Development and validation of an instrument to assess students' science, technology, engineering, and mathematics identity

Sijia Liu,[1] Shaorui Xu,[2] Qiuye Li,[1] Hua Xiao,[1] and Shaona Zhou[1,*]

[1]*Guangdong Provincial Key Laboratory of Quantum Engineering and Quantum Materials, National Demonstration Center for Experimental Physics Education, School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, People's Republic of China*

[2]*School of Electronics and Communication, Guangdong Mechanical & Electrical Polytechnic, Guangzhou, Guangdong 510550, People's Republic of China*

(Received 16 December 2022; accepted 18 April 2023; published 5 June 2023)

Despite the high demand for STEM (science, technology, engineering, and mathematics) professionals, the STEM talent pipeline continues to leak. A framework of STEM identity could provide insight into why and to what extent individuals engage in STEM-related activities. The present study describes the process of developing and validating an instrument for assessing students' STEM identity. This instrument was created using the conceptual framework of science identity as its foundation. It aims to conceptualize student STEM identity and extend the STEM identity model. This developmental instrument was examined through expert review, small-scale pilot tests, student interviews, and large-scale field tests. The present study involved 2100 Chinese adolescents from middle and high schools in 10 regions. Factor analysis and Rasch analysis provided evidence of reliability and validity. The final instrument contained four dimensions: recognition, belonging, performance and competence, and interest. Our findings provide multiple sets of evidence for reliability and validity, supporting the appropriate structure of the instrument. It exhibits strong psychometric properties, which can be used to assess sustainability in students' STEM fields.

DOI: 10.1103/PhysRevPhysEducRes.19.010138

I. INTRODUCTION

As modern societies become closely integrated into the development of science and technology, STEM (science, technology, engineering, and mathematics) education has an important effect on our daily lives. The STEM workforce has been globally considered to be the most important way of sustaining and enhancing national competitiveness to deal with the opportunities and challenges of the future [1]. Despite the high demand for STEM professionals, the pipeline of STEM talents continues to leak. The STEM identity of students has been identified as a significant construct that positively influences their learning engagement, career aspirations, and sustainable development in STEM. It has been determined that a STEM identity is the most accurate predictor of high school students pursuing a STEM undergraduate major [2–9]. Identity lenses support individuals in making decisions about learning behavior and career pursuit through the assessment of their own discipline fit [10]. Therefore, cultivating students' STEM identity has been a concern in educational fields [11–15].

Over the past decade, there has been a growing body of literature on identity development in science and science-related disciplines [3,10,16,17]. These include studies of individuals' perceptions of themselves based on their past life experiences, context-specific identity, and multiple identity perspectives (e.g., gender, ethnicity, etc.) [17,18]. Other studies have focused on the impact of identity on future career choices, and researchers have considered how individuals describe who they are and the efforts to be made for future choices [2,4]. Carlone and Johnson posited a framework for a three-dimensional degree of scientific identity, including academic performance, competence, and cognition. These interrelated perspectives constitute the ways in which individuals narrate features of their science identity [10]. Hazari *et al.* suggested interest as a newly construct of identity [19]. Both Hazari *et al.* and Verdín have emphasized that belonging played a key role in students' identity [15,20]. Although qualitative and quantitative research studies in educational contexts posit that individuals identify with STEM congruent with how they identify with other content-based domains [21–25], there is no existing theory-based quantitative model to evaluate students' STEM identity.

*zhou.shaona@m.scnu.edu.cn

To address these issues, the present study proposed a more comprehensive quantitative STEM identity development model that aims to quantitatively describe students' STEM identity built on the previous disciplinary identity framework. Conducting this work recognizes the uneven nature of student identity development at the adolescent level, which leads to a certain percentage of students leaving STEM majors; therefore, the present study addresses the STEM learning experiences of students in grades 7–11 (13–18 years old) to gain a deeper understanding of their STEM identity.

II. LITERATURE REVIEW

A. Identity research

Identity is a complex sociocultural concept that is determined by individuals based on their interactions with others and the environment in which they interact. In the education literature, identity is defined as "being recognized as a certain 'kind of person' in a given context" [3]. Thus, creating a role identity in STEM is a process of recursive social construction. STEM identity is not fixed, and its trajectory may change dynamically with the passage of time [26]. Moreover, STEM identity is socially constructed through students' daily practices [17]. In a situated learning framework, students' life experiences with school and social interactions with others in their family can help them develop and shift their STEM identity [12]. In terms of subject knowledge content, students develop their own understanding and shape an identity of the kind of STEM person they want to be through interaction with the scientific phenomena and social negotiation of the meaning of STEM concepts [11]. Studies show that how students think others see them (perceived other appraisals) depends more strongly on how they see themselves (self-image) rather than how others actually see them (actual appraisal) [27]. In other words, how people think of themselves strongly influences their behaviors as they interact with others.

Identity theory suggests that individuals' identity shapes their choices and behaviors. It may be an inclusive identity that engenders various levels of participation, or it may be a marginalized identity that shows rejection or exclusion [28]. The more a teen sees him- or herself as a "STEM person," the more likely he or she is to continue studying STEM and pursue STEM-related careers in the future. According to expectancy-value theory (EVT) [29], students' achievement-related activities should be explained by the interaction between their beliefs about how well they will do on the activity and the extent to which they value the activity [30]. When students have a strong sense of identity, they invest more in learning-related content, and the development process from participating to thinking that the input is more valuable maintains the sustainability of STEM learning. Previous studies have explored STEM identity's positive relation to outcomes such as persistence and career goals [31]. In addition, students' interactions when building conceptual understanding in the classroom not only shape STEM identity but also influence the quality of learning by developing disciplinary relationships as part of identity [32,33].

B. Conceptualizing STEM identity

A variety of approaches to making sense of identity in STEM contexts have arisen and focused on individuals' perceptions of the kind of person they believe themselves to be in relation to broad disciplines (e.g., science) or specific STEM disciplines (e.g., physics) [4,10]. The present study is to accurately measure the construct of STEM identity based on a systematic review process of STEM identity literature. To conceptualize STEM identity in the study, the relevant literature is evaluated to determine the importance level of the dimension.

Vincent-Ruz and Schunn proposed that an individual's science identity is composed of both the individual's internal view of self and the perceptions of external others. They found that science identity is distinct and separate from other attitudinal constructs [34]. Another oft-cited concept is that of Eccles and his colleagues, who use identity formation as part of an expectation-value model of achievement-related choices. Various social and psychological factors (such as personal beliefs, experiences, and abilities) influence individuals' expectations for success and the value they place on available task options. All of the above factors influence the decisions and choices individuals make, such as enrolling in STEM courses or pursuing STEM careers [35–37]. Gee holds identity to be an analytic lens for research in education. Informed by Gee's identity theory, Carlone and Johnson originally proposed that science identity was a three-dimensional construct comprising students' competence (refers to the ability to understand science content knowledge), recognition (which means one is recognized by oneself and/or by others as a science person), and performance (as the social performances of relevant scientific practices in the public area and culture of science) [3,10]. Hazari *et al.* first expanded the model of Carlone and Johnson through supplementing the interest dimension and combining performance and competence as one dimension [19]. Additionally, the modified model included the construct of physics "interest," which aligns with Bandura's adaptation of social cognitive theory (SCT). The SCT posits a relationship between constructs related to self-efficacy (such as performance competence) and those related to interest [38]. For students who have not yet committed to a particular major or career, their interest may be critically important in influencing the decisions of what kind of person they want to be. The model has also been validated in different STEM disciplines, including biology and chemistry [39], mathematics [40], engineering [41], and computer science [42]. Verdín examined how developing an engineering identity through the interplay between interest, recognition, and performance and competence beliefs and establishing a sense of belonging support students' persistent beliefs in engineering [20]. The results suggest that when students experience a sense of belonging, they are more likely to develop a stronger sense of identity. Hazari *et al.*, who used similar identity constructs in the physics context, found that for women physicists in their senior year (i.e., the fourth year of college or greater), a sense of belonging supports their belief in themselves as a physics person [15]. Therefore, four dimensions were involved in the framework of students' STEM identity: (i) recognition, (ii) competence and performance, (iii) interest, and (iv) sense of belonging.

(i) Recognition is both an external manifestation and an internal state required for identity development [10,39]. Gee noted that individuals develop their identity when they are recognized by themselves or others in a particular context [3]. It assesses students' perceptions of how well they and others are recognized as STEM people. Moreover, a person's recognition of him- or herself and others is based on his or her ability to see his or her ability and performance in a specific field [20]. (ii) Performance and competence belief measure students' confidence in their ability to understand STEM content and knowledge, expectations for success in STEM learning, and belief in performing well in the STEM learning process [43]. Competence and performance closely resemble beliefs of self-efficacy, which are students' assessments of themselves. However, "perceived self-efficacy is an important contributor to performance accomplishments, whatever the underlying skills might be." Because the two often appear together in previous studies, they are grouped into subdimensions of identity. (iii) Interest is a crucial part of identity and shows students' desire to acquire STEM knowledge and is a psychological state with emotional attributes and cognitive factors. In terms of identity, STEM interests are expressed as persisting in STEM learning, having a positive attitude toward STEM careers, or having career aspirations to a certain extent [12,19,44]. (iv) Finally, a sense of belonging depends on students' perceived social support on campus, a feeling or sensation of connectedness, and experiences of mattering or feeling cared about, accepted, respected, valued by, and important to the group (e.g., campus community) or others (e.g., faculty, peers) on campus [45]. The sense of belonging in STEM enhances students' STEM learning motivations and makes students feel comfortable integrating into the STEM learning environment, which is conducive to establishing STEM learning tendencies and maintaining STEM learning. Thus, belonging is also a crucial component of identity.

C. Need for developing a new instrument

Although many studies recognized the importance of STEM identity, it still needs to make a consensus on the concept and construct. Therefore, this particular identity domain needs to be defined, which is not only operationalized at the middle or high school level but also based on traditional identity theory [13]. Second, in the research on students' STEM identity or science identity, many existing instruments have equated identity with students' self-concepts (e.g., STEM "kind of person"). Certainly, a brief one- or two-item measure of STEM or science self-concept could reflect a portion of students' STEM or science identity and be useful for researchers who need a quick measure of identity. Nevertheless, the number of items in a measure is a function of reliability [46]. Thus, a single item is generally not the ideal way to measure educational- or psychological-related constructs. Furthermore, equating STEM self-concept and STEM identity hinders our understanding of the more extensive nature of the STEM identity construct. Third, most identity surveys have been developed with a traditional psychometric approach, such as calculating item-total correlations or exploratory factor analysis. Individuals' scores and item difficulty depend on each other, so a change in the respondent sample will influence the item statistics [47]. To solve these problems, science education researchers have called for using the Rasch model to develop attitudinal scales [47,48]. However, there is currently no tool developed or modified by the Rasch model to measure the STEM identity of young students. Finally, most studies focus on elementary school students and college students, without paying sufficient attention to middle school and high school students. However, for STEM identity, young children, who comprised preschool and elementary school, have difficulty in understanding the concept of identity because their cognitive abilities are insufficient, and it is difficult for students to have a clear idea of their future careers and identity. On the other hand, in college, students have already chosen their majors, and even if they have a low sense of identity, only a small percentage of students are able to change their majors. In contrast, adolescent students gradually have a clear understanding of identity. The adolescent student population consists mainly of middle school and high school students. At the beginning of middle school, students acquire progressively more knowledge and begin to develop a deeper understanding of the subject matter. In high school, they can choose their future majors or careers more autonomously. Identity is now receiving increasing attention. It is reported that the middle and high school years are critical times for students to develop STEM identity [13]. Therefore, there is an urgent need for a more traditional, theoretically grounded, psychometrically compliant measure of STEM identity.

D. Aim of the research

The aim of this research is to construct the STEM identity model based on previous studies. According to the model, an instrument was developed to measure the STEM identity of adolescent students. The instrument was combined with the robustness of Carlone and Johnson's, Hazari *et al.*'s, and Verdín's identity models across various research contexts [4,10,20]. The model of the instrument provides utility evidence for our guiding personal and social identity frameworks [49,50].

Specifically, we define the concept and constructs of the STEM identity measured in the study based on a solid theoretical framework. Items were selected and revised through a literature review, expert reviews, student interviews, small-scale pilot tests, and large-scale field tests. It is to pro-

TABLE I. Comments, advice on the validity of the survey's face, and modifications.

Evaluators	Comments and feedbacks	Modifications
Expert A	Regarding the item in the interest scale that reads "I want to engage in STEM-related work in the future," Its way of asking questions is too direct.	It was modified to "I'm interested in a career in science"
Expert B	The item in the performance and competence scale that reads "I am confident to learn STEM subjects well and get good grades in high school." refers to students' recognition rather than their competence beliefs.	This item was moved to the recognition dimension
Expert C	In the item in the recognition scale that reads "My classmates often ask me about science and technology," classmates recognize themselves but do not always ask questions	This item was redundant with "Classmates think I'm good at STEM discipline and was removed.
Teacher A	The item in the belonging scale that reads "I will follow news reports related to STEM fields on TV or online." is more like a description of a phenomenon.	This item was removed.
Teacher B	The item in the interest scale that reads "I like to read STEM-related books" suggests that using only books restricts their experiences.	It was modified to "I would like to learn more about STEM from various sources."
Teacher C	The concept of "a STEM person" in the scale may not be understood by students.	It was replaced with "fit for study STEM"

vide sufficient reliability and validity evidence using exploratory factor analysis (EFA) and Rasch analysis. Construct validity will be ensured by confirmatory factor analysis (CFA) and Rasch analysis. The present study intends to develop and evaluate the reliability and the validity of a STEM identity measurement instrument for adolescent students.

III. METHOD

The instrument was developed through the following procedures: development of the item pool, expert reviews, pilot test and student interviews, field test, and validation of the instrument. With the expert reviews and student interviews, the item pool was modified and put into the large-scale field test to ensure that the instrument satisfied the criteria of reliability and validity.

A. Development of the item pool

The concept and construction of STEM identity guided the development of the item pool. According to an extensive review of the theoretical literature and empirical studies, the essential components embodying students' STEM identity were found and adopted to define the four scales of this instrument: (i) recognition, (ii) performance and competence, (iii) interest, and (iv) sense of belonging [10,12,43,51]. The item pool was developed based on previous literature that developed or modified instruments, such as the Student Science Identity Questionnaire (SSI) [12]. The initial item pool was arranged on a five-point Likert scale to which respondents may respond in one of five ways: strongly agree, agree, not sure, disagree, strongly disagree, since most of the teenagers can understand and answer this type of question more easily. As this instrument was designed to be applied in the Chinese context, its items were first written in English and then translated into Chinese. The items include recognition (7 items), performance and competence (10 items), interest (6 items), and sense of belonging (7 items) (see the Appendix).

The initial items were submitted to three experts who studied STEM education and three experienced STEM teachers. To ensure content validity, three experts were asked to evaluate each item according to the wording and the definition of the dimensions to which it belonged to. A summary of their comments and suggestions as well as corresponding improvements in items are shown in Table I.

B. Small-scale pilot test and interviews

A small-scale pilot test with student interviews was conducted for the initial item pool. A total of 50 grade 7 students participated in the small-scale interview. These students were divided into groups of 10 for five sessions, with the same teacher inviting the same group of 10 people into the classroom at one time for group interviews to ensure that feedback was provided on each project. They were allowed to ask any questions if needed. Considering that students were able to understand the scale more broadly, efforts were made in the present study to ensure the minimum age (grade 7) of the first small sample of interviews. Based on the information obtained, items that did not work well or might confuse students were modified. For example, "I have difficulty digesting new knowledge in STEM" was modified to "It takes me a long time to understand new knowledge in STEM-related subjects," and "I will learn about engineering and technology in a variety of ways" was revised to "I will learn more about STEM through various sources of information." The remaining items were simply revised for individual words based on student feedback.

C. Participants

The participants in the present study were drawn from ten districts in southern and northeastern China for two large-scale field tests. Samples 1 and 2 each contain five different districts and sample 3 includes eight of them. The general principle for choosing students was (i) every sample involved students from southern and northeastern China and (ii) within one sample, students from southern or northeastern China were from two districts or more for each. Sample 1 was used for the first large-scale test and included 360 students from five districts with varying levels of academic achievement in STEM areas. From the 360 students, 357 valid questionnaires were collected. The second round of large-scale testing consisted of two sample groups, with the sample 2 data used for EFA to determine the structure of the instrument and included reports from 570 students from another five districts, with 552 valid questionnaires collected. Data from sample 3 were used for Rasch analysis and CFA to check the dimensional structure and reliability of the instrument and included reports from 1070 students from eight districts, collecting 1043 valid questionnaires. The gender and age stratification of students are shown in Table II.

In China, primary school students generally acquire scientific knowledge through formal learning of science course at school. From secondary school, science course begins to split into multiple disciplines. Students start to learn biology in the first year, physics in the second, and chemistry in the third. When students enter the first year of high school, the science curricula mainly include independent disciplines of physics, chemistry, and biology, in addition to the three necessary subjects of mathematics, Chinese, and English. By the end of the second year of high school, students are faced with a choice between science or liberal arts. This choice largely determines whether they choose STEM-related subjects when they enter college. It can also be said that this stage is the critical process of STEM pipeline leakage. Therefore, it is of great significance to study the STEM identity of adolescents from middle school to high school.

D. Large-scale field test and data analysis

For the two large-scale tests, IBM SPSS 24.0, Winsteps 3.81.0, and Conquest 2.0 software were used. Based on the data obtained, factor analysis and Rasch analysis were used to provide evidence for reliability and validity. In the first round of the large-scale field test, an EFA was first conducted on sample 1 to provide more information on structural validity. Rasch analysis was then undertaken to provide information on item-level modifications. Sample 2 from the second round of testing was subjected to EFA and Rasch analysis to determine the instrument structure. Then, a CFA was conducted using sample 3 to achieve satisfactory psychometric properties.

IV. RESULTS

A. Results of the first round of the large-scale field test

Data collected from the pilot study were used to conduct EFA by SPSS. It is used to explore the dimensionality of the scale and identify items and factors that did not fit. First, the multivariate normality and sampling adequacy of the data were tested. Bartlett's test of sphericity indicated that $\chi^2 = 6630.048$, which was statistically significant ($p < 0.001$). The Kaiser-Meyer-Olkin (KMO) measure of adequacy was high (KMO $= 0.941$). Based on Pallant's criteria, when the value of KMO is higher than 0.60, the recommended value, researchers can continue the further factor analysis [52]. Principal component analysis followed by oblique rotation (Promax rotation with Kappa $= 4$) was used because the potential factors in the scale were assumed to correlate with each other. A loading size of greater than 0.4 was considered acceptable. Cross loading was considered if the loading size was larger than 0.3 in more than one factor.

Considering the low explained variance and the theoretically different definitions of these dimensions, four factors (according to our theoretical constructs of STEM identity) were extracted and the results were good. The explained variance was 57.673%, and the items were categorized in accordance with the presumptive dimensions developed from the extensive literature review. The final results, with loading less than 0.4 not shown, factor loading scores were above 0.40 for all items except I6 and P8 and below 0.50 for R5, P4, P7, P10, and I5. Therefore, it was decided to delete these items. As shown in Table III, after removing the above inappropriate items, the explained variance rose to 62.967% (greater than 60%, satisfying the requirement), and the factor loadings for all dimensions were above 0.50, which is at a good level [53]. Factor 1 is recognition, factor 2 is performance and competence, factor 3 is interest, and factor 4 is the sense of belonging.

TABLE II. Participant demographics.

	Age						Gender		
	13	14	15	16	17	18	Boy	Girl	Total
Sample 1	9.5%	23.8%	18.8%	14.8%	26.9%	5.0%	53.5%	45.4%	357
Sample 2	9.1%	13.4%	27.4%	12.0%	29.2%	8.9%	55.7%	44.3%	552
Sample 3	1.4%	24.4%	18.6%	20.1%	27.0%	8.5%	50.0%	49.6%	1043

TABLE III. Factor loading in the first round of the field test.

		Factor loading			
Dimension	Item	1	2	3	4
Recognition	R2	0.884			
	R7	0.848			
	R3	0.843			
	R4	0.807			
	R1	0.804			
	R6	0.782			
	B7	0.699			
	P9	0.628			
	P5	0.582			
Performance and competence	P1		0.810		
	P3		0.729		
	P2		0.716		
	P6		0.537		
Interest	I1			0.777	
	I3			0.776	
	I4			0.688	
	I2			0.658	
Belonging	B2				0.906
	B3				0.886
	B4				0.737
	B6				0.697
	B1				0.688
	B5				0.660

TABLE IV. Correlation or correlation matrix among dimensions in the first round of the field test.

	Dimension			
Dimension	1	2	3	4
1		1.520	1.205	1.754
2	0.685		1.042	1.068
3	0.628	0.786		0.985
4	0.700	0.618	0.659	
Variance	3.217	1.532	1.146	1.950

Rasch analysis was then conducted on the 23 items generated by the EFA. Given the multidimensional nature of the instrument and the fact that all items were rated on a scale of 5, a multidimensional rating scale model under the multidimensional random coefficient multinomial logit model (MRCMLM) framework was selected and data were analyzed using Conquest 2.0. In addition, a one-dimensional rating scale model was applied to each dimension using Winsteps 3.81.0 to ensure the unidimensionality of each subdimension. In general, the Rasch model is a single-parameter model and unidimensionality was a prerequisite for items to meet the requirement. The first eigenvalues for the unexplained variance in the four dimensions reported in the principal component analysis (PCA) were 1.8, 1.6, 1.5, and 2.1. It showed that all four dimensions possessed unidimensionality.

TABLE V. Fit statistics in the first round of the field test.

Dimension	Item	Estimate	Model standard error	Unweighted MNSQ	Weighted MNSQ
Recognition	R1	−0.237	0.052	0.69	0.71
	R2	0.041	0.052	0.82	0.85
	R3	0.204	0.053	0.84	0.84
	R4	0.454	0.052	0.92	0.95
	R6	−0.730	0.053	1.16	1.22
	R7	0.287	0.052	0.87	0.87
	P5	0.102	0.052	1.05	0.98
	P9*	−0.248	0.148	1.63	1.43
	B7	0.128	0.053	1.18	1.17
Performance and competence	P1	−0.013	0.048	1.08	1.06
	P2	−0.028	0.050	0.89	0.89
	P3	−0.321	0.050	0.85	0.85
	P6*	0.362	0.085	1.16	1.13
Interest	I1	0.235	0.048	1.07	1.07
	I2	0.025	0.048	0.92	0.91
	I3	0.324	0.047	1.01	1.01
	I4*	−0.584	0.082	1.21	1.18
Belonging	B1	−0.154	0.046	1.13	1.13
	B2	−0.254	0.046	1.03	1.01
	B3	−0.070	0.042	1.36	1.30
	B4	0.117	0.044	1.01	1.01
	B5	0.166	0.044	0.93	0.92
	B6*	0.196	0.099	0.87	0.86

An asterisk next to a parameter estimate indicates that it is constrained. Separation reliability = 0.968, Chi-square test of parameter equality = 523.84, d.o.f. = 19, significant level <0.001.

To further test the single dimension, the Rasch model residual PCA method was used to test the dimension of the form. The characteristic value of the standardized residual of the first factor and the variance in the measurement interpretation are important indicators to measure the uniformity of the data structure. The range of the characteristic value of the standardized residual of the first component should be between 1.4 and 2.1. The higher the variance explained by Rasch model scores, the higher the likelihood that items measure the same dimension [54]. Table IV shows the intercorrelations among the four subscales for the multidimensional results. All the correlations between dimensions were moderate. Thus, the instrument measured closely related multidimensional constructs.

The model-data fit was assessed using unweighted and weighted mean square (MNSQ). The unweighted and weighted mean square MNSQ for all items were within the acceptable range (i.e., 0.6–1.4), except for P9; thus, item P9 was removed and 22 items remained (Table V). The expected a-posteriori/plausible value (EAP/PV) reliabilities for each dimension were all acceptable with the data of 0.927 (recognition), 0.863 (performance and competence), 0.804 (interest), and 0.847 (sense of belonging). An asterisk next to a parameter estimate indicates that it is constrained. Separationreliability = 0.968, Chi-square test of parameter equality = 523.84, d.o.f. = 19, significant level <0.001.

For the function of the five-point response category, a minimum of ten observations were made for each category ($N > 10$). The results showed that the step estimates increased monotonically, and the distance satisfied the 0.81–5 logit requirement without peak submergence. At the same time, the average measurement increases evenly

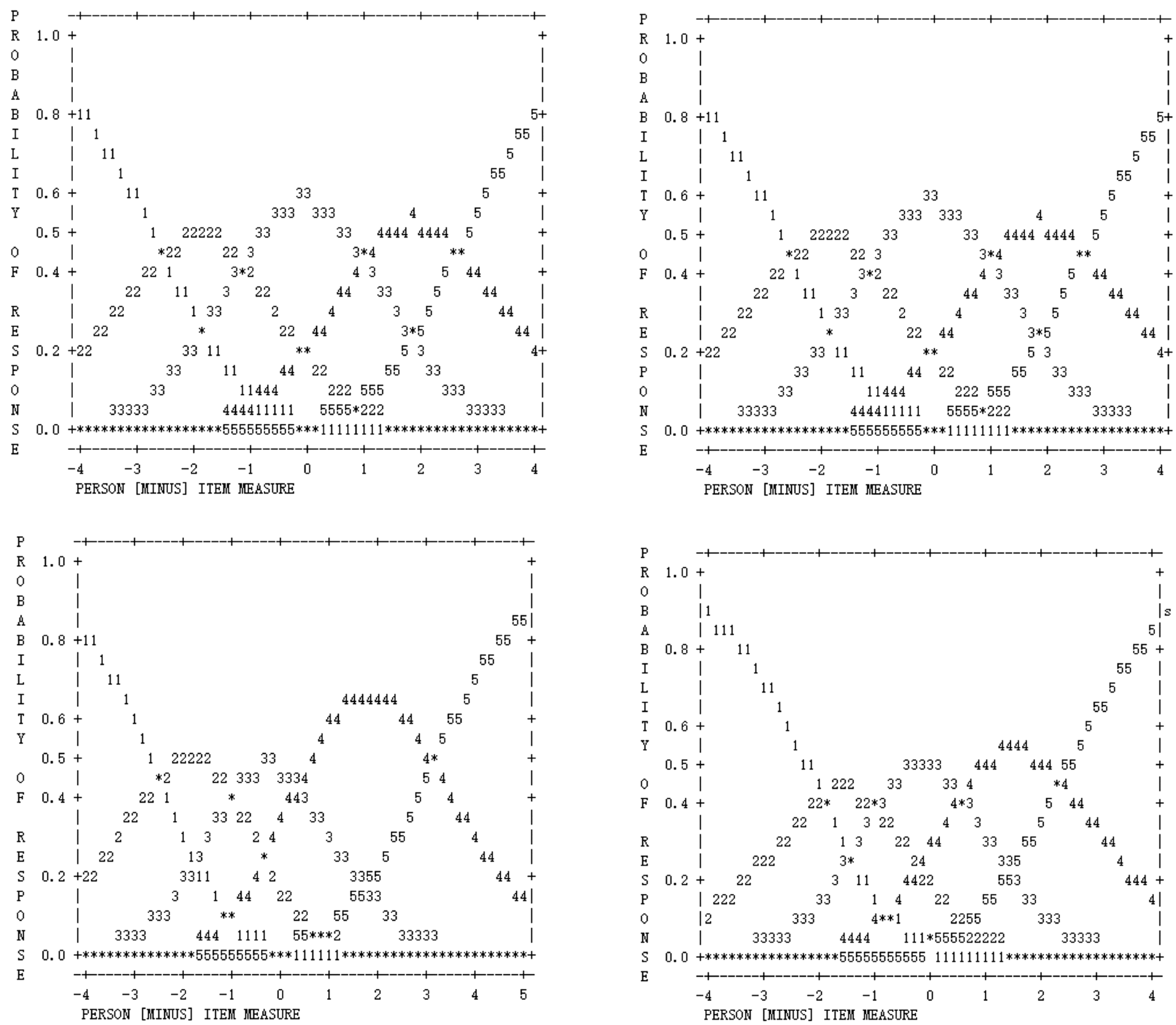

FIG. 1. Probability curves for subdimension of the instrument in the first round of the field test.

between adjacent options, indicating that the five-point scoring method of each subdimension is consistent with the overall distribution of items and participants. Figure 1 provides the probability curves for each subdimension. This evidence suggests that the five-point scale is suitable.

The Wright map shows the distribution of items and students along a one-logits scale with students' ability estimates on the left (represented by X) and item difficulties on the right (shown by the item number; Fig. 2). The students at the top have more positive STEM identity, while the items at the bottom are less endorsable. For the five-point scale, there were four thresholds for each item. For example, 1.1, 1.2, 1.3, and 1.4 were the four thresholds for item 1. They indicated the location at which the probability of achieving a higher category was 0.5. According to the Wright map, there is a discontinuity in the difficulty of the questions, and the three levels of thresholds are not coherent. There is also a discontinuity in the difficulty of the dimensions, which are not evenly distributed. For example, the difficulty of threshold 4 of the recognition dimension is concentrated above, while the difficulty of threshold 1 is generally lower.

```
MAP OF LATENT DISTRIBUTIONS AND THRESHOLDS
==============================================================================
                      Dimension                 Generalised-Item Thresholds
---------------------------------------
         1         2         3         4
------------------------------------------------------------------------------
                |         |         |         |
                |         |         |         |4.4 6.4 8.4
                |         |        X|        X|3.4
  3            X|        X|        X|        X|1.4 7.4 11.4 13.4
               X|        X|        X|        X|2.4 10.4 12.4 14.4
              XX|        X|       XX|        X|5.4 16.4
              XX|        X|       XX|       XX|15.4 22.4 23.4
  2           XX|       XX|      XXX|       XX|18.4 19.4 21.4
             XXX|      XXX|     XXXX|      XXX|
             XXX|     XXXX|     XXXX|     XXXX|3.3 4.3 17.4
             XXX|   XXXXXX|  XXXXXXX|    XXXXX|6.3 20.4
  1         XXXX|   XXXXXX|  XXXXXXX|   XXXXXX|2.3 7.3 8.3
            XXXX| XXXXXXXX| XXXXXXXX| XXXXXXXX|13.3 16.3 18.3
            XXXX| XXXXXXXX| XXXXXXXX|  XXXXXXX|1.3 11.3 14.3 15.3 21.3 22.3
           XXXXX|  XXXXXXX|XXXXXXXXX|  XXXXXXX|10.3 20.3 23.3
  0        XXXXX| XXXXXXXX|XXXXXXXXXXXXXXXXXXX|5.3 12.3 17.3 19.3
          XXXXXX|  XXXXXXX|   XXXXXX|  XXXXXXX|
           XXXXX| XXXXXXXX|    XXXXX|  XXXXXXX|13.2 16.2 20.2 22.2
           XXXXX|    XXXXX|     XXXX|    XXXXX|14.2 15.2 21.2 23.2
 -1        XXXXX|     XXXX|      XXX|      XXX|4.2 7.2 10.2 11.2 19.2
            XXXX|      XXX|       XX|       XX|1.2 2.2 6.2 8.2 12.2 18.2
             XXX|       XX|        X|       XX|3.2 17.2 20.1 21.1 22.1 23.1
           XXXXX|        X|        X|       XX|14.1 16.1
 -2          XXX|        X|        X|         |13.1 18.1
             XXX|        X|        X|         |4.1 5.2 10.1 15.1 17.1 19.1
              XX|         |         |        X|6.1 7.1
              XX|         |         |         |2.1 3.1 11.1 12.1
 -3            X|         |         |         |8.1
               X|         |         |        X|1.1
               X|         |         |         |5.1
                |         |         |         |
 -4            X|         |         |         |
                |         |         |         |
               X|         |         |         |
                |         |         |         |
==============================================================================
Each 'X' represents   4.1 cases
The labels for thresholds show the levels of
    item, and step, respectively
==============================================================================
```

FIG. 2. Wright map for the instrument in the first round of the field test (red fonts, green fonts, black fonts, and blue fonts are item thresholds for recognition, performance and competence, interest, and sense of belonging).

B. Additional insights from EFA and Rasch analysis for improving the instrument

According to the results of EFA after deleting items, the instrument (23 items) met the acceptable standards of good psychometric properties. The results from Rasch analysis showed that the degree of separation, unidimensionality, and response category had acceptable indicators. Based on the results of the MNSQ, inappropriate questions were removed (P9), leaving 22 items. The Wright map (Fig. 2) suggested that further improvements are needed. This resulted in more appropriate difficulty for each dimensional topic. According

TABLE VI. Factor loading in the second round of the field test.

		Factor loading			
Dimension	Item	1	2	3	4
Recognition	R1	0.860			
	R2	0.856			
	R4	0.852			
	R6	0.839			
	R3	0.830			
	R7	0.824			
	R8	0.728			
	B7	0.720			
	P5	0.597			
Performance and competence	P1		0.820		
	P11		0.757		
	P3		0.722		
	P2		0.648		
Interest	I2			0.809	
	I1			0.801	
	I3			0.751	
	I4			0.687	
	I7			0.574	
	I8			0.503	
Belonging	B2				0.905
	B3				0.872
	B4				0.701
	B1				0.669
	B5				0.665
	B6				0.598

to the item difficulties, item P6 was inappropriate with high difficulty. Therefore, item P6 was removed, leaving 21 items.

Given that it is better to control for scale length based on the results of data fitting and item estimation, four new items were developed and added to the second round of testing. With the item changes, a total of 25 questions were incorporated into the second round of large-scale testing. The four new items were also drawn from existing instruments [12,13,55]. Specifically, item R8 was referenced from questions 19 and 35 in the STEM identity survey [55]. Item P11 was referenced from C6 in SSI [12]. Item I7 was derived from items V4 and V30 in SciID [13]. Item I8 was derived from I8 in SSI [12]. Some expressions of the items were modified so that the distinctions between the option thresholds were close. For example, "I would get satisfaction from a course in a STEM-related specialist subject compared to other courses" rather than "I like STEM-related subjects." The whole instrument with additions and deletions of items has been described in the Appendix. Specifically, there were 9 items in the recognition dimension, 4 items in the performance and competence dimension, 6 items in the interest dimension after modification, and 6 items in the belonging dimension (see the Appendix for descriptions).

C. Results of the second round of the large-scale field test

First, sample 2 was selected for EFA again in the second round of testing, and the results showed KMO $= 0.941, \chi^2 = 7698.814$ ($p < 0.001$). Continuing with the four factors consistent with the above, the total explained variance was 61.02%. As shown in Table VI, the factor conformity for each dimension is above 0.50, which is at a good level.

Rasch analysis was conducted second to determine whether the problems found previously had been solved. Item selection was mainly based on the data fit statistics and the Wright map, with consideration of item context. For the finalized instrument, the unidimensional Rasch analysis conducted in each dimension indicated their unidimensionality, with the first eigenvalues in the PCA being 1.8 (recognition), 1.7 (performance and competence), 1.7 (interest), and 2.0 (belonging). According to that, the eigenvalue of the belonging dimension is the critical value of 2.1. We made the unidimensionality of it more certain in the second round. For model-data fit, the unweighted and weighted MNSQ of all items fell within the acceptable range (Table VII). In addition, the EAP/PVof each dimension was 0.883, 0.823, 0.802, and 0.814. The separation and reliability of the student

TABLE VII. Rasch fit statistics in the second round of the field test.

Dimension	Item	Estimate	Model standard error	Unweighted MNSQ	Weighted MNSQ
Recognition	R1	−0.044	0.030	0.74	0.75
	R2	0.119	0.030	0.79	0.81
	R3	0.172	0.031	0.76	0.78
	R4	0.046	0.031	0.84	0.87
	R6	0.052	0.030	0.79	0.79
	R7	0.113	0.030	0.84	0.85
	R8	−0.735	0.030	1.05	1.06
	B7	0.106	0.031	1.13	1.10
	P5*	−0.243	0.086	1.19	1.15
Performance and competence	P1	−0.030	0.030	1.19	1.16
	P2	0.104	0.031	0.89	0.89
	P3	−0.348	0.031	0.95	0.96
	P11*	0.274	0.052	1.21	1.19
Interest	I1	0.238	0.027	1.04	1.03
	I2	0.051	0.027	0.88	0.86
	I3	0.315	0.027	0.99	0.99
	I4	−0.806	0.028	1.27	1.30
	I7	0.143	0.027	1.02	1.01
	I8*	0.059	0.061	0.96	0.95
Belonging	B1	−0.229	0.026	1.37	1.30
	B2	0.044	0.027	0.97	0.95
	B3	0.057	0.027	0.86	0.85
	B4	0.085	0.027	0.86	0.86
	B5	−0.227	0.027	1.21	1.21
	B6*	0.270	0.060	0.92	0.92

An asterisk next to a parameter estimate indicates that it is constrained. Separation reliability = 0.991, Chi-square test of parameter equality = 2258.11, d.o.f. = 21, significant level <0.001.

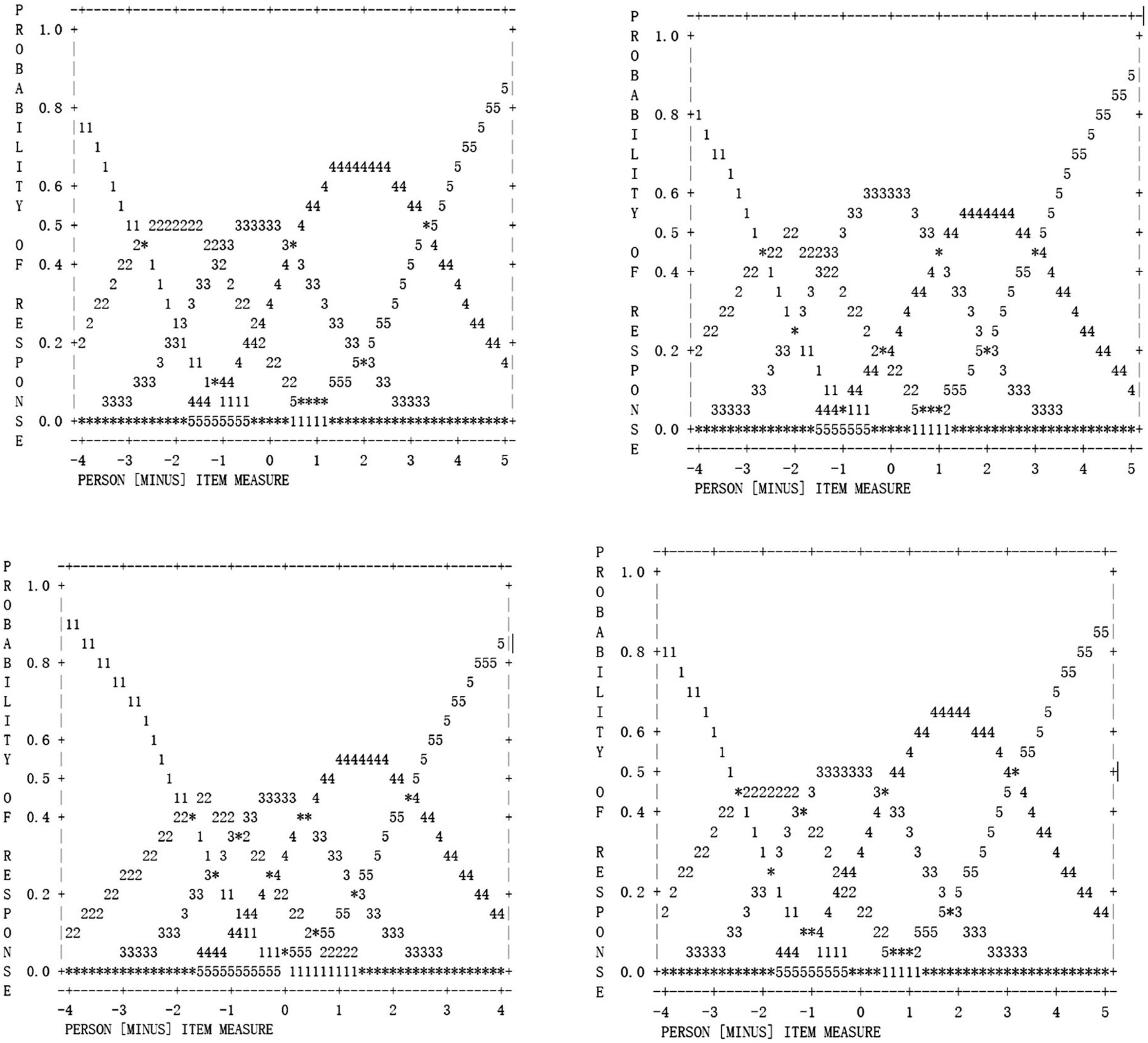

FIG. 3. Probability curves for subdimension of the instrument in the second round of the field test.

sample were also analyzed. The results show that the reliability of questions and participants is sufficient.

For the category function, each had a minimum of ten observations for every item ($N > 10$). The step estimates showed a monotonic increase, resulting in the distinct peaks of the probability curves of each category (Fig. 3). These observations suggest that the five-point response category was used optimally. Sufficiently modified topics did not make a difference.

The Wright map (Fig. 4) shows that the item difficulty was reasonably distributed with the thresholds covering most of the students' abilities, which means that the items target the sample well. There was a threshold break in the questions in the first large-scale test while the distribution of questions is more balanced after the modification in the second round.

Finally, to confirm the structure of the four-factor model from EFA, we administered this instrument to sample 3. Confirmatory factor analysis was conducted using Mplus 8.0 to verify the construct and ensure that the finalized instrument met the criterion. The weighted least squares mean and variance adjusted estimator option was used to deal with categorical and non-normal data to achieve a robust result. The results showed a good fit, with a root mean square error of approximation (RMSEA) of 0.045, a comparative fit index (CFI) of 0.901, and a Tucker-Lewis index (TLI) of 0.913, while the ideal values were as follows: RMSEA < 0.08, CFI > 0.9, TLI > 0.9 [56,57]. Additionally, the factor loadings of each item per subdimension were all above 0.6 (Fig. 5).

V. DISCUSSION

This research aimed to develop an instrument to measure STEM identity among adolescent students through pilot and field tests. The results indicated that the instrument was generally of high quality. To accomplish these

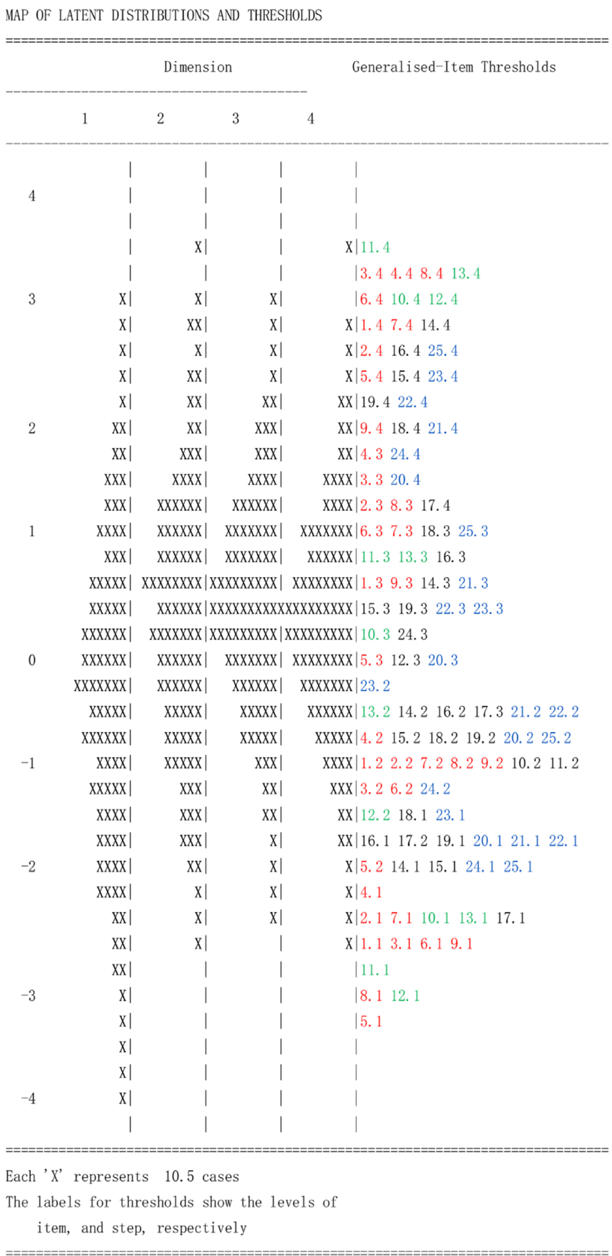

FIG. 4. Wright map for the instrument in the second round of the field test (red fonts, green fonts, black fonts, and blue fonts are item thresholds for recognition, performance and competence, interest, and belonging).

goals, we carefully reviewed the relevant structures for assessing the STEM identity of students, developed some items, and revised items according to opinions from experts and STEM teachers. The construct validity and reliability of items from the instrument were determined through EFA and Rasch analysis with sample 1. The results of EFA showed that the 23 items could be extracted into four salient factors. According to Rasch analysis, logit scores can also show students' STEM identity in various dimensions. We identified four significant factors: recognition, performance and competence, interest, and sense of belonging.

To confirm the fit of the four-factor model, EFA and Rasch analysis were reperformed on the final 25 items using the data from sample 2 to confirm the dimensional fit. The Rasch analysis verified the reasonable distribution of difficulty of the revised items and the optimal use of the five-point response category. CFA was conducted using sample 3 with factor loading values greater than 0.6. These results supported the idea that the development of STEM identity is a complex and multidimensional process that requires all dimensions to be adequately nurtured and developed [58].

The four aspects of STEM identity were consistent with prior work on identity development [10,15,19,41]. The structure of this research inherited the results of Carlone and Johnson and accepted the "interest" structure mentioned by Hazari *et al.* and the results of belonging obtained in Verdín [4,10,20]. This result echoed the research results of Hazari *et al.* and expanded the research group of adolescent students. It supported the suitability of the identity framework as a critical analysis lens for students interested in STEM. Studies have shown that one of the key areas found to influence persistence in the STEM discipline is the role of recognition. How others see a student is vitally important to how the student sees her- or himself and to her or his subsequent choices. Adolescent students are easily influenced by the perceptions of significant others in their lives regarding STEM, which can determine their subsequent academic choices. Many studies have shown that parents' perceptions and expectations regarding their children's STEM abilities influence the children's self-perceptions and expectations in the STEM disciplines and even have a non-negligible influence on the subsequent decision to pursue a STEM career [59,60].

Another factor that repeatedly arises in the literature addressing STEM identity is performance and competence. Marsh *et al.* found a reciprocal effect between performance and self-concept; that is, earlier performance affects academic self-concept and prior self-concept affects future performance [61]. In addition, studies show that performance and ability affect students' self-efficacy and individuals' judgments of their abilities to perform tasks successfully [62].

In addition to recognition and performance and competence, affect (e.g., interest) is a strong predictor of STEM identity [63]. Social cognitive career theory (SCCT) also supports this finding [64]. Some researchers have proven that in college students, there is an interaction between interest and performance and competence, which jointly affects identity [38,41,65]. Students' performance and competence are necessary but not sufficient for their construction of STEM identity. Students who feel able to perform and are competent need to be recognized or become interested to see themselves as "STEM persons" [15,40,41].

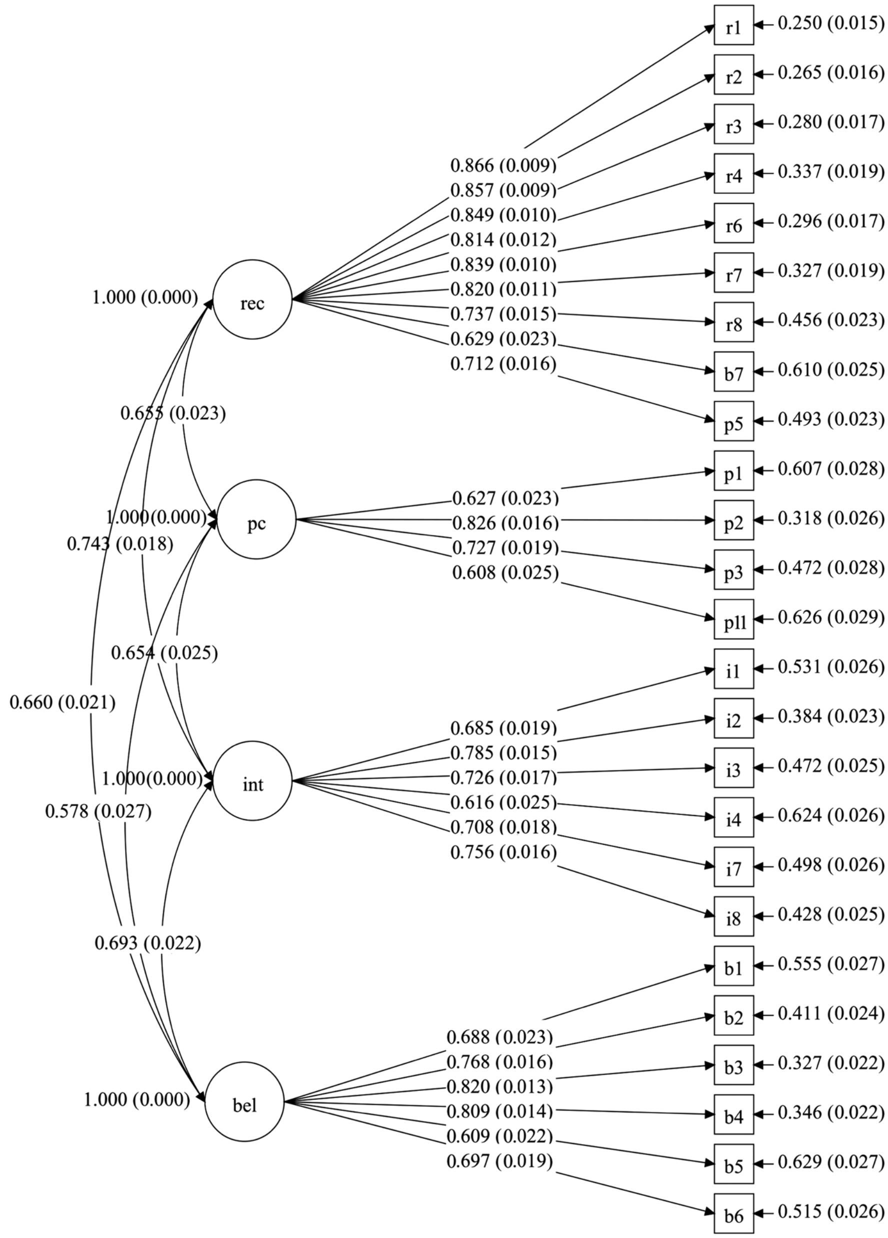

FIG. 5. Confirmatory factor analysis in the second round of the field test. Note: rec: Recognition, pc: performance and competence, int: interest, bel: belonging.

Previous work has demonstrated that identity construction is fluid in that it can change from moment to moment and from context to context [17,66–69]. When students interact with others or are influenced by the environment, they generate self-construction of STEM identity. Given the fluidity of disciplinary identity, it is also likely that how different constructs (e.g., performance and competence, interest, and recognition) influence identity will change with different contexts. Therefore, new constructs, such as a sense of belonging, will emerge as necessary for disciplinary identity construction. Although the sense of belonging is related to recognition, it is theoretically distinguishable from it. Others may recognize students as STEM persons (teachers, parents, or peers), but they still believe that they are unsuitable for STEM discipline. Other stereotypes of other identity attributes usually affect the sense of belonging, such as gender. To accommodate these disruptions, some women have been found to suppress parts of their discipline identity [70]. Regardless of acts of suppression, previous research has found that college students with identity as members of underrepresented groups still experience a lower sense of belonging than their peers. This lower sense of belonging has been found to have negative consequences for academic self-efficacy, valuing of academic tasks, and performance in disciplinary coursework, which likewise can affect students' disciplinary identity construction.

The major contribution of the present study lies in developing and validating an instrument that can be used to assess adolescent students' STEM identity. We considered this to be more relevant in the Chinese context. The final version of the instrument has high construct validity and reliability when used among students in grades 7–11 from ten regions of northern and southern China. Theoretically, with the development of the instrument, our work has provided a more complete structure of STEM identity, with multiple pieces of evidence to reveal the dimensionality. With the data obtained for the final version of the instrument, as well as considering the students and their cultural background, it is reasonable to divide identity into four potential dimensions, including recognition, performance and competence, interest, and sense of belonging. The multidimensionality of the data was established by both factor analysis and Rasch analysis, which reinforced the notion that students' identity toward STEM is multidimensional. The four dimensions in the instrument are correlated, but distinct parts, taken together, constituted a multidimensional measure of STEM identity. In practical applications, students can be helped to self-identify themselves as STEM persons. Meanwhile, teachers can strengthen teaching practices through students' STEM identity information, enhancing students' STEM participation and retention to a certain extent, and reducing STEM pipeline leakage. For researchers, a more comprehensive and in-depth understanding of STEM identity can be obtained, which is conducive to further empirical research. Meanwhile, the instrument can be used to explore influential factors of STEM identity for adolescent students, such as gender, grade, family background, and school context [71].

On the other hand, the present study expanded the age range of the participants. Most researchers have looked at the two extremes of the student population for STEM identity: college students and elementary school students [15,72]. However, in the process of dynamic development and self-construction of STEM identity, adolescent students are more likely to be influenced by the environment and others, so more attention should be given to students at this age. In addition, the majority of students had some basic understanding of STEM as a mix of two or more of its component disciplinary fields. In China, students begin to learn STEM concepts from mathematics in the seventh grade, physics in the eighth grade, chemistry in the ninth grade, and general technology in high school. Students' STEM concepts begin to become richer, and the main concepts gradually shift to physics or science and technology. Due to the complexity of course content at this stage, the dynamic change in student identity is obvious. The present study takes junior and senior high school students as the research object, which has significance and value for research on STEM identity.

VI. CONCLUSIONS

The present study describes the process of developing and validating an instrument for assessing students' STEM identity. The content of the instrument was determined using expert reviews, students' interviews, and factor and Rasch analyses. The present study provides multiple sets of evidence for reliability and validity, supporting the appropriate structure of the instrument. The final instrument is a five-point Likert scale in which students can answer using the responses. The instrument includes four subscales, including recognition, performance and competence, interest, and belonging (Appendix). As the results indicated, the instrument can be used to evaluate the identity of young students and the sustainability of STEM fields. At the same time, the construction of the STEM Identity tool formed in the present study shows that the four-factor model is reasonable.

The limitations of the present study should be mentioned. First, the correlation of each dimension is highly significant, and their discriminant validity needs to be improved, such as by increasing the number of items or modifying the expression method of items with large residual correlation. Second, the study population did not include elementary school students by considering reading and comprehensive abilities. However, research suggests that children begin to identify their interests related to STEM as early as elementary school. Primary students begin to shape their personal identity and start making decisions about who they are and could be in the future [72–77]. Therefore, future studies should be applied to and modified for elementary school students to provide more evidence for psychometric properties. Finally, the sample we used for the pilot and field tests was from the southern and northern parts of China and did not include students from the many other regions of China. Since China is a multiethnic country and there were no ethnic minority students in the sample, we do not know whether ethnic minority identity affects STEM identity. In the future, it would make sense to measure STEM identity among students from more diverse ethnic groups.

ACKNOWLEDGMENTS

The research is supported in part by the Chinese Ministry of Education's Advanced Basic Science Program 2.0 under the Grant No. 20222152 and Guangdong Province Education Research Planning Project (Higher Education Special) of P.R. China under the Grant No. 2022GXJK175.

APPENDIX: STEM-ID SCALE ORIGINAL 30 ITEMS AND NEWLY ADDED ITEM

The appendix is about a developmental survey of STEM identity aimed at middle and high school education groups. It is to provide support for the research by collecting accurate and reliable data. The items in the scale have no right or wrong answers and do not reveal personal information, so students could answer the items according to their real thoughts.

Dimension	Item No.	Item
Recognition	R1*	I think I'm good at STEM-related subjects.
	R2*	My classmates think I'm good at STEM-related subjects.
	R3*	Teachers of STEM-related subjects think I am good at STEM-related subjects.
	R4*	My family and friends think I am a master of STEM-related subjects.
	R5	Before I started a new STEM-related course, I was confident in it
	R6*	I think I am gifted in STEM-related subjects.
	R7*	My classmates would ask me for STEM-related knowledge or exercises.
	P5*	I can achieve high levels of success in STEM-related subjects.
	P9^{-}	It takes me a long time to understand new knowledge in STEM-related subjects.
	B7^{-}*	I think I am completely unsuitable for STEM-related subjects.
	R8^{+}*	I am confident that I will be able to learn STEM-related subjects at the high school.
Performance and Competence	P1*	I can use tools and equipment proficiently in lab classes.
	P2*	I can understand the laws and principles of STEM-related subjects well.
	P3*	I can use science to explain natural phenomena in everyday life.
	P4	I am good at designing and fixing things.
	P6	I am able to complete homework assignments in STEM-related subjects well.
	P7	I believe I can learn a lot in STEM-related subject classes.
	P8	I believe I can solve complex STEM-related problems.
	P10	I can apply my knowledge of mathematics flexibly in science subjects.
	P11^{+}*	I believe that I can learn even the most difficult science subjects if I work hard.
Interest	I1*	I would like to learn more about STEM-related knowledge through various sources of information.
	I2*	I am interested in STEM-related careers.
	I3*	I like to participate in various STEM-related activities.
	I4*	I think the STEM knowledge I learn in the classroom is important in the daily life.
	I5	I like to actively think about STEM issues.
	I6	STEM issues excite me.
	I7^{+}*	I plan to pursue a STEM-related major in college.
	I8^{+}*	I intend to be more satisfied from STEM-related major courses than other courses.
Belonging	B1*	I would be proud to excel in STEM-related subjects.
	B2*	I would like to be seen as a person in the field of mathematics or science and technology.
	B3*	I am willing to show competence in math or engineering in front of others.
	B4*	I feel pleasure when talking to others about content related to math or science subjects.
	B5*	I intend to follow the example of STEM-related scientists and engineers.
	B6*	I would feel comfortable when talking to people who work in STEM-related fields.

($^{+}$) indicates that the item was added in the second round; (*) indicates that the item was retained for the final project instrument; ($^{-}$) indicates that the item was reverse-coded during analysis.

[1] National Research Council, *Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future* (The National Academies Press, Washington, DC, 2007).

[2] A. C. Barton, H. Kang, E. Tan, T. B. O'Neill, J. Bautista-Guerra, and C. Brecklin, Crafting a future in science: Tracing middle school girls' identity work over time and space, Am. Educ. Res. J. **50**, 37 (2013).

[3] J. P. Gee, Identity as an analytic lens for research in education, Rev. Res. Educ. **25**, 99 (2000).

[4] Z. Hazari, G. Potvin, R. H. Tai, and J. Almarode, For the love of learning science: Connecting learning orientation and career productivity in physics and chemistry, Phys. Rev. ST Phys. Educ. Res. **6**, 010107 (2010).

[5] J. Moote, L. Archer, J. DeWitt, and E. MacLeod, Science capital or STEM capital? Exploring relationships between science capital and technology, engineering, and maths aspirations and attitudes among young people aged 17/18, J. Res. Sci. Teach. **57**, 1228 (2020).

[6] J. Rahm and J. C. Moore, A case study of long-term engagement and identity-in-practice: Insights into the STEM pathways of four underrepresented youths, J. Res. Sci. Teach. **53**, 768 (2016).

[7] S. D. Simpkins, C. D. Price, and K. Garcia, Parental support and high school students' motivation in biology, chemistry, and physics: Understanding differences among Latino and Caucasian boys and girls, J. Res. Sci. Teach. **52**, 1386 (2015).

[8] Z. Hazari, R. Dou, G. Sonnert, and P. M. Sadler, Examining the relationship between informal science experiences and physics identity: Unrealized possibilities, Phys. Rev. Phys. Educ. Res. **18**, 010107 (2022).

[9] Z. Hazari, C. Cass, and C. Beattie, Obscuring power structures in the physics classroom: Linking teacher positioning, student engagement, and physics identity development, J. Res. Sci. Teach. **52**, 735 (2015).

[10] H. B. Carlone and A. Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, J. Res. Sci Teach. **44**, 1187 (2007).

[11] S. L. Li, Learning in a physics classroom community: Physics learning identity construct development, measurement and validation, Degree of Doctor of Philosophy Granting Oregon State University, Corvallis, Oregon, USA,2012.

[12] S. Chen and B. Wei, Development and validation of an instrument to measure high school students' science identity in science learning, Res. Sci. Educ. **52**, 111 (2022).

[13] M. E. Lockhart, O. M. Kwok, M. Yoon, and R. Wong, An important component to investigating STEM persistence: The development and validation of the science identity (SciID) scale, Int. J. STEM Educ. **9**, 34 (2022).

[14] K. M. Paul, A. V. Maltese, and D. S. Valdivia, Development and validation of the role identity surveys in engineering (RIS-E) and STEM (RIS-STEM) for elementary students, Int. J. STEM Educ. **7**, 45 (2020).

[15] Z. Hazari, D. Chari, G. Potvin, and E. Brewe, The context dependence of physics identity: Examining the role of performance/competence, recognition, interest, and sense of belonging for lower and upper female physics undergraduates, J. Res. Sci. Teach. **57**, 1583 (2020).

[16] R. Tytler, Attitudes, identity, and aspirations toward science, in *Handbook of Research on Science Education*, edited by Norman G. Lederman and Sandra K. Abell (Routledge, London, 2014), Vol. II, pp. 96–117.

[17] A. C. Barton, H. Kang, E. Tan, T. B. O'Neill, J. B. Guerra, and C. Brecklin, Crafting a future in science: Tracing middle school girls' identity work over time and space, Am. Educ. Res. J. **50**, 37 (2013).

[18] N. Enyedy, J. Goldberg, and K. M. Welsh, Complex dilemmas of identity and practice, Sci. Educ. **90**, 68 (2006).

[19] Z. Hazari, G. Sonnert, P. M. Sadler, and M. C. Shanahan, Connecting high school physics experiences,outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47**, 978 (2010).

[20] D. Verdín, The power of interest: Minoritized women's interest in engineering fosters persistence beliefs beyond belongingness and engineering identity, Int. J. STEM Educ. **8**, 33 (2021).

[21] F. A. Herrera, S. Hurtado, G. A. Garcia, and J. Gasiewski, *American Educational Research Association Annual Conference* (2012), https://heri.ucla.edu/nih/downloads/AERA2012HerreraGraduateSTEMIdentity.pdf.

[22] R. M. Hughes, B. Nzekwe, and K. J. Molyneaux, The single sex debate for girls in science: A comparison between two informal science programs on middle school students' STEM identity formation, Res. Sci. Educ. **43**, 1979 (2013).

[23] L. Rosenthal, B. London, S. R. Levy, and M. Lobel, The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university, Sex Roles **65**, 725 (2011).

[24] V. Seyranian, A. Madva, N. Duong, N. Abramzon, Y. Tibbetts, and J. M. Harackiewicz, The longitudinal effects of STEM identity and gender on flourishing and achievement in college physics, Int. J. STEM Educ. **5**, 40 (2018).

[25] J. Steinke, Adolescent girls' STEM identity formation and media images of STEM professionals: Considering the influence of contextual cues, Front. Psychol. **8**, 716 (2017).

[26] N. W. Brickhouse, P. Lowery, and K. Schultz, What kind of a girl does science? The construction of school science identities, J. Res. Sci. Teach. **37**, 441 (2000).

[27] D. M. Tice and H. Wallace, The reflected self: Creating yourself as (you think) others see you, in *Handbook of Self and Identity*, edited by M. R. Leary and J. P. Tangney (Guilford Press, New York, 2003), pp. 91–105.

[28] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity* (Cambridge University Press, Cambridge, England, 1998).

[29] J. W. Atkinson, Motivational determinants of risk-taking behavior, Psychol. Rev. **64**, 359 (1957).

[30] A. Wigfield and J. S. Eccles, Expectancy-value theory of achievement motivation, Contemp. Educ. Psychol. **25**, 68 (2000).

[31] M. M. Chemers, E. L. Zurbriggen, M. Syed, B. K. Goza, and S. Bearman, The role of efficacy and identity in science career commitment among underrepresented minority students, J. Soc. Issues **67**, 469 (2011).

[32] G. Trujillo and K. D. Tanner, Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity, CBE Life Sci. Educ. **13**, 6 (2014).

[33] J. Boaler, Exploring situated insights into research and learning, J. Res. Math. Educ. **31**, 113 (2000).

[34] P. Vincent-Ruz and C. D. Schunn, The nature of science identity and its role as the driver of student choices, Int. J. STEM Educ. **5**, 48 (2018).

[35] J. S. Eccles, Where are all the women? Gender differences in participation in physical science, and engineering, in

Why Aren't More Women in Science? Top Researchers Debate the Evidence, edited by S. J. Ceci and W. M. Williams (American Psychological Association, Washington, DC, 2007), pp. 199–210.

[36] J. S. Eccles, Who am I and what am I going to do with my life? Personal and collective identities as motivators of action, Educ. Psychol. **44**, 78 (2009).

[37] J. S. Eccles, Gendered educational and occupational choices: Applying Eccles et. al model of achievement-related choices, Int. J. Behav. Dev. **35**, 195 (2011).

[38] A. Bandura, Social cognitive theory: An agentic perspective, Annu. Rev. Psychol. **52**, 1 (2001).

[39] G. Potvin and Z. Hazari, The development and measurement of identity across the physical sciences, *presented at PER Conf. 2013, Portland, OR*, 10.1119/perc.2013.pr.058.

[40] J. Cribbs, Z. Hazari, G. Sonnert, and P. M. Sadler, Establishing an explanatory model for mathematics identity, Child Development **86**, 1048 (2015).

[41] A. Godwin, G. Potvin, Z. Hazari, and R. Lock, Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice, J. Eng. Educ. **105**, 312 (2016).

[42] J. Mahadeo, Z. Hazari, and G. Potvin, Developing a computing identity framework: Understanding computer science and information technology career choice, ACM Trans. Comput. Educ. **20**, 1 (2020).

[43] R. A. Beghetto, Factors associated with middle and secondary students' perceived science competence, J. Res. Sci. Teach. **44**, 800 (2007).

[44] K. A. Renninger, Interest and identity development in instruction: An inductive model, Educ. Psychol. **44**, 105 (2009).

[45] T. L. Strayhorn, *College Students' Sense of Belonging: A Key to Educational Success for all Students* (Routledge, London, 2018).

[46] P. T. Oon and X. Fan, Rasch analysis for psychometric improvement of science attitude rating scales, Int. J. Sci. Educ. **39**, 683 (2017).

[47] W. J. Boone, M. S. Yale, and J. R. Staver, *Rasch Analysis in the Human Sciences* (Springer, Berlin, Germany, 2014).

[48] M. H. M. Cheng and Z. H. Wan, Unpacking the paradox of Chinese science learners: Insights from research into Asian Chinese school students' attitudes towards learning science, science learning strategies, and scientific epistemological views, Stud. Sci. Educ. **52**, 29 (2016).

[49] R. Dou, Z. Hazari, K. Dabney, G. Sonnert, and P. Sadler, Early informal STEM experiences and STEM identity: The importance of talking science, Sci. Educ. **103**, 623 (2019).

[50] R. Dou and H. Cian, Constructing STEM identity: An expanded structural model for STEM identity research, J. Res. Sci. Teach. **59**, 458 (2022).

[51] D. R. Williams, H. Brule, S. S. Kelley, and E. A. Skinner, Science in the Learning Gardens (SciLG): A study of students' motivation, achievement, and science identity in low-income middle schools, Int. J. STEM Educ. **5**, 8 (2018).

[52] J. Pallant, *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*, 7th ed. (Routledge, London, 2002).

[53] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis: A Global Perspective*, 7th ed. (Pearson Prentice Hall, Upper Saddle River, New Jersey, 2010).

[54] G. Raîche, Critical eigenvalue sizes in standardized residual principal components analysis, Rasch Meas. Trans. **19**, 1012 (2005), https://www.rasch.org/rmt/rmt191h.htm.

[55] F. N. Alhija, Factor analysis: An overview and some contemporary advances, in *International Encyclopedia of Education*, 3rd ed., edited by P. L. Peterson, E. Baker, and B. McGaw (Elsevier, London, 2010), pp. 162–170.

[56] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. (Prentice-Hall, Upper Saddle River, New Jersey, 2010).

[57] A. Singer, G. Montgomery, and S. Schmoll, How to foster the formation of STEM identity: Studying diversity in an authentic learning environment, Int. J. STEM Educ. **7**, 57 (2020).

[58] M. K. Eagan, S. Hurtado, M. J. Chang, G. A. Garcia, F. A. Herrera, and J. C. Garibay, Making a difference in science education: The impact of undergraduate research programs, Am. Educ. Res. J. **50**, 683 (2013).

[59] L. Sha, C. Schunn, M. Bathgate,and A. Ben-Eliyahu, Families support their children's success in science learning by influencing interest and self-efficacy, J. Res. Sci. Teach. **53**, 450 (2016).

[60] S. D. Simpkins, C. D. Price, and K. Garcia, Parental support and high school students' motivation in biology, chemistry, and physics: Understanding differences among Latino and Caucasian boys and girls, J. Res. Sci. Teach. **52**, 1386 (2015).

[61] H. W. Marsh, K. T. Hau, and C. K. Kong, Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English compared with Chinese) for Hong Kong students, Am. Educ. Res. J. **39**, 727 (2002).

[62] N. A. Fouad, P. L. Smith, and K. E. Zao, Across academic domains: Extensions of the social–cognitive career model, J. Counsel. Psychol. **49**, 164 (2002).

[63] R. H. Tai, C. Q. Liu, A. V. Maltese, and X. Fan, Planning early for careers in science, Science **312**, 1143 (2006).

[64] A. Bandura, *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1986).

[65] R. M. Lock, Z. Hazari, and G. Potvin, Impact of out-of-class science and engineering activities on physics identity and career intentions, Phys. Rev. Phys. Educ. Res. **15**, 020137 (2019).

[66] H. B. Carlone, A. Johnson, and C. M. Scott, Agency amidst formidable structures: How girls perform gender in science class, J. Res. Sci. Teach. **52**, 474 (2015).

[67] H. B. Carlone, C. M. Scott, and C. Lowder, Becoming (less) scientific: A longitudinal study of students' identity work from elementary to middle school science, J. Res. Sci. Teach. **51**, 836 (2014).

[68] E. B. Moje and D. R. Dillon, Adolescent identities as demanded by science classroom discourse communities, *Reconceptualizing the Literacies in Adolescents' Lives* (Routledge, London, 2006), pp. 85–106.

[69] K. A. Renninger, Interest and identity development in instruction: An inductive model, Educ. Psychol. **44**, 105 (2009).
[70] A. J. Gonsalves, Physics and the girly girl-there is a contradiction somewhere: Doctoral students' positioning around discourses of gender and competence in physics, Cult. Stud. Sci. Educ. **9**, 503 (2014).
[71] P. T. Oon and R. Subramaniam, Factors influencing Singapore students' choice of physics as a tertiary field of study: A Rasch analysis, Int. J. Sci. Educ. **35**, 86 (2013).
[72] L. Archer, J. DeWitt, J. Osborne, J. Dillon, B. Willis, and B. Wong, "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity, Sci. Educ. **94**, 617 (2010).
[73] J. DeWitt and L. Archer, Who aspires to a science career? A comparison of survey responses from primary and secondary school students, Int. J. Sci. Educ. **37**, 2170 (2015).
[74] A. V. Maltese and C. S. Cooper, STEM pathways: Do men and women differ in why they enter and exit?, AERA Open **3**, 1 (2017).
[75] A. V. Maltese, C. S. Melki, and H. Wiebke, The nature of experiences responsible for the generation and maintenance of interest in STEM, Sci. Educ. **98**, 937 (2014).
[76] A. V. Maltese and R. H. Tai, Eyeballs in the fridge: Sources of early interest in science, Int. J. Sci. Educ. **32**, 669 (2010).
[77] C. Murphy and J. Beggs, Children's perceptions of school science, Sch. Sci. Rev. **84**, 109 (2003), https://www.semanticscholar.org/paper/Children%27s-perceptions-of-school-science-Murphy-Beggs/cc0c672e362bb930cc6bc-4bbbb36c5baacd3d620.