International Journal of
STEM Education

RESEARCH   **Open Access**

Development and validation of the role identity surveys in engineering (RIS-E) and STEM (RIS-STEM) for elementary students

Kelli M. Paul[*] , Adam V. Maltese and Dubravka Svetina Valdivia

Abstract

Background: Despite the increasing number of science, technology, engineering, and mathematics (STEM) jobs available, concern continues to grow over the low number of students who choose to study and enter STEM fields. Research suggests that children begin to identify their interests and career aspirations related to STEM as early as elementary school when they begin to shape their personal identities and start making decisions about who they *are* and *could be* in the future, their role identities (e.g., scientist, engineer). Existing surveys that assess identity target high school or post-secondary students, with less work on elementary and middle school students. This paper describes the development and validation of survey instruments to assess engineering identity in elementary students and its adaptation to a more general STEM context.

Results: The role identity survey in engineering (RIS-E) was developed across four phases of pilot testing where it was administered to 634 students in third–sixth grade enrolled in classrooms in the West, Midwest, and Northeastern United States. Exploratory modeling approaches and scale reliability were used to narrow down items, while confirmatory factor analyses (CFA) and item response theory (IRT) approaches were used to examine item performance. The final survey contained four scales that assess aspects of one's identity (competence, interest, self-recognition, and recognition by others), all of which demonstrated strong psychometric properties. The RIS-E was then adapted to assess STEM identity (RIS-STEM), and it was administered to 678 fourth–fifth grade students enrolled in classrooms in the Southwestern United States. CFA and IRT analyses provided support for use of the RIS-STEM in a more general STEM context.

Conclusions: The RIS-E and RIS-STEM appear to produce reliable scores that measure aspects of identity (engineering and STEM) in elementary students. Suggestions are made for future studies to examine how the RIS-E and RIS-STEM function across diverse student populations and the impact on one's identity as a result of curricula or programs designed to encourage and support identity development in youth, especially in engineering and STEM.

Keywords: Identity, Engineering, STEM, Survey development, Validity, Elementary students

Introduction

Context

Despite the increasing number of science, technology, engineering, and mathematics (STEM) jobs available (U.S. Bureau of Labor Statistics, n.d.; Lockard & Wolf,

2012), concern continues to grow whether there will be graduates to fill these positions (President's Council of Advisors on Science and Technology, 2012). Student interest and attitudes toward STEM play an influential role in students choosing to continue to study or pursue a STEM career (Doerschuk et al., 2016; Maltese & Tai, 2011). However, some reports suggest that interest in STEM fields and careers among students is only modest

* Correspondence: kelpaul@iu.edu
School of Education, Indiana University, Bloomington, IN 47405, USA

Paul *et al. International Journal of STEM Education*   (2020) 7:45

Page 2 of 17

and has changed very little over the past few years (American College Testing, 2017; Guzey, Harwell, & Moore, 2014). As a result, many outreach programs target high school students with the hope of increasing their interest in STEM as students begin to make decisions related to attending college or pursuing a career. However, research suggests that children begin to identify their interests and career aspirations related to STEM as early as elementary school (Archer et al., 2010; DeWitt & Archer, 2015; Maltese & Cooper, 2017; Maltese, Melki, & Wiebke, 2014; Maltese & Tai, 2010; Murphy & Beggs, 2003) when they begin to shape their personal identities and start making decisions about who they *are* and *could be* in the future. In other words, students begin to develop their role identities (e.g., scientist, engineer). Unlike constructs such as attitudes or self-efficacy, one's role or science identity has received less attention (Kim & Sinatra, 2018), despite its potential for predicting pursuing science in one's future (Godwin, Potvin, Hazari, & Lock, 2016; Hazari, Sonnert, Sadler, & Shanahan, 2010; Stets et al., 2017; Vincent-Ruz & Schunn, 2018; Wagstaff, 2014). Identity theory posits that we have multiple identities (personal, social, role) that shape our choices and behaviors (Carlone & Johnson, 2007; Pierrakos, Beam, Constantz, Johri, & Anderson, 2009). When we see ourselves as someone who can "do" or "be" a certain type of person, we are more likely to act in a manner consistent with these beliefs. So, the greater degree to which an adolescent sees herself as a "science person," the greater the likelihood that she will continue to engage in and possibly pursue science in her future. While girls and boys have similar interests in STEM, during the transition to upper elementary and middle school, students face social norms that influence their continued interest in and identification with STEM (Archer et al., 2010; Murphy & Beggs, 2003). Social pressure often forces girls to enact the identity of "female" over that of "scientist" or "engineer" (Carlone, Johnson, & Scott, 2015; Carlone, Scott, & Lowder, 2014) and contributes to a decline in girls' identification with STEM.

Theoretical framework

Science identity has been examined from different disciplinary perspectives, resulting in multiple conceptualizations and definitions of identity. Disagreement exists in the literature as to whether attitudinal constructs (e.g., interest, competence) influence identity development or whether these constructs make up one's identity. Vincent-Ruz and Schunn (2018) conceptualized an individual's science identity as composed of both one's internal view of self and the perceptions of external others. In other words, one's science identity is a singular construct that consists of self-recognition and

recognition by others as a science person. They found that science identity was distinct and separate from other attitudinal constructs such as competence, values, and fascination.

Another oft-cited conceptualization is that of Eccles and colleagues who identified identity formation as part of their expectancy value model of achievement-related choices. Various social and psychological factors (such as cultural norms, personal beliefs, experiences, and aptitudes) influence an individual's expectations for success and the value that they place on available task options, all of which impact the decisions and choices an individual makes, such as enrollment in science courses or pursuing an engineering career. Within the model, expectations of success and values influence the formation of one's collective identities, or aspects of the one's self related to social groups and relationships, and one's personal identities, or aspects that defines how individuals are unique from others (Eccles, 2007, 2009, 2011; Wigfield & Eccles, 2000).

Unlike Eccles' expectancy value model in which attitudinal constructs influence identity formation, Carlone and Johnson (2007) conceptualize identity as consisting of three interrelated components that work together to form one's science (or role) identity: *performance*, *competence*, and *recognition*. Specifically, one cannot be a particular kind of person (e.g., a science person) unless she is able to demonstrate the behaviors (perform) and be knowledgeable in the practices of a field (competence), all of which must be found to be credible by others (recognition). Although they may be exhibited to varying degrees, all three components are necessary aspects of one's role identity. While they initially posited recognition to be a singular component, Carlone and Johnson (2007) emphasized dual types of recognition. They acknowledged that recognition by others is important in developing a science identity, but just as important is whether one views one's self as a science person (self-recognition).

Drawing from a social-cognitive perspective and based on the relationship between interest and persistence in science, Hazari and colleagues built upon Carlone and Johnson's framework and included interest as a fourth component of science identity (Hazari et al., 2010; Lent, Brown, & Hackett, 1994, 1996). Interest reflects one's desire or curiosity to think about or understand a subject. While Carlone and Johnson acknowledged the importance of interest in the development of one's identity, they did not include it in their model due to the assumption that scientists in their study were interested in the content in which they practiced (Hazari et al., 2010). However, interest could not be assumed in the study by Hazari and colleagues who examined the development of physics identity in high school students, an age at

which interest in science (or lack thereof) is a strong predictor of persistence and future success in science. We draw on the work of Carlone and Johnson and Hazari and colleagues, and we took on their identity frameworks to guide our conceptions of science, engineering, and STEM identity.

Much of the literature on role identity examines either science, writ large or STEM identity, without separating out individual sub-disciplines, such as engineering (Pierrakos et al., 2009). Efforts to attract and retain more girls and minorities in specific fields such as engineering rely on the extent to which these individuals see themselves as someone who does or can do engineering (Brickhouse, Lowery, & Schultz, 2000). Therefore, recent attention has focused specifically on engineering identity, though much of the research in this emergent field focuses on high school or post-secondary students (Capobianco, French, & Diefes-Dux, 2012; Patrick & Borrego, 2016), with less research on elementary students.

In addition to increasing focus on engineering, more educators and outreach providers are trying to engage students at the elementary level in the broader notion of STEM. Such a focus warrants examination of students' STEM identity development over time. Namely, do elementary students develop a "STEM" identity whereby they come to see themselves as a "STEM person" or someone who does STEM? Research on STEM identity has increased over the past decade (Simpson & Bouhafa, 2020) and typically focuses on STEM identity interchangeably with science identity or examines STEM attitudes or interests, focusing on some or all of the individual components of STEM separately (Kier, Blanchard, Osborne, & Albert, 2014; Simpson & Bouhafa, 2020; Staus, Lesseig, Lamb, Falk, & Dierking, 2019; Tyler-Wood, Knezek, & Christensen, 2010). Some work examines attitudes toward STEM and STEM integration (Guzey et al., 2014), though few studies examine STEM identity as a singular construct (Appianing & Van Eck, 2018), with little work done at the elementary level (Simpson & Bouhafa, 2020).

Current study

The purpose of the current study is twofold. First, we sought to develop an instrument to assess engineering identity in elementary students in order to study the efficacy of interventions targeted at improving identification with engineering in children. Second, we explored the usefulness of the instrument in a more general context. As more educators begin to focus on students' interest in a general notion of STEM, we sought to examine how well the survey would function if adapted from assessing "engineering" to "STEM." We summarize the survey development process undertaken to create the engineering-specific version of the survey to allow the reader to understand how and why we included various survey items. We then describe the minor

adaptations to the survey to fit items to a more general STEM context. Finally, we present and compare evidence on the psychometric properties of both survey versions.

We drew from identity theory to inform the development of an engineering identity survey, with additional work conducted to adapt this instrument to a more general STEM context. Specifically, we focused on dimensions of one's role- or career-related identity, rather than personal (e.g., individual characteristics) and social (e.g., group membership characteristics) identities (Burke & Stets, 2009; Carlone & Johnson, 2007; Hazari et al., 2010; Patrick & Borrego, 2016; Stets et al., 2017). Our primary interest was in assessing aspects of role-related identity as this often encompasses roles undertaken as part of a career, and thus we do not include these other important aspects of identity. Rather than focusing on only one aspect of role identity (e.g., interest), we focused on the larger identity framework (i.e., Carlone & Johnson; Hazari et al., 2010) as doing so provides a more complete lens through which to examine students' identification with a role or career area.

The addition of engineering and STEM into school and outreach programs has led to increased focus on curriculum development with lesser focus on measures to assess impacts of such efforts on one's identity. Studies that examine the development of one's science and engineering identity often utilize interviews or case-study methodology (Archer, Moote, Francis, DeWitt, & Yeomans, 2017; Aschbacher, Li, & Roth, 2010; Patrick & Borrego, 2016) which are less pragmatic for classroom or outreach settings with large numbers of students (Trujillo & Tanner, 2014). Many of the existing surveys that assess identity tend to target high school or post-secondary students (Godwin, Potvin, Hazari, & Lock, 2013; Hazari et al., 2013), with much less work on elementary and middle school students (DeWitt & Archer, 2015).

The role identity survey–engineering (RIS-E) was developed within the context of a multi-year research project examining the impact of an engineering outreach program on elementary students. We needed an age-appropriate instrument capable of examining changes in elementary students' engineering identities as a result of having participated in the outreach program. After developing the survey and through discussion with a range of practitioners and researchers, we believed that the engineering identity constructs could also apply to the broader context of STEM. Below, we describe the methods and phases of development of the RIS-E (study 1) and its adaption to a more general STEM context (RIS-STEM; study 2). For each study, we present the methods, data analysis, results, and validity evidence for each phase of development for both versions of the survey. All research was conducted with permission from

our institution's Internal Review Board with informed consent obtained from all participants whose data were used in the study.

Study 1

Study 1 describes the multi-phase process through which we developed and pilot tested the role identity survey–engineering (RIS-E). Initially, we began by using the Engineering Identity Development Scale (EIDS; Capobianco & Deemer, 2017). Our initial results suggested that the EIDS and a modified version where we expanded the response categories might not be appropriate for our purposes (see "Phase 1" section). As such, we developed the RIS-E to better meet the needs of our project.

Phase 1
Methods

Participants Participants in phase 1 were 80 elementary students enrolled in 3rd–5th grade classrooms in suburban schools participating in an engineering outreach program in the Northeastern United States (36 girls, 44 boys). Students self-reported their race/ethnicity with most students identifying themselves as White/Caucasian (45%), followed by Black/African American (15%), multi-racial (9%), Asian (8%), Hispanic/Latino(a) (6%), and other (e.g., Native American, Pacific Islander; 1%). Thirteen students (16%) did not report their race/ethnicity.

Data sources and collection procedures In phase 1, we utilized the 16-item revised Engineering Identity Development Scale (EIDS) which measured three components of identity: *academic identity*, *occupational identity*, and *engineering aspirations* (Capobianco & Deemer, 2017). Students responded to each item using a 3-point scale (1 = "no," 2 = "not sure," and 3 = "yes"). We utilized the EIDS as a pre-survey with students participating in the engineering outreach program because it had been specifically designed to assess engineering identity in elementary age students and had established reliability and validity evidence (Capobianco et al., 2012). A paper-pencil survey was administered to students by classroom teachers in the fall of 2017.

On this and all subsequent versions of the survey, we did not define "engineering" for students. We recognize that students vary in their interpretation and understanding of engineering and what it means to be an engineer or do engineering. We also recognize that students' conceptualization of engineering might skew towards that of science. However, we were interested in students' perceptions based on their understanding of engineering.

Analyses

Review of the data collected from our administration of the EIDS suggested limitations of the instrument. We examined frequency distributions and found that students in our sample did not utilize the full three-point response scale on most items. We were also concerned that the "unsure" option was utilized more for an "I don't know" rather than a middle point between "no" and "yes". Second, we calculated Cronbach's alpha coefficients to examine the internal consistency of the items for each component of identity, and we found that scale reliabilities were lower than desired on the *academic identity* component ($\alpha < .70$). Finally, we conducted exploratory factor analysis (EFA) and found that factor loadings were inconsistent with multiple items cross-loading across factors on the *academic identity* and *occupational identity* components.

Results and discussion

Our initial literature review focused on measures of identity for elementary age students. We chose to use the EIDS because it was specifically designed to assess engineering identity in elementary students and had established reliability and validity evidence (Capobianco et al., 2012). However, throughout data collection during phase 1, we continued to examine the literature on identity in order to broaden out item pool. This expanded review included work on identity at all age levels and across disciplines (e.g., science, physics) exploring the possibility of adapting items to an engineering context if necessary. Based on the findings from our analyses, we felt that the EIDS was limited in assessing elementary students' engineering identity for two primary reasons. First, our results suggested inconsistent factor loadings and lower reliability compared to previous research using the EIDS. Second, closer examinations of the items suggested that the EIDS measured academic identity and knowledge of engineering careers more so than engineering identity. Our expanded literature review identified aspects of one's role-related identity not assessed by the EIDS that we felt were important for describing students' engineering identity. Specifically, the EIDS did not include aspects of one's identity such as interest or competence in doing engineering. Therefore, we sought to generate new items for use in phase 2.

Phase 2
Methods

Participants Participants in phase 2 were 89 elementary students enrolled in 3rd–5th grade classrooms in suburban schools participating in an engineering outreach program in the Northeastern United States. Demographic information was not collected from students as

the survey was used only for pilot testing items. A paper-pencil survey was administered to students by classroom teachers in late fall of 2017.

Data sources and collection procedures Based on the findings from phase 1, we revised the survey instrument in three ways. First, the response options were changed from a 3-point to a 4-point scale (1 = NO!, 2 = no, 3 = yes, 4 = YES!). This scale represents a measure of intensity of agreement and addressed the issue of interpreting what the "unsure" response means on the previous 3-point scale. We changed to the "NO/no" and "YES/yes" anchor points as these have been found to be easier to interpret by younger children than traditional scales that use strongly agree to strongly disagree anchors (Moore, Bathgate, Chung, & Cannady, 2011; Sha et al., 2015). Additionally, the meaning of the anchor points was explained to students, along with an example, in the item instructions. We experimented with graphics in previous research and found that youth were not always sure how to interpret them, particularly when it came to negatively worded items. Second, the *academic identity* items of the EIDS were eliminated. These items assess one's general identity as a student or learner but do not reflect one's specific engineering identity, which was our goal in developing the survey. Additionally, this scale had lower than desired reliability ($\alpha < .70$); thus, we eliminated these items. We kept items from the EIDS that assessed *engineering aspirations* (four items) and *occupational identity* (six items) as the content of these items related to one's engineering identity and had acceptable internal consistency values ($\alpha > .70$). Finally, 12 new items were added based on review of the literature. We primarily adapted these items from Godwin's measure of engineering identity which was designed to assess engineering identity in post-secondary students (Godwin, 2016), with an additional item drawn from the Engineering Interest and Attitudes Survey (EIA; Higgins, Hertel, Shams, Lachapelle, & Cunningham, 2015). The new items were specifically chosen to assess *recognition* (four items), *interest* (five items), and *performance/competence* (three items). The phase 2 version of the survey consisted of 22 items.

Analyses

We followed the same analytic procedures described in phase 1 to examine the data with one exception. Namely, we did not conduct exploratory factor analyses. Examination of descriptive statistics (means, standard deviations, frequency distributions) found greater variance using the 4-point scale such that all four response options were used on almost all items. Scale reliabilities (Cronbach's α) were within acceptable ranges for four of the five scales (ranged from $\alpha = .74–.83$); however,

occupational identity had lower than acceptable reliability ($\alpha = .51$).

Results and discussion

With the exception of the *occupational identity* items, we felt that the remaining items piloted during phase 2 showed promise for assessing students' engineering identity. However, we believed it was important to over-sample items on each construct in order to narrow down each scale to a minimum of the four or five best performing items. We continued to examine the literature to generate additional items to pilot in phase 3.

Phase 3
Methods

Participants Phase 3 was conducted with 323 third–fifth grade students enrolled in urban, suburban, and rural classrooms in the West, Midwest, and Northeastern United States (159 girls, 126 boys, 38 did not provide gender information). Students self-reported their race/ethnicity with most students identifying themselves as White/Caucasian (36%), followed by multi-racial (10%), Asian (5%), Black/African American (2%), Hispanic/Latino(a) (< 1%), and other (including Native American, Pacific Islander; 4%). Just under half of the students did not report their race/ethnicity (43%) with some students indicating that they preferred not to answer the item (28%) and other students skipping the item entirely (15%).

Data sources and collection procedures During phase 3, we retained all of the newly added items related to *recognition*, *interest*, *performance*, and *competence* and all of the *engineering aspirations* items from the EIDS. We eliminated the *occupational identity* items due to their low reliability. We created a larger pool of items to assess each construct, continuing to draw and adapt items from a variety of existing measures including the STEM Fascination and Competence/Self-efficacy Scales (Chen, Cannady, Schunn, & Dorph, 2017; Chung, Cannady, Schunn, Dorph, & Vincent-Ruz, 2016), the STEM Career Interest Survey (STEM-CIS; Kier et al., 2014), the Modified Attitudes toward Science Inventory (M-ATSI; Weinburgh & Steele, 2000), and the Persistence Research in Science & Engineering survey (PRiSE; https://www.cfa.harvard.edu/sed/projects/PRiSE_survey_proof.pdf). In total, we added 21 items in the following areas: *performance and competence* (eight items), *STEM fascination* (six items), *interest* (four items), *outcome expectations* (two items), and *recognition* (one item). The phase 3 survey consisted of 37 items.

Prior to administering the survey to students, we submitted the items to three research faculty who are expert

in engineering education and identity development for their feedback on the content validity and construct coverage of the items. We also submitted the items to two elementary teachers to review the items for content appropriateness and readability, especially for the targeted age groups. Because we drew from validated instruments, we sought to keep item wording as close as possible to the original wording. However, since the existing instruments often were developed and validated for secondary students or older, we wanted to ensure that elementary students would understand them. Minor wording changes were made only for a few items to simplify and make them age-appropriate for elementary students. Despite these efforts, the nature of the content of the questions resulted in the readability continuing to skew toward older students. The revised version of the survey in phase 3 was administered to students using Qualtrics in the spring of 2018.

Analyses

We used exploratory modeling approaches as well as reliability analyses to identify underlying factors and to further reduce the number of items. Principal components analysis (PCA) with oblique rotation was selected as we expected the factors to be related to one another (Tabachnick & Fiddell, 2001). Pairwise deletion was used to maximize the sample size available for each comparison. Utilizing iterative rounds of analyses, we eliminated items that did not load on any factor within each round. This process continued until all items loaded on at least one factor with a loading of |.40| or higher.

Results and discussion

After 11 iterations, the final model retained consisted of 24 items across five factors: *competence* (seven items), *interest* (six items), *self-recognition* (five items), *recognition by others* (three items), and *negative perceptions of engineering* (three items). The five factors accounted for 63% of the variances. Factor correlations were weak-to-moderate and ranged between |.216| and |.459|, with the largest correlations found between *competence* and *interest* and between *competence* and *self-recognition* (see Table 1).

Four of the five factors were consistent with what we would expect from the existing literature on identity.

Table 1 Correlations among factors

Factors	1	2	3	4	5
1. Competence	-				
2. Interest	− .459	-			
3. Self-recognition	.424	− .381	-		
4. Recognition by others	− .322	.220	− .322	-	
5. Negative perceptions of engineering	− .358	.327	− .291	.216	-

The *interest* factor reflected an individual's enjoyment in doing engineering activities, while the *competence* factor reflected students' beliefs in their ability when doing engineering activities. Both factors are similar to constructs found in the literature (Carlone & Johnson, 2007; Godwin, 2016; Hazari et al., 2010). The two recognition factors (*recognition by others* and *self-recognition*) were consistent with work by Carlone and Johnson (2007) which suggests that while recognition by others is important in developing a science identity, just as important is whether one views one's self as a science person (self-recognition). The fifth factor, *negative perceptions of engineering*, was unexpected. These items were intended to serve as negatively worded items to assess the constructs of *interest* and *competence*. While these items could reflect a negative perception or view related to the difficulty of engineering or due to students' lack of a mental model for engineering, we believe it is more likely that these items loaded on one factor simply due to their negative wording. In other words, if the items had been worded positively (e.g., "I understand engineering"), they might have loaded on the originally intended factor. Negatively worded items have been found to be psychometrically problematic because of the added level of difficulty when answering them. Some researchers have found that negatively worded items create distinct, albeit artificial, factors (Schmitt & Stults, 1985; Swain, Weathers, & Niedrich, 2008; Woods, 2006).

Phase 4
Methods

Participants Phase 4 was conducted with 242 third–sixth grade students enrolled in urban, suburban, and rural classrooms in the Northeastern United States (73 girls, 81 boys, 88 did not provide gender information). Just under half of the students (39%) were a part of the target outreach program while the remaining students were not (61%). Students self-reported their race/ethnicity with most students identifying themselves as White/Caucasian (19%), Asian (5%), Hispanic/Latino(a) (5%), other (including Native American, Pacific Islander; 4%), multi-racial (3%), and Black/African American (2%). An additional 13% indicated an ethnicity other than one listed. Just under half of the students did not report their race/ethnicity (49%) with some students indicating that they preferred not to answer the item (18%) and other students skipping the item entirely (31%).

Data sources and collection procedures Based on phase 3 analyses, new items were added to three of the scales to further test out possible items that represented multiple aspects of the construct. In reviewing the items that assessed *recognition by others*, we felt it was

necessary to include other individuals with whom students this age interact who could potentially recognize an adolescent individual as an engineer (e.g., kids in my class, parents). Similarly, we added items to *self-recognition* that reflected more ways an individual could self-recognize as an engineer (e.g., "I feel like an engineer when I apply engineering ideas to my life"). Items added to *interest* reflected a wider variety of ways that students could demonstrate their interest in engineering (e.g., talking to others about engineering, searching for information about how things work). No additional items were added to *competence* as we believed that this construct was adequately assessed. Finally, based on the literature, we continued to conceptualize the negatively worded items as reflecting the constructs of *interest* and *competence* in this final round of testing. We re-worded one of the negatively worded items to make it less difficult to interpret. In total, we kept or re-worded all 24 of the previous items and added 18 new items, such that the phase 4 survey now consisted of 42 items across the four scales. In phase 4, the survey was administered using Qualtrics to students in the target outreach program in the fall of 2018 and to non-program students early in the spring of 2019.

Analyses

The overarching goal of study 1 was to develop an instrument to measure elementary students' engineering identity. During phase 4, several approaches were utilized in order to narrow down the items on individual scales and to arrive at a psychometrically sound instrument that measures the four domains. Specifically, we examined items via reliability analysis using Cronbach's alpha and item-total relationship via point biserial (*pbis*) via `reliability` function in the classical test theory (CTT) package (Willse, 2018) in R (R Core Team, 2019). Criteria to remove items was based on alpha increasing if an item in question was removed and if *pbis* was negative or low (below .30).

Additionally, confirmatory factor analysis (CFA) and item response theory (IRT) approaches were used to further examine items' performance. Since correlations were somewhat weak, in order to conduct IRT analyses, we decided to analyze each construct separately. Specifically, single-factor CFA models were fit using lavaan package (Rosseel, 2012) in R (R Core Team, 2019) using the `cfa` function. In addition to chi-square and its significance (which are known to be sensitive to large sample size), we reported a variety of traditional fit indices. There are several different recommendations in the literature of what is considered acceptable fit, and Hu and Bentler's (1999) is often used to evaluate model fit. For general guidance following Hu and Bentler, we deemed model fit as acceptable if the root mean square error of

approximation (RMSEA) was close to .06, if the comparative fit index (CFI) and Tucker Lewis index (TLI) were close to .95 or above, and if the standardized root mean square residual (SRMSR) was .08. These model fit indices were used together to evaluate CFA model fit to investigate whether a single factor model was attainable, but we report the obtained statistics completely to allow readers to draw conclusions about the model fit.

If a reasonable single factor model was obtained, we fit a graded response IRT model to the item responses using the mirt package (Chalmers, 2012) in R (R Core Team, 2019). A graded response model (GRM; Samejima, 1969) was fit using a `mirt` function, with default options. The GRM model deals with ordered polytomous categories and thus is an appropriate IRT model to be used in the analysis. In order to examine the overall model fit, we used C2, a variant for M2 statistic (Maydeu-Olivares & Joe, 2006) that is appropriate for polytomous response models that do not have sufficient degrees of freedom to compute M2*. We also used the `itemfit` function to obtain signed chi-square test (S_X2; Orlando & Thissen, 2000, 2003; Kang & Chen's, 2007) and signed chi-squared test to evaluate individual item fit. We adjusted *p* values for multiple comparisons for the items using the `p.adjust` function. Lastly, we examined overall information functions for each of the scales at the test level to further examine final items' ability to provide reliable parameter estimates (i.e., information in IRT is analogous to reliability).

Results and discussion

We next describe the iterative process that ultimately led to the final set of items for each of the four identity scales: *competence*, *interest*, *self-recognition*, and *recognition by others*. Table 2 reports reliability evidence for the four scales. Results of CFA/IRT analyses of overall model fit and item level fit and parameter estimates are presented in Tables 3 and 4, respectively.

Competence A total of nine items were retained from the earlier phases of analyses. Based on reliability and item analysis, two items were removed (i.e., if deleted,

Table 2 Reliability and item analysis for identity scales—engineering (phase 4)

Scales	n	No. items	*a*	*a* if deleted	*pbis* range
Competence	192	7	.818	None	.48–.62
Interest	180	7	.883	None	.57–.79
Self-recognition	179	6	.921	None	.72–.84
Recognition by others	179	6	.901	None	.61–.81

pbis range represents the point biserial correlation between the item in question and the total scale score. The ranges presented here are across all items within a scale
a Cronbach's alpha estimate

alpha would increase and *pbis* values for those two items were near zero or negative). Once the two items were removed, the remaining seven items yielded an alpha of .818 (an increase from the alpha of .673 for the nine items), with reasonably strong *pbis* values ranging from .48 to .62 across items. Furthermore, alpha would not increase by deletion of any other items. Based on the CFA and IRT analyses of overall model fit for the seven items, the competence scale yielded somewhat adequate fit. While the chi-square/C2 statistics were statistically significant, the RMSEA values were .123 and .095, respectively. Further, CFI/TLI values ranged from .923 to .962, suggesting only slightly reasonable model fit. Lastly, the SRMSR value of .068 supported earlier findings largely that a single factor model for *competence* is reasonable (although a better overall model fit may be desirable).

Focusing on the item-level IRT analyses, it was observed that the final seven *competence* items yielded acceptable item fit. Furthermore, the item discriminations were all high (1.3 or above), and item step difficulties covered a wide range of the construct continuum (in particular for those two standard deviations below the mean and about one standard deviation above the mean). As noted in Fig. 1, the overall test information (i.e., scale reliability) suggests that these seven items yield the most reliable estimates for people at around average levels of *competence*.

Interest Based on the previous phases, a total of 15 items with alpha of .886 remained on the interest scale prior to final analyses. In order to reduce the number of items, two separate iterations were conducted based on classical item analysis (i.e., looking at change in alpha and *pbis*). In the first iteration, five items were identified for removal. The remaining 10 items yielded an alpha of .908, with most items having high *pbis* (upper .40s or higher). The second iteration of removal was conducted to obtain a slightly shorter scale. Three items were removed due to negative or low *pbis*. The remaining seven items had a reasonable alpha of .883, with *pbis* ranging from .57 to .79 across the items.

The overall model fit of the *interest* scale was adequate as suggested by CFA and IRT model fit indices. Specifically, in both cases, the chi-square (C2) statistics were nonsignificant, the RMSEA values were small (.064 and .017, respectively), while CFI/TLI indices were over .99 for CFA and IRT analyses, and SRMSR was .042. Additionally, all items on the interest scale yielded good item level fit, and item parameter estimates suggested items with strong discriminations and reasonable spread of the item step difficulties. Similar to the competence scale, the interest scale items tended to cover the lower end of the construct continuum (see Fig. 1) and up to just above the average. Standard errors were similarly very low across the continuum, further suggesting that these items indeed measure the construct across a wide range of the continuum.

Self-recognition The original self-recognition scale contained nine items, which yielded an alpha of .924, with strong *pbis* values of .628 or higher for any one item. Removing any of the items would not yield a higher alpha, and due to the high *pbis*, there were no clear candidates to eliminate in order to shorten the scale. Thus, after fitting the CFA model, the three items with the lowest loadings were eliminated (we examined the content of the removed items to ensure that it was not crucial to keep in the scale). Thus, the scale was reduced to six final items. Although removing strong items reduced the alpha, it remained very high at .921, with very strong *pbis* values in the range of .72 to .84 across the items. Based on the overall fit at the scale level, all fit indices across the CFA and IRT analyses suggested a strong model fit (i.e., non-significant chi-square and C2; RMSEA values of .042 or lower, CFI/TLI values greater than .999, and SRMSR of .030, the lowest of all four scales). Examining individual items, all six retained items had good item fit, displayed a range of step difficulties, and had high discrimination values (ranging from 2.387 to 5.216). Based on these analyses, we concluded that the six items retained on the self-recognition scale do a good job covering the trait continuum (see Fig. 1).

Table 3 Overall model fit for identity scales—engineering

| Scales | CFA | | | | | IRT | | | | | |
	χ	df	RMSEA	CFI	TLI	C2 [a]	df	RMSEA	CFI	TLI	SRMSR
COMP	54.26*	14	.123	.948	.923	37.97*	14	.095	.962	.943	.068
INT	24.40	14	.064	.994	.991	14.73	14	.017	.999	.999	.042
SELF	11.89	9	.042	.999	.999	9.76	9	.022	.999	.999	.030
OTHERS	42.55*	9	.145	.989	.981	31.83*	9	.119	.977	.961	.047

COMP competence, *INT* interest, *SELF* self-recognition, *OTHERS* recognition by others
*$p < .01$
[a]C2 is a variant of the M2 statistic (Maydeu-Olivares & Joe, 2006) for evaluating model fit; C2 is appropriate for polytomous models when there is not sufficient degrees for freedom (e.g., due to low small number of items)

Table 4 Item fit and item parameter estimates for identity scales—engineering

Scales	Item ID	Item fit			Item parameter estimates[a]			
		S_X2*	df	RMSEA	a	b1	b2	b3
Competence	Comp1	19.62	15	.040	1.700	− 2.012	− 0.779	1.076
	Comp2	10.75	10	.020	1.818	− 2.710	− 1.744	0.217
	Comp3	13.75	14	.000	1.358	− 2.282	− 0.435	1.979
	Comp4	16.60	12	.045	1.647	− 2.812	− 1.454	0.497
	Comp5	10.45	11	.000	1.884	− 2.401	− 1.428	0.641
	Comp6	7.97	11	.000	1.943	− 2.904	− 1.555	0.484
	Comp7	12.93	11	.030	1.827	− 2.976	− 1.695	0.451
Interest	Int1	9.96	12	.000	2.015	− 2.499	− 1.362	− 0.117
	Int2	15.54	11	.048	1.943	− 2.775	− 1.669	− 0.122
	Int3	17.01	11	.055	2.010	− 2.348	− 1.663	0.117
	Int4	9.51	10	.000	1.723	− 1.813	− 0.444	0.729
	Int5	11.19	9	.037	3.968	− 1.994	− 1.128	0.187
	Int6	14.88	6	.091	3.970	− 2.263	− 1.468	− 0.244
	Int7	4.54	8	.000	2.212	− 1.834	− 1.148	0.425
Self-recognition	Self1	18.19	14	.041	2.703	− 1.035	− 0.165	1.021
	Self2	16.72	13	.040	2.623	− 1.866	− 0.486	0.57
	Self3	14.73	12	.036	3.700	− 0.885	0.265	1.069
	Self4	4.57	8	.000	5.216	− 0.821	0.188	1.002
	Self5	10.30	13	.000	3.829	− 0.68	0.164	0.928
	Self6	12.74	17	.000	2.387	− 1.358	− 0.482	0.669
Recognition by others	Oth1	18.00	19	.000	1.789	− 1.474	− 0.268	1.107
	Oth2	10.57	12	.000	4.183	− 0.613	0.321	1.140
	Oth3	12.40	13	.000	3.519	− 0.81	0.018	0.857
	Oth4	34.46	17	.076	1.884	− 1.391	− 0.146	1.076
	Oth5	21.93	13	.062	3.026	− 0.968	− 0.157	1.122
	Oth6	13.95	11	.039	3.396	− 0.676	0.338	1.467

*$p < .01$
[a]Item parameters in graded response model include a—discrimination, and b1–b3—step difficulties (analogous to thresholds in CFA)

Additionally, for this scale, standard errors are extremely low for the large part of the continuum covered.

Recognition by others The original nine items on the *recognition by others* scale yielded an alpha of .917 and had strong *pbis*, although the overall CFA model fit was not as strong. Additionally, classical reliability analyses did not reveal clear item candidates for removal. Thus, a combination of the overall model fit and an item level (mis)fit in IRT was used to reduce the scale to six items. While removing items reduced the alpha, it remained high at .901 with strong *pbis* values in the range of .61 to .81. Out of all four scales, the model fit for the *recognition by others* was the poorest and with the most mixed results. While the CFI/TLI values remained somewhat high, the RMSEAs and significant chi-square suggested a poor model fit (see Table 3).

Study 2
Study 2 consisted of adapting the final version of the role identity survey–engineering (RIS-E) that we described in study 1 to be used to assess STEM identity more generally. The sections that follow describe changes made to adapt the survey and the results of pilot testing the role identity survey–STEM (RIS-STEM).

Methods
Participants
Study 2 was conducted with 678 fourth–fifth grade students enrolled in a district located in the Southwestern United States (318 girls, 317 boys, 37 preferred not to answer or indicated "Other"). Students self-reported their race/ethnicity with most students identifying themselves as Hispanic/Latino(a) (34%), followed by "other" (including Pacific Islander, Indian Subcontinent, etc.; 13%), multi-racial (13%), White/Caucasian (5%),

Fig. 1 Test information functions and standard errors for identity subscales—engineering
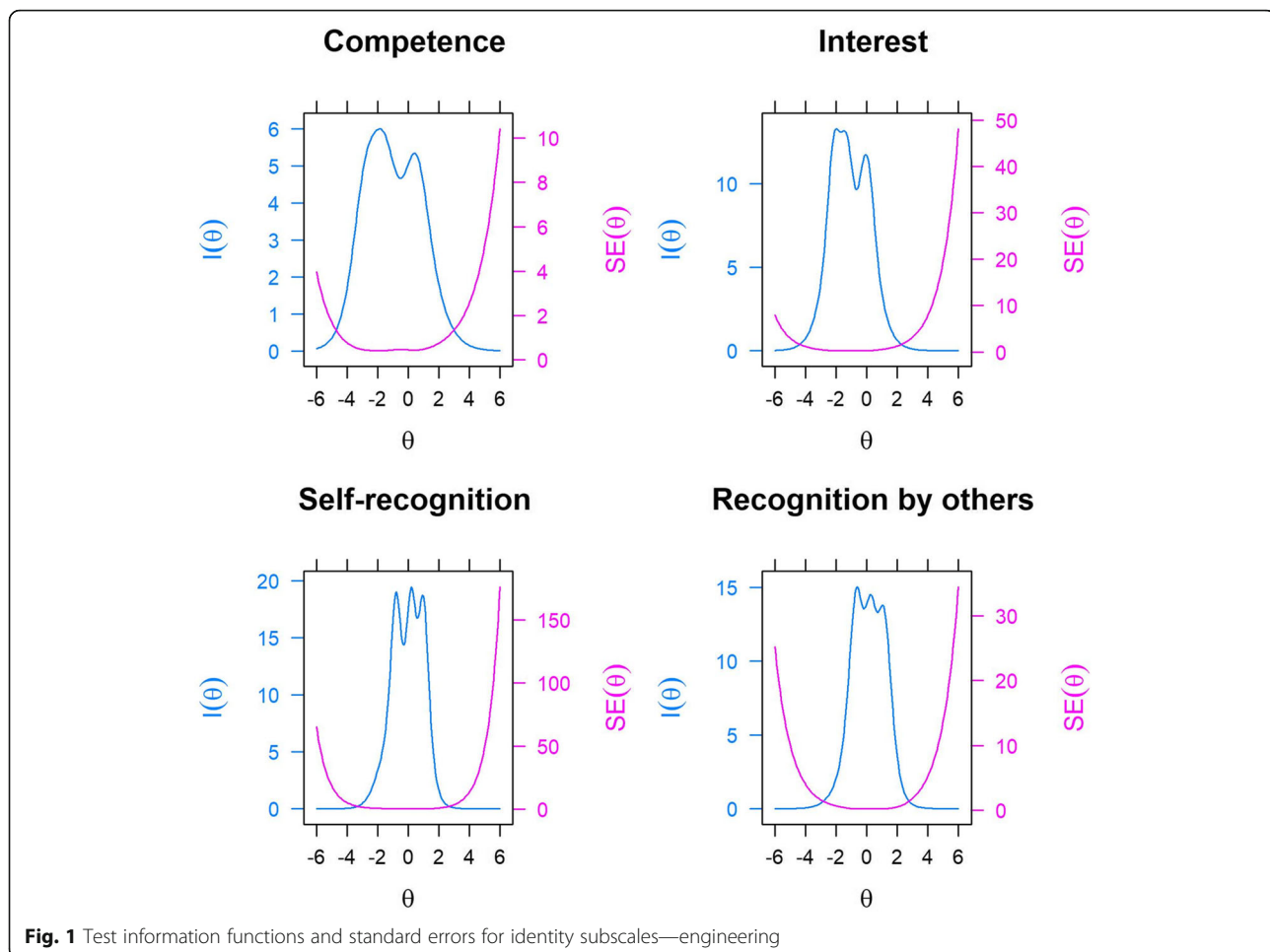
American Indian/Alaska Native (4%), Black/African American (3%), and Asian (1%). One-fourth of the students did not report their race/ethnicity (27%).

Survey adaptation and data collection procedures
As described above, the role identity survey–engineering consisted of 4 sub-scales: *competence, interest, recognition by others*, and *self-recognition*. Minor wording changes were necessary to shift the items to assess the broader notion of *STEM identity*. Less of a shift was needed for *competence* and *interest* as these items referred to engineering activities or content, which easily converted to STEM activities or content. For example, the item "I enjoy learning about *engineering*" referred to one's interest in the content of engineering. With a simple word replacement, we easily adapted this item to reflect one's interest in STEM: "I enjoy learning about *STEM*." A greater shift in item wording was necessary for the *recognition* constructs (both self and by others). In the engineering survey, items reflected students being recognized by self or others as a member/potential member of a specific career, namely as an engineer. Because there is not a career where a person is a

"STEM," simply replacing "engineering" with "STEM" was problematic. Therefore, we changed the wording from referencing a specific career (i.e., engineering) to referencing a type of person, namely a "STEM person" or a "STEM professional." Finally, we added definitions of both "STEM" and "STEM Person" to the instructions. While the acronym STEM likely is familiar to most students as school curricula and outreach activities increasingly incorporate or use this or other similar terms (e.g., STEM, STEAM), we did not want to assume that all students had the same understanding of STEM. Therefore, we identified that STEM referred to science, technology, engineering, and mathematics. Additionally, we defined the term "STEM Person" as someone who does or is good at STEM activities now or who might do STEM as part of a job one day in the future. Table 5 presents examples of item wording changes from the engineering to the STEM survey.

Analyses
We utilized the same approaches in study 2 as we did in study 1, with a caveat that our goal was to seek evidence of support for the use of engineering items into a

Table 5 Example items and wording changes by identity construct

Scales	Example items	
	RIS-E	RIS-STEM
Competence	"I am able to do well in activities that involve **engineering.**"	"I am able to do well in activities that involve **STEM.**"
Interest	"I enjoy learning about **engineering**!"	"I enjoy learning about **STEM**!"
Self-recognition	"I see myself as **an engineer**."	"I see myself as **a STEM person**."
Recognition by others	"My best friends see me as **an engineer**."	"My best friends see me as **a STEM person**."

Bold font identifies words that were changed between the RIS-E and RIS-STEM

broader STEM context, rather than further scale reduction. In other words, we wanted to see if items developed for the engineering survey (RIS-E) can be applied to a more general STEM context (RIS-STEM) while keeping the scales useful.

Results and discussion
Item analysis
Classical item analyses were conducted for each scale separately prior to fitting a single-factor CFA and IRT models. Table 6 reports reliability analyses using alpha, alpha if item deleted and *pbis*. As noted, across the scales for the construct of identity, alpha values ranged from .787 (competence) to .884 (recognition by others). Given a small number of items per any one scale (6–7 items), these alpha values are considered to be high. For all scales except interest, deletion of any one item would yield lower reliability estimates, further supporting the homogeneity of items as a set per any one scale. For the interest scale, removing one of the items would increase alpha to .818. However, after reviewing that item's content, we decided to retain all items on the interest scale since alpha was still reasonably high (.807). Lastly, the ranges of *pbis* correlations were moderate to high for all scales (see the last column of Table 6), further suggesting that the items on each scale were reasonably related to each other. The observation of somewhat weaker *pbis* values for the interest scale was consistent with the findings related to the scale's reliability.

Overall model fit
Overall CFA and IRT model fit indices for the identity scales are reported in Table 7. The agreement between the two sets of indices was high, suggesting that items associated with each scale generally supported conclusions of individual single factor models. Specifically, the

RMSEA values ranges from .068 (competence) to .141 (recognition by others), and CFI/TLI values were all reasonably high (lowest CFI/TLI values were found for self-recognition with .975 and .958, respectfully). Across all scales, *competence* and *interest* yielded more adequate model fit than the *self-recognition* and *recognition by others* scales. The SRMSR values were low for each of the four scales, with the lowest value of .040 for competence and highest of .057 for recognition by others. While χ^2 and C2 values were statistically significant for all four scales, we were less concerned given the larger sample size and the known general sensitivity of the statistics to large sample sizes.

IRT item and test level analysis
In addition to examining overall scale level model fit, we examined the item level fit using the S-X2 and RMSEA statistics. As noted in Table 8, individual items had very good model fit as most items yielded nonsignificant p values and low RMSEAs. Even when individual items yielded poorer item fit, corresponding RMSEA values were still reasonably low. Furthermore, estimates of item parameters (i.e., discrimination and item step-difficulties) suggested that items were discriminating well among the respondents (high values of a parameter) and that collectively, the items covered a reasonably large portion of the construct continuum (both negative and positive b values).

Furthermore, it is typical to examine overall test level functioning of the items at test or subtest levels via test information functions (and corresponding standard errors). As noted in Fig. 2, test information functions for the various scales reliably captured respondents within or around ± 3 range of the continuum on the standard score metric, with the most precision (reliability) around the middle of the continuum. Stated differently, the lowest standard errors were found across the desirable ± 3 range, where most of the respondents would be expected to be located.

pbis range represents the point biserial correlation between the item in question and the total scale score. The ranges presented here are across all items within a scale

α Cronbach's alpha estimate

Table 6 Reliability and item analysis for identity scales—STEM

Scales	n	No. items	α	α if deleted	*pbis* range
Competence	604	7	.787	None	.43–.58
Interest	594	7	.807	(.818)	.33–.69
Self-recognition	603	6	.837	None	.55–.66
Recognition by others	597	6	.884	None	.59–.75

Table 7 Overall model fit for identity scales—STEM

Scales	CFA					IRT					
	χ	df	RMSEA	CFI	TLI	C2[a]	df	RMSEA	CFI	TLI	SRMSR
COMP	52.5*	14	.068	.978	.966	40.5*	14	.056	.983	.975	.040
INT	55.9*	14	.071	.987	.981	32.5*	14	.047	.99	.986	.044
SELF	89.5*	9	.122	.975	.958	67.8*	9	.104	.970	.950	.049
OTHERS	115.5*	9	.141	.982	.971	79.1*	9	.114	.975	.959	.057

COMP competence, INT interest, SELF self-recognition, OTHERS recognition by others

*p < .05

[a]C2 is a variant of the M2 statistic (Maydeu-Olivares & Joe, 2006) for evaluating model fit; C2 is appropriate for polytomous models when there is not sufficient degrees for freedom (e.g., due to low small number of items)

Discussion

Our goals in the current study were twofold. First, we sought to develop a measure of engineering identity appropriate to use with elementary age youth. Given the increasing focus on engineering at younger grades, a measure appropriate to this age group is needed as most existing instruments target and are validated for students in high school and beyond. Although engineering curricula and programs may be increasing in schools, many schools and programs focus more generally on STEM. As such, our second goal was to adapt the engineering instrument to assess STEM identity and examine its

Table 8 Item fit and item parameter estimates for identity scales—STEM

Scales	Item ID	Item fit			Item parameter estimates[a]			
		S_X2*	df	RMSEA	a	b1	b2	b3
Competence	Comp1	40.02	31	.02	1.27	− 1.51	0.04	1.67
	Comp2	28.41	30	.00	1.59	− 2.01	− 1.00	0.76
	Comp3	45.39	32	.03	1.06	− 1.55	0.15	2.16
	Comp4	34.86	31	.01	1.44	− 1.86	− 0.73	0.89
	Comp5	22.52	29	.00	1.82	− 1.81	− 0.82	0.74
	Comp6	29.61	28	.01	1.78	− 2.01	− 0.84	0.91
	Comp7	32.61	30	.01	1.65	− 2.05	− 0.67	1.16
Interest	Int1	65.10	40	.03	0.75	− 3.62	− 1.97	− 0.11
	Int2	39.96	34	.02	0.85	− 4.15	− 2.66	0.18
	Int3	27.38	29	.00	2.08	− 1.51	− 0.57	0.84
	Int4	46.78	33	.03	1.28	− 1.42	0.11	1.31
	Int5	53.56	26	.04	2.73	− 1.45	− 0.72	0.64
	Int6	34.70	26	.02	2.99	− 1.37	− 0.67	0.56
	Int7	38.21	28	.02	2.32	− 1.33	− 0.45	0.69
Self-recognition	Self1	32.77	29	.01	1.59	− 1.39	− 0.18	1.05
	Self2	29.25	29	.00	1.44	− 1.80	− 0.55	0.85
	Self3	27.79	24	.02	2.46	− 0.67	0.42	1.23
	Self4	25.55	24	.01	2.71	− 0.64	0.46	1.23
	Self5	33.46	25	.02	2.29	− 0.72	0.43	1.24
	Self6	33.11	30	.01	1.59	− 1.20	− 0.14	1.07
Recognition by others	Oth1	107.60*	27	.07	2.26	− 1.07	− 0.03	0.91
	Oth2	42.54	22	.04	3.36	− 0.71	0.19	0.96
	Oth3	49.98*	23	.04	3.09	− 0.78	0.10	0.77
	Oth4	63.02*	29	.04	1.62	− 1.34	− 0.11	0.97
	Oth5	85.10*	27	.06	2.25	− 1.01	0.05	1.13
	Oth6	34.50	19	.04	3.39	− 0.63	0.30	1.14

*p < .01

[a]Item parameters in graded response model include a—discrimination and b1–b3—step difficulties (analogous to thresholds in CFA)
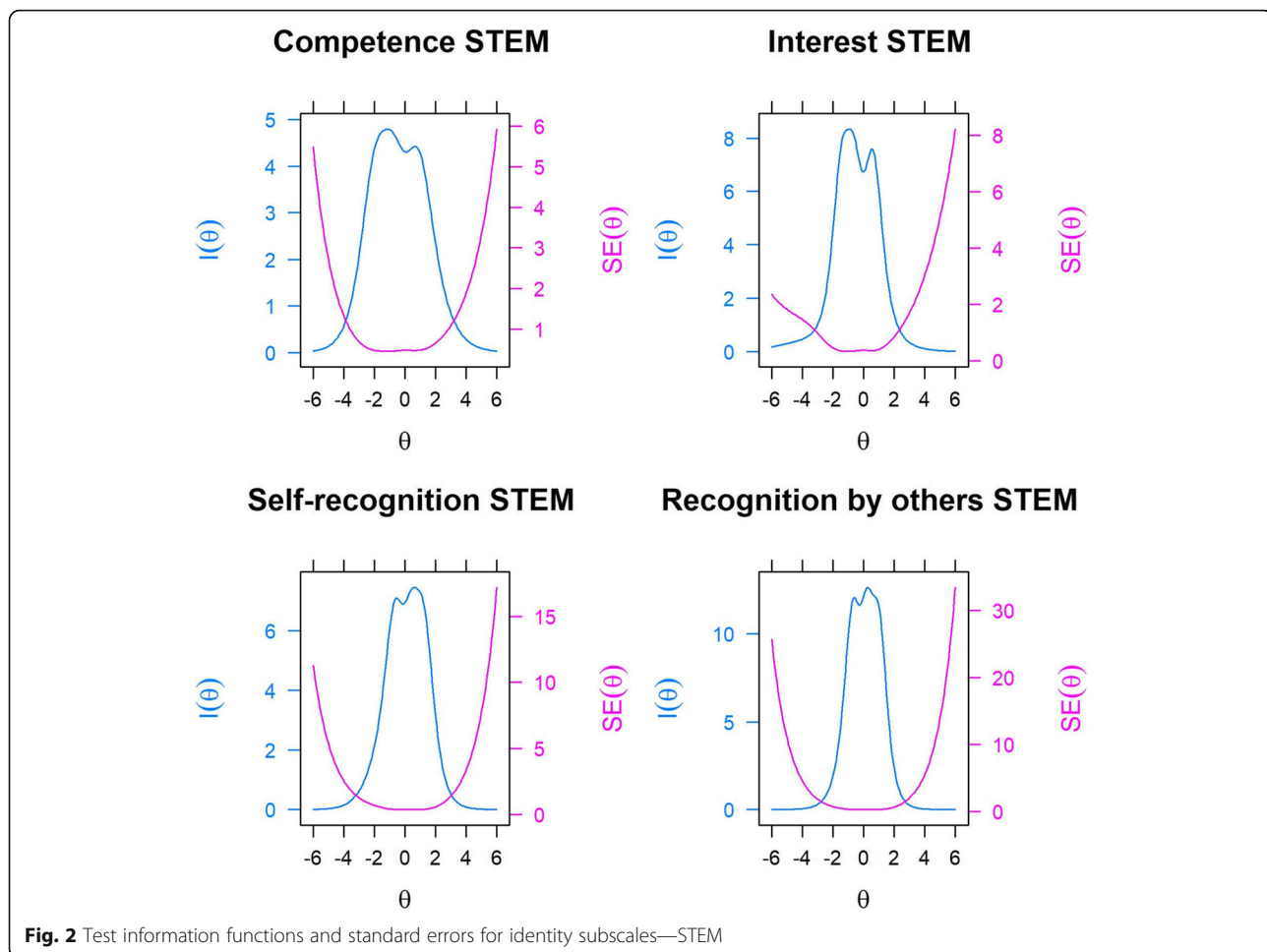
Fig. 2 Test information functions and standard errors for identity subscales—STEM

appropriateness in this more general context. To accomplish these goals, we carefully developed items to assess constructs related to one's role-related identity, drawing from previously validated instruments when possible. After multiple rounds of pilot testing, both the role identity survey–engineering (RIS-E) and role identity survey–STEM (RIS-STEM) were found to consist of four aspects of one's identity (*interest, competence, self-recognition*, and *recognition by others*) and to demonstrate strong psychometric properties. As such, we are confident that both versions produce reliable scores that measure aspects of identity (engineering and STEM) in elementary students. We are hopeful that these instruments can be used to measure changes over time, but the reliability of the RIS-E and RIS-STEM has not been examined over extended periods. Future research should continue to examine how scores vary over time.

Components of identity

The four aspects of identity on the RIS-E and RIS-STEM were consistent with prior work on identity development (Carlone & Johnson, 2007; Godwin, 2016; Hazari et al., 2010). *Interest* captured an individual's enjoyment in doing engineering or STEM activities, while *competence* reflected students' ability when doing these activities. While our surveys are consistent with work by Carlone and Johnson (2007) in that recognition was differentiated into *self-recognition* and *recognition by others*, we differed slightly in how we assessed recognition by others. One limitation of the Carlone and Johnson framework is its focus on recognition by meaningful others in science. Students in elementary and middle school may not have contact with such individuals in their lives. Kim and Sinatra (2018) suggest that it is more useful to view one's science identity instead as a social identity, thus allowing for recognition by others to include family and peers, both of which may have a stronger influence on students in younger age groups. Our survey items are consistent with this view such that we measure recognition by others by asking about teachers, family, and members of one's peer group, and we only ask one item that assesses recognition by a "meaningful other in science" when we ask students about recognition by a STEM teacher or STEM outreach provider.

Conceptualizations of engineering and STEM

We developed the survey to assess students' perceptions of themselves as engineers. We then extended this to the context of STEM more broadly. However, we recognize the likelihood that elementary students will interpret and understand engineering and STEM differently. We fully expect that students (similar to many adults) will have alternative conceptions about engineering that follow many common stereotypes (e.g., engineers build bridges, repair cars) and that they will not be fully aware of the breadth of the field of engineering, which is consistent with those that have been documented elsewhere (Capobianco, Diefes-Dux, Mena, & Weller, 2011; Cunningham, Lachapelle, & Lindgren-Streicher, 2005; Fralick, Kearn, Thompson, & Lyons, 2009; Lachapelle, Phadnis, Hertel, & Cunningham, 2012). Similarly, we do not expect elementary students to have a nuanced understanding of STEM, especially since defining STEM is an issue that adults also wrestle with (Staus et al., 2019). However, given the increased focus on and inclusion of "STEM" in elementary grades, we felt that the majority of students will have some basic familiarity with STEM and have at least a general sense that it is a mix of two or more of its component disciplinary fields. We suspect that it is most likely that students conceptualize STEM predominately in terms of science, followed by mathematics, and then varying levels of technology and engineering (e.g., SteM), as this reflects their greater familiarity with science and mathematics due to the emphasis on these subjects in both formal and informal education settings. Conceptualizations of engineering, though, are likely to be even more constrained than those of STEM, especially in elementary students. We expect that the inclusion of engineering practices in the Next Generation Science Standards (NGSS) has improved this situation over the last few years, though our data and findings of others confirm that many stereotypes about engineering continue to exist in this population.

Despite the fact that students likely hold alternative conceptions of engineering (and STEM), we feel that there is enough of a common thread in the understandings held by students that asking about their perceptions still has value. In an ideal world, we would assess students' identity based on a universally agreed upon understanding of engineering/STEM, but this is unrealistic, for both youth and adults. Students' knowledge and conceptions of content areas and related careers constantly change as they continue to learn more about various disciplines. Although knowing how students understand engineering/STEM is important, and is indeed part of our research, this is a different question than knowing whether students perceive themselves as an engineer or STEM person. Whether or not their understanding and

core beliefs about engineering/STEM are accurate, it is these core beliefs that influence students' decisions about who they are and want to be in their future.

Limitations and next steps

We developed our survey building on the framework proposed by Carlone and Johnson (2007) and expanded upon by Hazari et al. (2010). However, some researchers have suggested gaps with this framework in assessing identity, namely that it does not include the role of the environment on an individual's developing STEM identity (Kim & Sinatra, 2018; Kim, Sinatra, & Seyranian, 2018). Therefore, one limitation is that our survey utilizes only one representation of identity but that other frameworks may provide different perspectives (e.g., Social Cognitive Career Theory [SCCT]; Lent et al., 1994; Lent et al., 2003) that include additional factors related to STEM identity development, such as environment or contextual influences.

A second limitation relates to how students conceptualize engineering and STEM. While we would like students to respond from a common and "accurate" understanding of engineering and/or STEM, we know this is not the reality. Students may not have a perfect understanding of engineering, but they do seem to be generally oriented in the right direction and distinguish between engineers and scientists in meaningful ways, such as the type of work engaged in (Fralick et al., 2009). Additionally, greater numbers of students have the opportunity to learn about engineering and STEM as schools increasingly incorporate the Next Generation Science Standards and programs such as Engineering is Elementary and Project Lead the Way into their curricula. Whether their core beliefs about engineering/STEM are accurate or even consistent with those of others, students draw upon these core beliefs as they form their identities and begin to define who they are and want to be in the future. Therefore, we think there is still value in assessing their role-related identities.

Because the RIS-STEM was adapted from the more specific engineering survey, many of the items skew toward engineering. Additionally, we did not attempt to have an equal number of items to represent each individual discipline of STEM. As such, the "STEM" version may skew more towards "StEm." Our results indicated that the STEM survey is appropriate for using in a more general STEM setting. Future work could continue to investigate the adaptability of the survey to other STEM-related domains (e.g., computer science).

Despite the instruments being piloted on 3rd–5th grade students, we are confident that it can be used with students up through 12th grade. Future work should examine the usefulness of the surveys for students with middle and high school students. For researchers

considering using the survey with students younger than 3rd grade, we recommend that you do so with caution as it may be too difficult for them in at least two foreseeable ways. First, some of the concepts may be new to younger students who have not yet started to develop role-related interests or identities. Second, the reading level skews older, especially for the engineering survey (primarily due to the number of syllables and length of some of the items).

Finally, while the age groups of the students in both studies was the same (3rd–5th grade), the demographics of the students differed between participants who piloted the engineering and STEM surveys. Next steps should include continuing to investigate how both surveys function across diverse student populations.

Conclusions

We sought to develop an instrument to assess engineering identity in elementary students in order to study the efficacy of interventions targeted at improving identification with engineering in children. Next, we explored its usefulness in a STEM context more broadly. The results of the two studies indicate that our surveys are appropriate measures of engineering and STEM identity in youth as young as elementary school. The resulting survey instruments, the RIS-E and RIS-STEM, assess four aspects of identity (competence, interest, self-recognition, and recognition by others) that are consistent with and build upon conceptions of identity as posited in the literature (Carlone & Johnson, 2007; Hazari et al., 2010; Kim & Sinatra, 2018). The novel contribution to STEM education is the development of instruments that are age-appropriate for elementary students that can be used to assess identity related to specific disciplines (e.g., engineering) or STEM more broadly. As programs and curricula are designed and implemented at younger ages, it will be important to be able to assess the impact such interventions have on the development of a STEM identity in younger students. These surveys can provide a singular indicator of identity at one point in time or can be used to assess changes in youth's STEM identity (e.g., before and after an intervention or across grades as youth progress through school). Additionally, assessing four aspects of one's identity can help identify specific areas that programs and curriculum can target to help increase youth's identity, such as fostering greater interest or building youth's competence in engineering or STEM. The RIS-E and RIS-STEM can support researchers and practitioners to better understand youth's identity development, especially in engineering and STEM. Because youth begin to shape their role-related identities as early as elementary school, greater insight is needed about youth's identity at these ages as we continue to examine the connection between identity and

intentions to pursue future study or careers in engineering or STEM.

Supplementary information

> **Additional file 1.** Items on the Role Identity Survey – Engineering (RIS-E). Items on the Role Identity Survey – STEM (RIS-STEM)

Abbreviations

CFA: Confirmatory factor analyses; CFI: Comparative fit index; CTT: Classical test theory; df: Degrees of freedom; EFA: Exploratory factor analyses; EIA: Engineering Interest and Attitudes Survey; EIDS: Engineering Identity Development Scale; GRM: Graded response model; IRT: Item response theory; M-ATSI: Modified Attitudes toward Science Inventory; NGSS: Next Generation Science Standards; pbis: Point biserial; PCA: Principal components analysis; PRiSE: Persistence Research in Science & Engineering survey; RMSEA: Root mean square error of approximation; SCCT: Social Cognitive Career Theory; SRMSR: Standardized root mean square residual; STEM-CIS: STEM Career Interest Survey; S_X2: Signed chi-square test; TLI: Tucker Lewis index

Authors' contributions

KP contributed to the literature review, survey development, data collection, data analyses, and writing the manuscript. AM contributed to the survey development and editing and review of the manuscript. DSV contributed to conducting quantitative data analysis (item analysis, CFA, IRT) and writing of the manuscript related to results from those analyses. All authors read and approved the final manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author upon request.

Ethics approval and consent to participate

Parents and students participating in the outreach program provided consent prior to data collection. Additionally, students were given the opportunity to decline to participate in the survey during administration. The Tufts University IRB approved the methodology for data collected from students participating in study 1.

For study 2, the survey was utilized by a large district in the Southwestern United States as part of their efforts to monitor their STEM programs. Students were given the opportunity to decline to participate in the survey during administration. Authors requested and were granted access to these anonymized data after they had been collected.

Competing interests

The authors declare that they have no competing interests.

References

American College Testing (2017). STEM education in the U.S.: Where we are and what we can do. ACT. https://www.act.org/content/dam/act/unsecured/documents/STEM/2017/STEM-Education-in-the-US-2017.pdf

Appianing, J., & Van Eck, R.N. (2018). Development and validation of the value-expectancy STEM assessment scale for students in higher education,

International Journal of STEM Education, 5(24). doi:https://doi.org/10.1186/s40594-018-0121-8

Archer, L., DeWitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). "Doing" science versus "being" a scientist: Examining 10/11-year-old schoolchildren's constructions of science through the lens of identity. *Science Education, 94,* 617–639.

Archer, L., Moote, J., Francis, B., DeWitt, J., & Yeomans, L. (2017). The 'exceptional' physics girl: A sociological analysis of multimethod data from young women aged 10-16 to explore gendered patterns of post-16 participation. *American Educational Research Journal, 54*(1), 88–126.

Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation and aspirations in science, engineering, and medicine. *Journal of Research in Science Teaching, 47*(5), 564–582.

Brickhouse, N. W., Lowery, P., & Schultz, K. (2000). What kind of girl does science? The construction of school science identities. *Journal of Research in Science Teaching, 37*(5), 441–458.

Burke, P. J., & Stets, J. E. (2009). *Identity theory.* New York: Oxford University Press.

Capobianco, B. M., & Deemer, E. (2017). Analyzing predictors of children's formative engineering identity development. *International Journal of Engineering Education, 33*(1A), 44–54.

Capobianco, B. M., Diefes-Dux, H. A., Mena, I., & Weller, J. (2011). What is an engineer? Implications of elementary school student conceptions for engineering education. *Journal of Engineering Education, 100*(2), 304–328.

Capobianco, B. M., French, B. F., & Diefes-Dux, H. A. (2012). Engineering identity development among pre-adolescent learners. *Journal of Engineering Education, 101*(4), 698–716.

Carlone, H. B., & Johnson, A. (2007). Understanding the science experiences of successful women of color: Science identity as an analytic lens. *Journal of Research in Science Teaching, 44*(8), 1187–1218.

Carlone, H. B., Johnson, A. C., & Scott, C. M. (2015). Agency amidst formidable structures: How girls perform gender in science class. *Journal of Research in Science Teaching, 52*(4), 474–488.

Carlone, H. B., Scott, C. M., & Lowder, C. (2014). Becoming (less) scientific: A longitudinal study of students' identity work from elementary to middle school science. *Journal of Research in Science Teaching, 51*(7), 836–869.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06.

Chen, Y.F., Cannady, M.A., Schunn, C., & Dorph, R. (2017). Measures technical brief: Fascination in STEM. Retrieved from: http://activationlab.org/wp-content/uploads/2018/03/Fascination_STEM-Report_20170403.pdf. Accessed 2 Feb 2018.

Chung, J., Cannady, M.A., Schunn, C., Dorph, R., & Vincent-Ruz, P. (2016). Measures technical brief: Competency beliefs in science. Retrieved from: http://activationlab.org/wp-content/uploads/2018/03/Competency-Beliefs-Report-3.2-20160331.pdf. Accessed 6 Feb 2018.

Cunningham, C.M., Lachapelle, C., & Lindgren-Streicher, A. (2005). Assessing elementary school students' conceptions of engineering and technology. *Proceedings from* the American Society for Engineering Education Annual Conference and Exposition, Portland, OR.

DeWitt, J., & Archer, L. (2015). Who aspires to a science career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education, 37*(12), 2170–2192.

Doerschuk, P., Bahrim, C., Daniel, J., Kruger, J., Mann, J., & Martin, C. (2016). Closing the gaps and filling the STEM pipeline: A multidisciplinary approach. *Journal of Science Education and Technology, 25*(4), 682-695.

Eccles, J. S. (2007). Where are all the women? Gender differences in participation in physical science and engineering. In S. J. Ceci, & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence*, (pp. 199–210). Washington, DC: American Psychological Association.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44,* 77–89.

Eccles, J.S., (2011). Gendered educational and occupational choices: Applying Eccles et. al model of achievement-related choices. *International Journal of Behavioral Development, 35*(3), 195-201.

Fralick, B., Kearn, J., Thompson, S., & Lyons, J. (2009). How middle schoolers draw engineers and scientists. *Journal of Science Education and Technology, 18*(1), 60–73.

Godwin, A. (2016). The development of a measure of engineering identity. *Proceedings from the American Society for Engineering Education Annual Conference and Exposition,* New Orleans, LA.

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2013). Understanding engineering identity through structural equation modeling, in *2013 IEEE Frontiers in Education Conference (FIE).* doi:https://doi.org/10.1109/FIE.2013.6684787

Godwin, A., Potvin, G., Hazari, Z., & Lock, R. (2016). Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice. *Journal of Engineering Education, 105*(2), 312–340.

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.

Hazari, Z., Sonnert, G., Sadler, P. M., & Shanahan, M.-C. (2010). Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study. *Journal of Research in Science Teaching, 47*(8), 978–1003.

Higgins, M., Hertel, J. D., Shams, M. F., Lachapelle, C. P., & Cunningham, C. M. (2015). *NASA MISSION grant: Engineering adventures unit development (Final Report).* Boston, MA: Museum of Science http://www.eie.org/sites/default/files/downloads/EiE/ResearchPublications/higgins_et_al_2015_nasa_grant.pdf.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Kang, T. & Chen, T. (2007). An investigation of the performance of the generalized S-X2 item-fit index for polytomous IRT models. ACT Research Report Series. https://files.eric.ed.gov/fulltext/ED510479.pdf. Accessed 4 Mar 2020.

Kier, M. W., Blanchard, M. R., Osborne, J. W., & Albert, J. L. (2014). The development of the STEM Career Interest Survey (STEM-CIS). *Research in Science Education, 44,* 461–481.

Kim, A. Y., & Sinatra, G. M. (2018). Science identity development: An interactionist approach. *International Journal of STEM Education, 5*(51). https://doi.org/10.1186/s40594-018-0121-8.

Kim, A. Y., Sinatra, G. M., & Seyranian, V. (2018). Developing a STEM identity among young women: A social identity perspective. *Review of Educational Research, 88*(4), 589–625.

Lachapelle, C.P., Phadnis, P., Hertel, J., & Cunningham, C.M. (2012). What is engineering? A survey of elementary students. Paper presented at the 2nd P-12 Engineering and Design Education Research Summit, Washington, DC.

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Towards a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45,* 79–122.

Lent, R. W., Brown, S. D., & Hackett, G. (1996). Career development from a social cognitive perspective. In D. Brown, L. Brooks, et al. (Eds.), *Career choice and development,* (3rd ed., pp. 373–422). San Francisco: Josey-Bass.

Lent, R. W., Brown, S. D., Schmidt, J., Brenner, B., Lyons, H., & Treistman, D. (2003). Relation of contextual supports and barriers to choice behavior in engineering majors: Test of alternative social cognitive models. *Journal of Counseling Psychology, 50*(4), 458–465.

Lockard, B., & Wolf, M. (2012). Employment outlook: Occupational employment projections to 2020. *Monthly Labor Review, January,* 84-108. Retrieved from https://www.bls.gov/opub/mlr/2012/01/mlr201201.pdf

Maltese, A. V., & Cooper, C. S. (2017). STEM pathways: Do men and women differ in why they enter and exit? *AERA Open, 3*(3), 1–16.

Maltese, A. V., Melki, C. S., & Wiebke, H. (2014). The nature of experiences responsible for the generation and maintenance of interest in STEM. *Science Education, 98*(6), 937–962.

Maltese, A.V., & Tai, R.H. (2010). Eyeballs in the fridge: Sources of early interest in science. *International Journal of Science Education, 32*(5), 669-685. http://dx.doi.org/https://doi.org/10.1080/09500690902792385

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*(5), 877–907.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71,* 713–732.

Moore, D.B., Bathgate, M., Chung, J., & Cannady, M.A., (2011). Technical report: Measuring activation and engagement. Activation Lab, Enables Success Study.

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science* Review, 84, 109-116.

Orlando, M., & Thissen, D. (2000). New item fit indices for dichotomous item response theory models. *Applied Psychological Measurement, 24,* 50–64.

Orlando, M., & Thissen, D. (2003). Further investigation of the performance of S-X2: An item fit index for use with dichotomous item response theory models. *Applied Psychological Measurement, 27,* 289–298.

Patrick, A.D., & Borrego, M. (2016). A review of the literature relevant to engineering identity, in Proceedings of the American Society for Engineering Education Annual Conference and Exposition, New Orleans, LA.

Pierrakos, O., Beam, T.K., Constantz, J., Johri, A. & Anderson, R. (2009). On the development of a professional identity: Engineering persisters vs. engineering switchers. *39th Annual Frontiers in Education Conference,* San Antonio, TX.

President's Council of Advisors on Science and Technology (2012). *Report to the president.* Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/fact_sheet_final.pdf

R Core Team (2019). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundations for Statistical Computing https://www.R-project.org.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. http://www.jstatsoft.org/v48/i02

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*, 100–114.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*(4), 367–373.

Sha, L., Schunn, C., Bathgate, M., & Ben-Eliyahu, A. (2015). Families support their children's success in science learning by influencing interest and self-efficacy. *Journal of Research in Science Teaching*, *53*(3), 450–472.

Simpson, A., & Bouhafa, Y. (2020). Youths' and adults' identity in STEM: A systematic literature review. *Journal for STEM Education Research*. https://doi.org/https://doi.org/10.1007/s41979-020-00034-y

Staus, N. L., Lesseig, K., Lamb, R., Falk, J., & Dierking, L. (2019). Validation of a measure of STEM interest for adolescents. *International Journal of Science and Mathematics Education*. https://doi.org/10.1007/s10763-019-09970-7.

Stets, J. E., Brenner, P. S., Burke, P. J., & Serpe, R. T. (2017). The science identity and entering a science occupation. *Social Science Research*, *63*, 1–14.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*, 116–131.

Tabachnick, B. G., & Fiddell, L. S. (2001). *Using multivariate statistics (4th Ed.)*. New York: Allyn and Bacon.

Trujillo, G., & Tanner, K. D. (2014). Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity. *CBE – Life Sciences Education*, *13*, 6–15.

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education*, *18*(2), 345–368.

U.S. Bureau of Labor Statistics (n.d.). *Employment in STEM occupations*. U.S. Bureau of Labor Statistics. Retrieved March 3, 2020, from https://www.bls.gov/emp/tables/stem-employment.htm

Vincent-Ruz, P., & Schunn, C. D. (2018). The nature of science identity and its role as the driver of student choices. *International Journal of STEM Education*, *5*, 48.

Wagstaff, I. R. (2014). *Predicting 9th graders' science self-efficacy and STEM career intent: A multilevel approach* (Publication No. 3584390) [Doctoral dissertation, North Carolina State University]. ProQuest Dissertations Publishing.

Weinburgh, M. E., & Steele, D. (2000). The modified attitudes toward science inventory: Developing an instrument to be used with fifth grade urban students. *Journal of Women and Minorities in Science and Engineering*, *6*(1), 87–94.

Wigfield, A., & Eccles, J.S., (2000). Expectancy: Value theory of motivation. *Contemporary Educational Psychology, 25*, 68-81.

Willse, J. T. (2018). CTT: Classical test theory functions. *R Package Version* 2.3.3 https://CRAN.R-project.org/package=CTT.

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*, 189–194.

Publisher's Note