminating power, confirmed the findings of other studies regarding the effect of negatively worded items on the response styles [47]. Converting these to positively worded items for the updated version of the instrument, which was administered to the post-secondary sample, reiterated this as it yielded a better fitting model. And finally, the overall reliability and the sub-scale specific reliabilities were high for both versions of the APT-STEM. While there is overlap in the roles some of the stakeholders identified as for the research, such as being a parent of a student in STEM while also being a STEM professional, this study only focused on what they identified as their primary role in the STEM education field in order to minimize issues given this overlap. Additionally, all stakeholder groups in this study rated all items as being important to develop attitudes and persistence towards STEM.

Conclusion

While this study utilized purposive and convenient sampling at various stages of the concept mapping process, the stakeholders from the local STEM collaborative may not be representative of a national sample. For a greater external validity, it would be important to replicate this process with stakeholders at the national level. Also, this was an exploratory and cross-sectional study and it would be important for future work to continue the instrument validation by establishing its concurrent and predictive validity. With Federal funding agencies, such as NSF, focus on broadening STEM participation for under-represented groups, it would also be important to validate this instrument in terms of investigating each item's performance across these different groups to study possible bias. The validity and reliability of the concept maps, though different from traditional methods, were established for this study and the details can be found in the primary study [12].

The implications of this study are multifaceted yet relevant to the engineering education (EE) community. First, in terms of the methodology, this study has introduced the EE community to a novel and innovative methodology, which enables the inclusion of diverse perspectives and experiences of stakeholders to conceptualize the constructs measured in a structured, systematic, and rigorous manner to add to the content validity of a developed instrument. This is important, especially in developing instruments to measure perceptions or prevalence of sense of belonging or harassing behaviors that involve the voices of all groups with diverse experiences. Second, its theoretical implications stem from the use of a collaborative and participatory framework employed in community-based participatory action research. This inclusion of stakeholders serves to increase the knowledge, understanding, and response to scientific research and may inform policies by capturing the wisdom and experiences of the STEM community and as such makes this a practice-based approach [48]. Finally, its practical implication is evident as it provides a tool to measure the attitudes and persistence of students and can be used to assess the efficacy of STEM-related education programs both in post-secondary and K-12 education. Research shows that persistence is a crucial behavioral characteristic in all academic settings [49] and later needed within STEM careers. Because this study adds to the literature on measuring non-cognitive skills, it supports the development of interventions and individualized support of all students, not just those in engineering, and thus has the potential to be part of efforts towards broadening participation in STEM.

Acknowledgement: This study was not funded. I would like to thank OICP at the University of Cincinnati and the Greater Cincinnati STEM Collaborative (leadership and stakeholders) for its participation through the various stages of the CM process. Additionally, I would also like to thank Dr. Dan McLinden for sharing his expertise in the field of CM.

References

[1] Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences* (Vol. 12). Boulder, CO: Westview Press.

[2] Freeman, R. B. (2005). Does globalization of the scientific/engineering workforce threaten US economic leadership? *Innovation Policy and the Economy*, Volume 6 (pp. 123-158). The MIT Press.

[3] Denney, J. J. (2011). *Priming the innovation pump: America needs more scientists, engineers, and basic research*. Fort Belvoir, VA: Defense Acquisition University

[4] Olson, S., & Riordan, D. G. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Report to the President. *Executive Office of the President*.

[5] Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: an analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952-984. doi:10.1002/tea.20439

[6] Koenig, K., Schen, M., Edwards, M., & Bao, L. (2012). Addressing STEM retention through a scientific thought and methods course. *Journal of College Science Teaching*, *41*(4), 23-29.

[7] Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). *School Science and Mathematics*, *114*(6), 271-279.

[8] Gibson, H. L., & Chase, C. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education, 86*(5), 693-705. doi:10.1002/sce.10039

[9] Tai, R. H., Liu, Q. C., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science*, *312*, 1143–1144. doi:10.1126/science.1128690

[10] National Academies of Sciences, Engineering, and Medicine. (2017). *Supporting students' college success: The role of assessment of intrapersonal and interpersonal competencies*. National Academies Press.

[11] Trochim, W. M. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and program planning*, *12*(1), 1-16. doi:10.1016/0149-7189(89)90016-5

[12] Sunny, C.E. (2018). Stakeholders' conceptualization of students' attitudes and persistence towards STEM: A mixed methods instrument development and validation study (Publication No. 10891760) [Doctoral dissertation, University of Cincinnati]. ProQuest Dissertations and Theses Global.

[13] Coneus, K., Laucht, M., & Reuß, K. (2012). The role of parental investments for cognitive and noncognitive skill formation—Evidence for the first 11 years of life. *Economics & Human Biology*, *10*(2), 189-209. doi:10.1016/j.ehb.2011.01.003

[14] Egalite, A. J., Mills, J. N., & Greene, J. P. (2016). The softer side of learning: Measuring students' non-cognitive skills. *Improving Schools*, 1-14. doi:10.1177/1365480215616313

[15] Reid, K. J. (2009). *Development of the student attitudinal success instrument: assessment of first year engineering students including differences by gender* [Doctoral dissertation, Purdue University]. ProQuest Dissertations and Theses Global.

[16] Reid, K., Imbrie, P. K., Lin, J. J., Reed, T., & Immekus, J. C. (2016). Psychometric properties and stability of the student attitudinal success instrument: the SASI-I. *The International Journal of Engineering Education*, *32*(6), 2470-2486.

[17] Berger, E., Godwin, A., Scheidt, M., Chen, J., Senkpeil, R., Ge, J., ... & Gates, A. (2018, October). Collaborative survey construction for national data collection: Coordination, negotiation, and delivery. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.

[18] Sutkus, J., Carpenter, D., Finelli, C., & Harding, T. (2008, October). Work in progress-Building the survey of engineering ethical development (SEED) instrument. In *2008 38th Annual Frontiers in Education Conference* (pp. S4C-14). IEEE.

[19] Rosas, S. R., & Camphausen, L. C. (2007). The use of concept mapping for scale development and validation in evaluation. *Evaluation and Program Planning*, *30*(2), 125-135. doi:10.1016/j.evalprogplan.2007.01.003

[20] Kane, M. & Trochim, W. M. (2007). *Concept mapping for planning and evaluation* (Vol. 50). London: Sage.

[21] Rosas, S. R. (2012). The utility of concept mapping for actualizing participatory research. *Hispanic Journal of Psychology*, *12*(2), 7-24.

[22] Ewan, L., McLinden, D., Biro, F., DeJonckheere, M., & Vaughn, L. (2015). 131. Mapping the views of adolescent health stakeholders. *Journal of Adolescent Health*, *56*(2), S68-S69. doi:10.1016/j.jadohealth.2014.10.137

[23] Anderson, L. A., Day, K.L., & Vandenberg, A. E. (2011). Using a concept map as a tool for strategic planning: The healthy brain initiative. *Preventing Chronic Disease*, *8*(5), 1-7.

[24] Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (GRIT–S). *Journal of Personality Assessment*, *91*(2), 166-174. doi:10.1080/00223890802634290

[25] Rojas, J. P., Reser, J. A., Usher, E. L., & Toland, M. D. (2012). *Psychometric properties of the academic grit scale.* Lexington, University of Kentucky. Retrieved from http://sites.education.uky.edu/motivation/files/2013/08/PojasPeserTolandUsher.pdf

[26] Pretty, J. N. (1995). Participatory learning for sustainable agriculture. *World Development*, *23*(8), 1247-1263. doi:10.1016/0305-750X(95)00046-F

[27] Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson Education.

[28] Collins, K. M. T., Onwuegbuzie, A. J., & Jiao, Q. G. (2007). A mixed methods investigation of mixed methods sampling designs in social and health science research. *Journal of Mixed Methods Research*, *1*(3), 267-294. doi:10.1177/1558689807299526

[29] Vaughn, L. M., & McLinden, D. (2015). Concept Mapping. In L.A. Jason & D.S. Glenwick (Ed.), *Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods* (pp. 305-314). Madison Avenue, NY: Oxford University Press.

[30] NVivo qualitative data analysis software (2012). QSR International Pty Ltd. Version 10.

[31] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/

[32] Borg, I., & Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). Springer Science & Business Media.

[33] Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (Vol. 11). Beverly Hills, CA: Sage.

[34] Sturrock, K., & Rocha, J. (2000). A multidimensional scaling stress evaluation table. *Field methods*, *12*(1), 49-60. doi:10.1177/1525822X0001200104

[35] De Leeuw, J. D., & Mair, P. (2008). *Multidimensional scaling using majorization: SMACOF in R*. Retrieved from https://escholarship.org/uc/item/6q9542qc

[36] Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). Package 'NbClust'. *Journal of Statistical Software*, 61, 1-36.

[37] Chouikhi, H., Charrad, M., & Ghazzali, N. (2015, June). A comparison study of clustering validity indices. In *Computer & Information Technology (GSCIT), 2015 Global Summit* (pp. 1-4). *IEEE.*

[38] Szaflarski, M., Vaughn, L. M., McLinden, D., Wess, Y., & Ruffner, A. (2015). Using concept mapping to mobilize a black faith community to address HIV. *International Public Health Journal*, *7*(1), 117-130.

[39] Vaughn, L. M., Jacquez, F., & McLinden, D. (2012). The use of concept mapping to identify community-driven intervention strategies for physical and mental health. *Health Promotion Practice*, 1-11. doi:10.1177/1524839912462871

[40] Nunnally, J.C. (1967). *Psychometric theory*. New York: McGraw-Hill Book Company.

[41] Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, (17).

[42] Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.

[43] Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424-453.

[44] de Ayala, R. J. (2009). *The theory and practice of item response theory.* New York: Guilford Press.

[45] Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, *18*(1), 25-34. doi:10.1016/j.jval.2014.10.005

[46] Rosas, S. R., & Ridings, J. W. (2017). The use of concept mapping in measurement development and evaluation: Application and future directions. *Evaluation and Program Planning*, *60*, 265-276. doi:10.1016/j.evalprogplan.2016.08.016

[47] DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*(3), 440-464.

[48] Park, P. (2001). Knowledge and participatory research. In P. Reason & H. Bradbury (EDS.), *Handbook of Action Research: Participative Inquiry and Practice* (pp. 81-90). Thousand Oaks, CA: Sage.

[49] Cooley, E., Beaird, J., & Ayres, R. (1994). Academic persistence and attributional style in fifth graders. *Psychology in the Schools*, *31*(2), 156-163. doi:10.1002/1520-6807(199404)31:2<156::AID-PITS2310310210>3.0.CO;2-S

Appendix: 28-item APT-STEM Survey 2021

Instructions: Please answer the following questions by clicking on one of the bubbles that best describes you. There are no right or wrong answers.	**Strongly disagree**	**Disagree**	**Neither agree nor disagree**	**Agree**	**Strongly agree**
I believe science, mathematics and technology have real-world applications.					
I like to better myself in science and mathematics by seeking opportunities in these areas.					
I like to go above and beyond by exploring science and mathematics outside my courses.					
I am good at science, mathematics, and technology.					
I could one day contribute meaningfully to science and mathematics fields.					
I like to problem solve or find solutions to science and mathematics problems.					
I would like internships which allow me to work with an engineer or a doctor or a scientist.					
I would like to have a career in science, technology, engineering, or mathematics in the future.					
I believe scientists and engineers have made our lives more comfortable through new technologies.					
I believe science, technology, engineering, or mathematics solve problems in society and help people.					
I believe science, technology, engineering or mathematics courses teach critical thinking skills that will help me later in life.					
I like to learn about science and mathematics ideas that support new technologies.					
I like to help my peers with questions while working on science or mathematics problems.					
I like the challenges that mathematics and science courses offer me as it helps me improve my knowledge and skills.					
I like to work with my peers while doing science, technology, engineering, or mathematics problems.					
I like how I can test out my ideas in science, mathematics, engineering, or technology.					
I believe learning science, technology, engineering, or mathematics helps me understand many different ideas.					
I like to come prepared to my mathematics and science courses.					
While working on science or mathematics problems, I don't give up trying even when I am at first unsuccessful.					
I like to work on science and mathematics projects that push me beyond my comfort zone.					
I get better at science and mathematics skills when I practice solving a lot of problems.					
I don't lose focus while working on science or mathematics problems that take a long time to complete.					
Because I have a strong desire for a better life and to learn, it makes me work harder in science, technology, engineering, and mathematics courses.					
I set long term goals and don't get frustrated when it takes a long time to achieve my goals.					
I am responsible for my own learning and experiences.					
I am willing to make sacrifices in order to do well in science, technology, engineering, or mathematics courses.					
I am not discouraged by criticism while working on science or mathematics problems.					
I don't stop believing in myself when I face setbacks while working on science or mathematics problems.					