IJEMST
International Journal of Education in Mathematics, Science and Technology

ISSN: 2147-611X

International Journal of Education in Mathematics, Science and Technology (IJEMST)

www.ijemst.com

An Instrument for Examining Elementary Engineering Student Interests and Attitudes

Cathy P. Lachapelle[1], Robert T. Brennan[2]
[1]Museum of Science, Boston
[2]Harvard T. H. Chan School of Public Health

To cite this article:

Lachapelle, C.P. & Brennan, R.T. (2018). An instrument for examining elementary engineering student interests and attitudes. *International Journal of Education in Mathematics, Science and Technology (IJEMST)*, *6*(3), 221-240. DOI: 10.18404/ijemst.428171

International Journal of Education in Mathematics, Science and Technology

Volume 6, Number 3, 2018 DOI:10.18404/ijemst.428171

An Instrument for Examining Elementary Students' Engineering Interests and Attitudes

Cathy P. Lachapelle, Robert T. Brennan

Article Info

Article History

Received:
22 July 2017

Accepted:
01 February 2018

Keywords

Elementary school
Engineering education
Student interests
Student attitudes
Gender differences

Abstract

Engineers and policymakers have expressed concern that too few students enter the engineering pipeline. This has led to many efforts to engage students in engineering in after-school programs, summer programs, and more recently, in school curricula. The expectation is that, through these efforts, greater numbers of more demographically diverse children will become aware of engineering as a possible career option, and some will decide to pursue it, thereby increasing and diversifying the population pursuing engineering careers. This expectation makes the assumption that students will become more interested in and form more positive attitudes towards engineering as they encounter it in formal and informal settings. To measure this assumption, we have developed an Engineering Interest and Attitudes (EIA) survey, drawing from earlier surveys used to measure student interest in and attitudes toward science. We show that the subscales developed from EFA and CFA are reliable, and considerable evidence is present for the validity of use of EIA for measuring young students' engineering interests and attitudes. We also present evidence that EIA can be used by researchers and curriculum developers with students ages 8-11 to measure change in student interests and attitudes towards the goal of evaluating engineering activities, programs, and curricula.

Introduction

Context

Engineers and policymakers have expressed concern that too few students enter the engineering pipeline. The Organization for Economic Co-operation and Development notes that the proportion of students in OECD countries choosing to enter all STEM fields has been dropping since the mid-1990s (OECD, 2008). A recent report by the U.S. National Academies summarizes reports and surveys of employers, industry groups, and government agencies that have expressed concern about an insufficient supply of engineers and other skilled technology workers (National Academies of Sciences, Engineering, and Medicine, 2017).

An additional concern is the lack of diversity among engineers (Buccheri, Gurber, & Bruhwiler, 2011; National Academy of Sciences, National Academy of Engineering (NAE), & Institute of Medicine, 2010; National Research Council & NAE, 2014). Women are underrepresented in most nations, and in the United States there is particular concern about the paucity of African American and Latino/a engineers compared to the general population.

It can be argued that an important time to introduce children to career options is during childhood. Research shows that many engineers and scientists form their career choices before adolescence (Maltese & Tai, 2010; Royal Society, 2006; Venville, Wallace, Rennie, & Malone, 2002). Using longitudinal datasets and methods, several researchers have found that before children reach middle school they already have settled on a career path, whether or not that will be a STEM field (Lindahl, 2007; Lyons, 2006; Tai, Liu, Maltese, & Fan, 2006). Children's interest in and aptitude for science has generally been found to be high for both girls and boys younger than 10; however, interest drops over time as children progress through school (Murphy & Beggs, 2003). The drop in interest is particularly pronounced for female students—among adolescent and older students, the attitudes of males toward the physical sciences and engineering are consistently more positive than those of females (Tytler, 2014). This may be due to the content of the curriculum, which often does not connect well to the concerns of people and societies—concerns that girls consistently rate as more compelling than content detached from such concerns, particularly as compared to boys (Burke, 2007; Häussler & Hoffmann, 2002). College-bound girls have been shown to prefer biological sciences and engineering majors, particularly those relating to health careers, the environment, and other "helping" professions (Buccheri, Gurber, & Bruhwiler, 2011; Drechsel, Carstensen, & Prenzel, 2011; Miller, Blessing, & Schwartz, 2006).

Children's Interests in and Attitudes toward Science and Engineering

Vaughan and Hogg (2013, p. 169) explain that "Theories of attitude structure generally agree that attitudes are lasting general evaluations of socially significant objects (including people and issues)." In engineering education, socially significant objects include engineers and the work of engineering, which can have social significance both for students themselves—their life experiences and future expectations—and in their effects on aspects of the world that matter to students, such as transportation, the environment, or medicine. In this paper, we focus on student attitudes toward three socially significant objects: engineers, engineering as a profession, and learning experiences in engineering. Research on student attitudes conducted in science education shows that it is important to attend separately to students' attitudes toward school science versus science and scientists more generally, because these can be quite different, and students' attitudes toward each can vary accordingly (Lindahl, 2007; Tytler, 2014), and we have taken this finding into consideration as we investigate children's attitudes toward engineering.

Most research has focused on interest in and attitudes toward science, though some findings have extended to STEM careers more generally. Despite the relative lack of work specifically on young students' engineering interests, engineering advocates interested in increasing the flow of students through the engineering pipeline have chosen to see science findings as applicable to engineering, leading to many efforts to engage pre-adolescent students in engineering in after-school programs, summer programs, and, more recently, in-school curricula. The expectation is that by engaging students in engineering greater numbers of more demographically diverse children will become aware of it as a career option, and some students will find a special affinity to engineering and ultimately pursue it. Given the goal of increasing interest in engineering through interventions, it is important to develop instruments capable of measuring change in student attitudes toward and interest in engineering for a given intervention.

Most available STEM attitude measures, like most STEM research studies, have focused on student attitudes toward science, as well as their interest in (or aspirations toward) future study of science or careers in STEM fields. According to two recent literature reviews, the quality and validity evidence for these surveys of attitudes varies greatly (Blalock et al., 2008; Tytler, 2014). Among the topics surveyed by science attitudes instruments as reported in these reviews are: (1) desire to learn science, (2) interest in science careers, (3) positive emotions (e.g., enjoyment or "liking") toward science generally, (4) positive emotions toward doing science in school, (5) valuing of the work of scientists and the outcomes of science, and (6) valuing scientific perspectives.

The most common instrument found in the literature (Tytler, 2014) is the *Scientific Attitude Inventory* (SAI), which was designed for use with middle and high school students (Moore & Sutman, 1970). This instrument was later revised and improved as the SAI-II (Moore & Foy, 1997). However, the revision did not result in a factor structure that matched the author's original theorized structure of 12 factors (six factors with a positive and negative version of each). Lichtenstein and colleagues revisited the survey (2008) with a new sample collected from more than 500 middle and high school students; using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), they found and confirmed a structure of three factors with only one of the three having acceptable psychometric properties.

The instrument with the strongest characteristics (Blalock et al., 2008) was found to be the *Attitude toward Science in School Assessment* (ATSSA), which was designed for use with high school students (Germann, 1988). The purpose of the ATSSA is to assess the attitudes of adolescent students toward school science. As with all the instruments we surveyed, the ATSSA employs a five-point Likert scale. Germann chose 14 items from an initial set of 34 based on expert review. EFA on those 14 items resulted in a single factor with high internal consistency (Cronbach's α>.95 in four experimental samples).

The *Simpson-Troost Attitude Questionnaire* (STAQ), which was developed originally for use with high school students, was designed to measure changes in students' commitment to learning science over time, and to identify influences on students' commitment to and interest in science among the student's teachers, peers, and family (Simpson & Troost, 1982). Recently, the instrument was reevaluated and shortened from 58 items in 14 subscales to 22 items in 5 subscales using methods of EFA and CFA (Owen et al., 2008).

Because of our interest in younger students, we paid particular attention to the *modified Attitudes Toward Science Inventory* (mATSI), which assesses changes in the attitudes of urban fifth-grade students (ages 10–11) due to an intervention (Weinburgh & Steele, 2000). This instrument is a simplified version of the ATSI, developed for use with college students not majoring in science (Gogolin & Swartz, 1992); Weinburgh and Steele cut questions and simplified phrasing of items to be appropriate for pre-adolescent children. The mATSI designates 5 subscales, including one addressing students' attitudes toward school science and another asking about the value of science to the world.

The *Middle School Students' Attitude to Mathematics, Science, and Engineering* (MS-AMSE) Survey was developed to investigate students' interest in and knowledge about potential careers in engineering (Gibbons, Hirsch, Kimmel, Rockland, & Bloom, 2004). The survey was adapted from a longer version developed for use with high school students (Hirsch, Gibbons, Kimmel, Rockland, & Bloom, 2003). In addition to asking students about engineering careers, the survey included items probing students' attitudes and feelings of efficacy toward mathematics and science.

We adapted the MS-AMSE (Cunningham & Lachapelle, 2010) for use with elementary school students to measure attitudes toward and interest in engineering careers before and after participation in the Engineering is Elementary curriculum (EiE). We used this instrument, the Elementary Engineering Attitudes (EEA) survey, a precursor to the EIA, throughout the development of EiE as part of formative evaluation, and found that girls showed interest in more socially or environmentally responsive engineering fields (e.g., biomedical engineering) while boys were more likely to express interest in engineering of vehicles or structures. We also found that interests and attitudes of EiE participants became more positive, with the attitudes of girls lower on the pretest than those of boys; however, the gap in interest and attitudes closed after participation. Pretest scores for subscales had much lower reliability, however, than posttest scores.

Our research presumes that student interest and attitudes toward engineering will vary with the context. A personally relevant, engaging context is likely to affect students' attitudes positively (Ainley & Ainley, 2011); but some of the impact may be of short duration. An intervention that focuses primarily on "fun," in particular, may have only short-term effects (Appelbaum & Clark, 2001). Therefore, an instrument (and an intervention) should focus on more aspects of attitude and interest than simply the emotional impact of an intervention.

Purpose

Many proponents of increasing student exposure to engineering claim that introducing engineering to greater numbers of young students will increase and diversify the population pursuing engineering careers. An important assumption of this claim is that students will become more interested in engineering and more positive in their attitudes as they engage in engineering experiences in and out of school. To measure this assumption, we have developed an Engineering Interest and Attitudes (EIA) questionnaire, intended to be used to measure the impact of an engineering intervention on the interests, attitudes, and gender biases of elementary school students.

In this paper, we lay out the evidence for the quality of the EIA instrument. This includes evidence for internal consistency reliability and validity of the subscales. We detail evidence of content validity, including the use of prior instruments and research to form the questionnaire, and the interpretation and judgment of survey questions by content experts. We also include evidence for response processes gathered from individual interviews with students, and we describe evidence based on the internal structure of the instrument by comparing our original theoretical constructs with the results of EFA and CFA.

Method

Research

The EIA questionnaire was designed in the context of an efficacy study of the EiE curriculum, Exploring the Efficacy of Engineering is Elementary (E4). EiE had been under intensive development and formative evaluation from 2004 to 2010. The E4 study was designed as a cluster randomized trial (CRT) between EiE and a comparison curriculum. The study collected data on student achievement outcomes and fidelity of implementation, as well as student interests and attitudes in engineering.

Students participated in either the treatment or comparison engineering curriculum. Teacher volunteers were recruited for this study through their principals and superintendents. Teachers applied to participate as teams of 2–4 teachers from the same school. Only teachers from schools that had not implemented engineering curricula were accepted. Once the recruitment and acceptance process was completed, cohorts of teachers at the school level were randomized into either the treatment or comparison group.

The treatment curriculum is designed from a social constructivist theoretical framework, based on the belief that students learn deeply the key practices and content of a discipline through meaningful engagement in its epistemic practices at a developmentally appropriate level (Duit & Treagust, 1998; Sawyer, 2006). The treatment curriculum meets the criteria for project-based learning, where students focus on a design challenge that engages them with key ideas in science and engineering. The central project is open-ended, where students are engaged in the problem with a realistic context, and heavy scaffolding is provided to students, to support them as they use engineering practices and reasoning. Although the comparison curriculum also includes hands-on challenges, the challenges are not motivated with a context, no scaffolding is provided, many challenges are not open-ended, and information is given through direct instruction.

Instrument Development

The E4 project was to collect data from upper elementary students aged 8–11, so the E4 project team searched the literature for instruments addressing interest in and attitudes toward engineering and science that were designed for younger populations of students. Some of the best instruments we found, however, were designed for older students, those in middle and high school. From such instruments, we chose scales with simpler phrasing, and avoided scales with complex language that we deemed would exceed the reading abilities of our younger subjects.

Instruments addressing science were more abundant and better tested than those addressing engineering, so we decided to duplicate our chosen science scales and items, replacing the word "science" with "engineering." We initially worked to assemble an instrument that included both science and engineering items, because our units were testing both science and engineering content knowledge—teachers participating in E4 were required to teach science content that was relevant to their engineering unit. In looking for scales and items, we chose to address five of the most common topics surveyed in science attitude instruments as reported by recent literature reviews (Blalock et al., 2008; Tytler, 2014), as noted in the first column of Table 1. The second and third columns of Table 1 show the evolution of subscales over time, and will be explained further in subsequent sections.

From the SAI-II (Lichtenstein et al., 2008; Moore & Foy, 1997), we chose to use items from the "I want to be a scientist" scale, as identified by Lichtenstein et al. (2008), to measure students' interest in pursuing engineering and science careers. This scale consists of eight items (Cronbach's α=.810), some of which are expressed as negatives, such as, "Scientific work would be too hard for me." It also includes an item more aligned with attitude than aspirations, "I enjoy studying science."

Table 1. Subscales identified and named at each stage of analysis

After Literature Review	After Qualitative Analysis	After Item Reduction
Desire to Learn Science	Value of Engineering to Me	Value of Engineering to Me
Self-Efficacy in Science	Self-Efficacy in Engineering	(not retained)
Enjoyment of Science	Enjoyment of Engineering	Enjoyment of Engineering
Interest in Science Careers	Aspirations for Engineering	Aspirations for Engineering
Attitudes toward School Science	Attitudes toward School Engineering	Attitudes toward School Engineering
Value of Science in Society	Value of Engineering to Society	Value of Engineering to Society
(added)	Gender Bias	Gender Bias

We chose to test the 14 items addressing attitudes toward school science from the ATSSA (Germann, 1988). Items include "I would like to learn more about science," and "Science is fascinating and fun." From the mATSI, we pulled 17 items from three of five subscales as candidates for testing (Weinburgh & Steele, 2000); scales included "Value of Science in Society," "Self-Concept of Science," and "Desire to Do Science." From the STAQ (Owen et al., 2008), we chose to examine 14 items from the three subscales "Motivating Science Class," "Self-Directed Effort," and "Science is Fun for Me." Some items were redundant across scales, either exactly or similarly, but we used such items only once in the questionnaire, blending or choosing between similar questions. We also incorporated eleven of the items from the EEA (2010) addressing the value of science and engineering. Finally, we chose to develop five new questions to assess gender biases in engineering attitudes.

We chose to implement the survey as a post-only Likert-scale questionnaire, with a range of prompts from "Strongly Disagree" to "Strongly Agree" (see Figure 1). Students were asked to answer each question twice: once to the prompt "Last summer, I would have said:" and also to the prompt "Now I would say." We chose to implement the survey in this way knowing that the students in our study were likely to know little to nothing about engineering before engaging in the curriculum, and we had learned from prior experience with the EEA that children's responses regarding engineering before engineering instruction tended to be unreliable given their lack of a clear sense of what engineering is. We considered that this may be due to a "response shift," whereby students have a better sense of how to self-evaluate at the end of an intervention than they do prior to the intervention (the "retrospective pre"), as other researchers have found to be the case (e.g., Bhanji, Gottesman, de Grave, Steinert, & Winer, 2012; Sibthorp, Paisley, Gookin, & Ward, 2007). By asking about "before" and "now" after engineering instruction, we hoped to get more reliable data about students' change in attitudes by having them compare their current attitudes and interests to what they remembered of their prior interests and attitudes. Though we expect this retrospective will introduce some bias to "before" responses, we expect this will be outweighed by students' ability to give more informed responses.

We are interested in learning about your opinions of engineering. Please answer each question honestly. Mark how strongly you agree or disagree after each statement. Thank you very much!

		Strongly Disagree	Disagree Somewhat	Not Sure	Agree Somewhat	Strongly Agree
1. It is important for me to understand engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
2. Engineering helps me to understand today's world.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4

Figure 1. Image from the EIA assessment

Qualitative Evidence of Content Validity

Our first goal was to gather expert opinions on the content validity of the 64 items we had collected, plus ten more we developed to investigate gender stereotypes. We solicited opinions from four experts on science and engineering assessment and education within our institution, as well as from a former engineer, now an educator. We asked the experts to read the items, think about how they and their students might answer them, and comment on possible problems with content, readability, or wording of the items. We also asked engineers to comment on the original subscale naming. With the assistance of these experts, we flagged items that were possibly inappropriate or likely to be misinterpreted by our target age group, made adjustments to the scale names, and confirmed that assignments of items to subscales was considered appropriate. Thirty-three items from this list were dropped, generally because they were not appropriate for the age group, for example, items that referenced a "science course" or "science teacher," because American elementary school children are often taught all subjects by one or two teachers in their primary classroom. Seven items were modified to simplify vocabulary or sentence structure; for example, "Science is of great importance to a country's development" was modified to read "Science is of great importance to my country." Twenty-one items were added to the list, duplicating some items but referencing "engineering" instead of "science"; for example, "I enjoy studying engineering" was added to the list in parallel to the item "I enjoy studying science."

After expert review, we had a list of 62 items to test with students. To test the items for validity of response processes, we conducted cognitive interviews with 15 students in the target grade range (grades 3–5, ages 8–11), some from classrooms that had implemented in-school engineering curricula, and others from out-of-school time (OST) programs engaged in engineering units of exploration. During the interviews, we asked students to read each question aloud and talk about it. We asked them to explain any confusing aspects of the question, and to talk aloud about what they were thinking as they chose answers from the Likert scales. Based on the results of these interviews, we dropped one question that students had difficulty reading, "Engineering solutions to problems would be boring work." We revised eight questions to simplify the phrasing and re-tested them: for example, "No matter how hard I try, I cannot understand engineering" was changed to "Engineering is really hard to understand." We also had several classes of students in the target age range complete the questionnaire without interviews, and found that it was taking them much too long—more than 45 minutes. To shorten the questionnaire we decided to drop all 31 questions that referenced science rather than engineering, as data about science attitudes were less important to us than data about engineering interests and attitudes for our engineering curriculum study.

Once again, after our revisions based on testing of suitability with the target population, we asked our experts to review the resulting 30 candidate items and subscales for validity of content. After combining and dropping items, we finalized six candidate subscales (see Table 1). The second review resulted in few additional suggestions for revision—all minor edits of wording.

Data Collection

As part of the Exploring the Efficacy of Engineering is Elementary (E4) study, we collected post-surveys of students' interests in and attitudes toward engineering. Over two years, we collected surveys from almost 11,000 students in grades 3, 4, and 5. Students spanned a wide range of racial and SES demographic groups, from rural, urban, and suburban areas of several geographically non-contiguous American states. See Table 2 for the demographic breakdown of the sample.

Table 2. Student demographic breakdown of sample, reported as percentages

	Male	Minority[1]	FRL[2]	English Learners	Grade 3	Grade 4	Grade 5	Total N
Initial Sample								
Comparison	50.3	37.1	46.3	6.5	23.8	31.4	44.7	5,994
Treatment	51.4	31.7	43.8	5.8	32.3	36.7	31.0	4,912
Total	50.8	34.6	45.1	6.2	27.6	33.8	38.5	10,906
Final Sample after drops (due to incomplete surveys):								
Comparison	49.7	35.4	45.4	6.3	22.6	31.9	45.6	5,385
Treatment	51.4	30.2	42.6	5.5	31.9	36.8	31.3	4,417
Total	50.5	33.1	44.1	5.9	26.8	34.1	39.1	9,802

[1]Percentage of students from underrepresented minority groups (African American, Latino/a, Mixed-race, Other).
[2]Percentage of students receiving Free or Reduced-Price Lunch.

Item Reduction

Initially, we tested the instrument with a portion of the first year of data collected for the E4 study, returned in the first three months of the study. The purpose of the initial testing was to drop items not adding to the value of the questionnaire, to shorten it and reduce the burden on class time. The initial sample totaled 1,563 students from grade 3–5 classrooms that had implemented one or two engineering curricular units. Students completed the questionnaire independently as a written assessment. To gather evidence for the validity of the internal structure of the questionnaire, we examined the internal consistency reliability of items (Cronbach's α) and conducted a Principal Components Analysis (PCA) in SPSS version 22 (IBM Corporation, 2013) to determine which items contributed least to the total variance within the set of items (Dillon & Goldstein, 1984). Toward our goals, we dropped six items from the 30 tested that performed particularly poorly, failing to load onto a component, or detracting from internal consistency reliability. For example, the item "I do not want to be an engineer" was dropped because it had particularly low initial and extraction communalities (<.1); its removal increased Cronbach's α, and it failed to load onto any component. E4 subjects completing the EIA after item reduction analysis received the 24-question version.

Exploratory Factor Analysis

Before final analysis of the instrument, we randomly split our sample in half to conduct an EFA and CFA on separate samples. The purpose of EFA is to describe or explore the relationships between items that are interrelated, to describe common factors (groupings of items) that are expected to correspond to theorized latent (unobserved) variables. EFA was conducted for this study because the items used in the instrument had not been previously analyzed together for the purpose of ensuring that an interpretable factor structure was possible, and CFA should not be run until the structure has been studied using EFA with a separate, independent sample (Bandalos & Finney, 2010).

With the first subsample, we used Parallel Analysis (PA), a method of comparing the eigenvalues of a specific sample with estimated population eigenvalues, to determine what number of factors was likely to be significant. To conduct PA, we used a script from https://people.ok.ubc.ca/brioconn/nfactors/nfactors.html (O'Connor, 2000) in SPSS 24 to assist in estimating the number of factors, backed up by examination of the scree plot and eigenvalues, and comparison to the intended subscales; Bandalos and Finney (2010) recommend the use of multiple methods and the criterion of theoretical plausibility to determine the number of factors, with preference given to a choice for which multiple methods and theoretical plausibility converge.

EFA was conducted in Mplus 7.4 (Muthén & Muthén, 2015). We used the MLR estimator, an extension of Maximum Likelihood (ML) estimation that is robust to multivariate non-normality, and adjusts for missing data using Full Information Maximum Likelihood (FIML), to handle the non-normality of our ordinal 5-point Likert-scale data. The ratio of sample size (5,390) to expected factors (<10) is quite high (539:1) so we expect that the sample size is sufficient for this procedure, even when extracted communalities are low (MacCallum, Widaman, Zhang, & Hong, 1999). We used an oblique rotation (Geomin) with the EFA because we expected the resulting factors to be correlated. We examined the structure matrices for correlations between items and factors, and the pattern matrices for item loadings and cross-loadings, using base thresholds of structure coefficient >.450 and pattern coefficient >.300 for considering an item as loading onto a factor; because our factor correlations are strong and sample size is large, structure coefficients generally are expected to be larger than pattern coefficients (Bandalos & Finney, 2010; Brown, 2006).

We also considered goodness-of-fit information that is available with an ML-based EFA. Three kinds of fitness measures are available for testing models: measures of absolute fit, comparative fit, and parsimonious fit (Kelloway, 2015). Measures of absolute fit test how closely the covariance matrix for the model matches the covariance matrix for the input (baseline) data. Measures of comparative fit give information about which of two competing models better matches the covariance matrix for the input data. Measures of parsimonious fit are a type of comparative measure that adjust negatively for the loss of degrees of freedom due to specifying more parameters for a model—because, all else being equal, the specification of more parameters will always lead to a better fit to the covariance matrix. We used the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) as measures of absolute fit, with cutoffs of <.05 for the former and <.80 for the latter; we used the comparative fit index (CFI) as a measure of comparative fit, with a cutoff of >.95 indicating good fit; and we used the Akaike information criterion (AIC) as a measure of parsimonious fit to compare models, with a smaller value indicating a superior model (Kelloway, 2015). We also report the χ^2 statistic, which can be used as a measure of absolute fit, with a difference between the fitted model and baseline model of $p<.05$ traditionally indicating good fit. Our purpose was to explore possible factor structures and compare the fit of a variety of candidate factor solutions, so we could ensure that a structure could be specified where the correlations of items with factors was interpretable and reasonably congruent to theorized latent dimensions before embarking on CFA.

Confirmatory Factor Analysis

Our purpose in conducting CFA was to cross-validate the factor structure developed by theory and refined with EFA (Brown, 2006). For CFA, in contrast to EFA, all indicators (survey items) and their relationships to latent variables (factors) must be specified in advance, to "confirm" the validity of the theorized model. Kelloway (2015) recommends that, because CFA is strongest for comparing models, the best approach is to identify ambiguous aspects of the model to be tested, and to specify nested models that remove ambiguous aspects of the full model to test their contribution to the model. The theorized relationships between latent variables can be tested by specifying nested models that contain a subset of the parent model's parameters. The parent and nested models can then be compared to determine which model specification is the best fit for the data (Brown, 2006). Therefore, before beginning CFA analysis, we generated a nested, competing model from a single, fully specified parent model containing the relationships between all latent variables freely specified, as well as the full set of relationships between observed indicators and latent variables that resulted from EFA. The nested model set all cross-loading parameters to zero, effectively removing them from the model. We used CFA to compare the fit of the parent and nested models.

Using the second subsample that we generated before EFA, we conducted CFA using Mplus 7.4. Sample data was input to Mplus, which generated variance-covariance matrices for analysis. All models used MLR as estimator. We used RMSEA and SRMR as measures of absolute fit, CFI as a measure of comparative fit, and AIC as a measure of parsimonious fit to compare models. We also report the χ^2 statistic; however, we compared our nested models using an adjusted scaled difference χ^2 test statistic (Satorra & Bentler, 2010), which is necessary because the simple difference between two scaled χ^2 statistics from MLR does not have a χ^2 distribution.

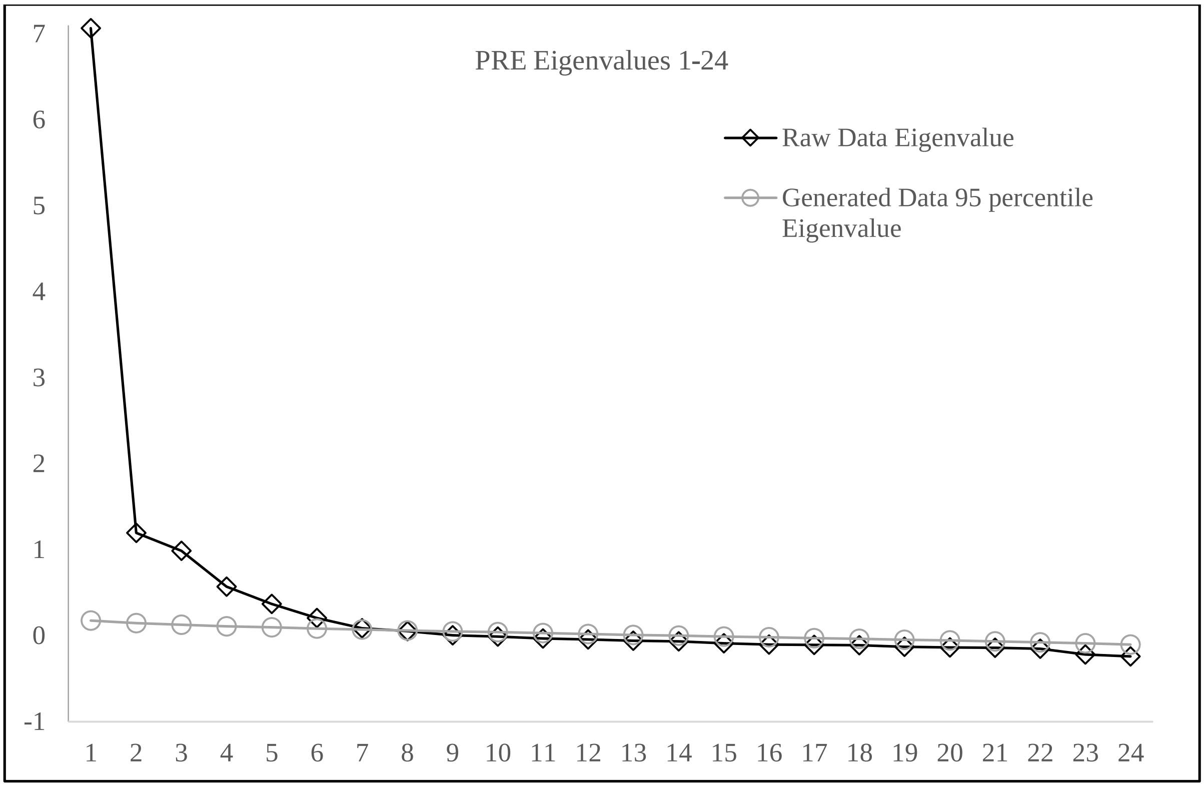
To determine the quality of the model, we examined the parameter estimates for significance and interpretability (Brown, 2006). Mplus provides, in addition to the unstandardized and standardized parameter estimates, the standard error, z-statistic, and z-test *p*-value for each parameter for each parameter estimate, which we inspected and reported. Non-significant parameters should be considered for removal from the model. Standard errors were inspected for excessively large values, which would indicate an unreliable parameter estimate. For items that are not cross-loading, the completely standardized factor loading represents the correlation between the item and factor, and the R^2 statistic represents the communality, that is the proportion of variance of the item that is explained by the factor. We examined the size of factor loadings and R^2 values as further evidence for whether item-factor relationships are strong enough to be meaningful. Finally, we examined the factor determinacy of factor scores as a measure of factor score quality, with a threshold of >.8 for a good-quality factor score and >.9 as preferred (Grice, 2001). Mplus provides factor score determinacies, which are a measure of the correlation between generated factor scores and the latent factor estimate (available on request by email with the first author).

Results and Discussion

Exploratory Factor Analysis and Subscale Development

EFA was conducted concurrently on each set of items (PRE and NOW), to ensure a factor structure that fit both the PRE, “Last summer, I would have said,” and NOW, “Now I would say,” responses. We expected to see differences between the PRE and NOW sets of items, because we knew that students were likely not to have had prior engineering experiences so were likely to report weaker opinions (or possibly stronger in the case of Gender bias) on the PRE questions. However, to facilitate PRE-NOW comparisons, we would need one consistent model. Therefore, throughout the EFA process, we worked to find the best-fitting factor structures for both the PRE and NOW sets of items that also made the most sense theoretically. Where there were differences in the pattern of coefficients for each set, a compromise was made, and we chose the thematically most sensible placement. This led to the choice of factor structure that may not have been the statistically best fit for either set of items. However, the final structure chosen was a good fit for each set of items, and made sense given the theoretical framework.

To begin, we ran PA with a 99% probability cutoff on the random half-sample 1 prior to EFA. To determine the number of factors indicated for analysis, we compared the sample data eigenvalues to the randomly generated data percentile eigenvalues (Table 3). A factor where the sample data eigenvalue exceeds the random data percentile eigenvalue is retained. Analysis indicates that seven factors can be extracted for the PRE variables, and eight for the NOW items. With the traditional cutoff of eigenvalues >1, however, only 2 factors are indicated. Examination of the scree plots with the PA simulated data superimposed as a gently sloped line (Figure 2) shows a sharp drop in sample data eigenvalues after one factor, with a softer elbow curving down to near horizontal including another five factors. To examine a range of possible factor structures as indicated by PA (7-8 factors), our theorized factor structure (6 factors), the scree plots (6 factors), and the much smaller number of eigenvalues>1 (2 factors), we decided to conduct EFA to fit 4 to 8 factors. We chose not to examine 2 or 3 factors because this was so much less than our theorized 6 factors and the results of other methods for estimating the number of factors.



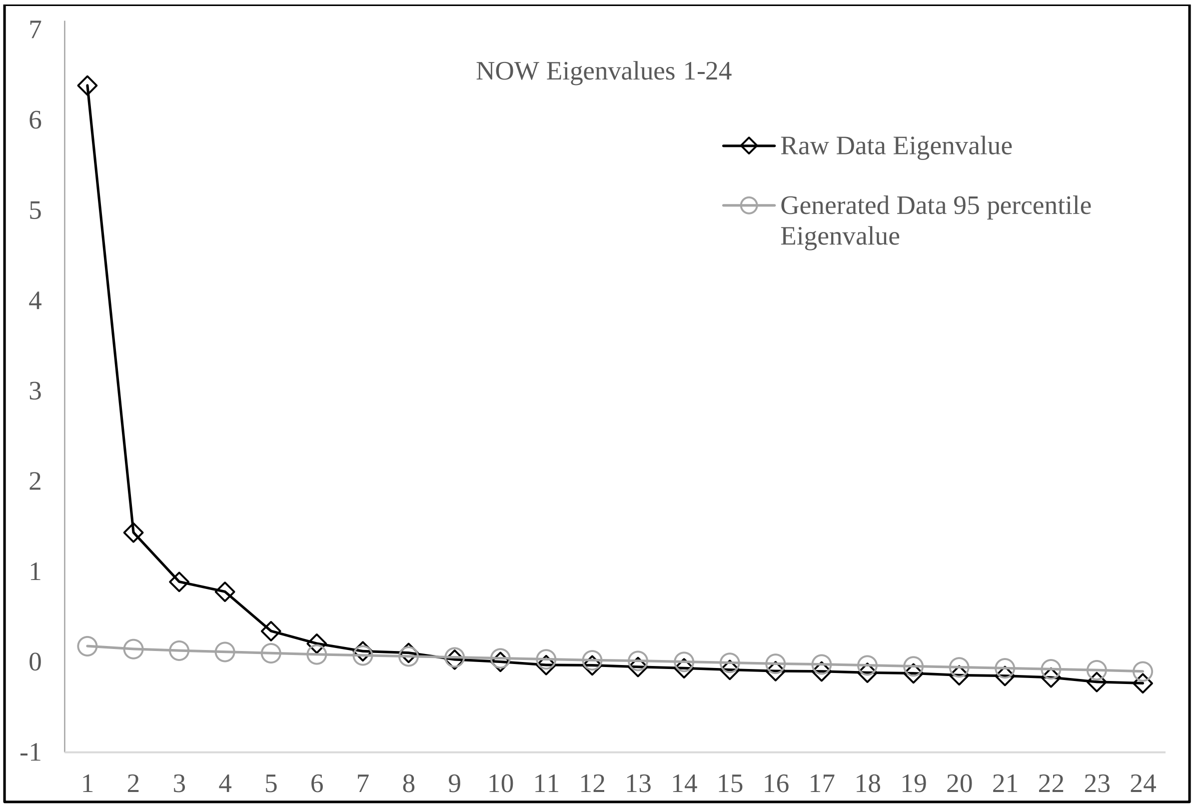


Figure 2. Scree Plots of "Pre" (top) and "Now" factors extracted, Random Half 1

Examination of goodness-of-fit estimates showed that structures with more factors tended to fit the data better, especially as compared to 4 or 5 factors, although the difference between the best-fitting 7- and 8-factor structures was very small (see Table 4). Table 5 shows the structure and pattern matrices for PRE/NOW, with coefficients for the chosen scales in **bold**, and cross-loading items shown in *italics* for the secondary loading. We were unable to calculate χ^2 or other fit coefficients for 5 factors for the PRE data, but as other factor structures fit the data better and had better fit statistics than the 5-factor NOW structure, we decided not to explore this further. The factor structure that best mirrored our theorized subscales, which we had tested qualitatively and examined using PCA (Table 1), included 7 factors. All 6 factors we had theorized were identifiable in the analysis for NOW data, and for the PRE data 5 of 6 were present (the seventh factor had no significant loadings). We expect that students were expressing weaker opinions about items loading on the two factors "Enjoyment" and "Aspirations" before they participated in engineering in school. Another difference between the factor structure and our theorized structure is that the Gender bias factors were, in each case, split across two factors: Male bias (items asking whether boys were better/girls had a harder time with engineering) and Female bias (with girls and boys swapped in the items, but using the same phrasing). The split of the Gender bias factor accounts for the remaining factor in each case.

Table 3. Parallel Analysis for PRE and NOW items

	"PRE" Eigenvalues			"NOW" Eigenvalues		
Factor	Sample Data	Random Data Mean	Random Data Percentile	Sample Data	Random Data Mean	Random Data Percentile
1	7.067699	.142976	.175269	6.384192	.142844	.174591
2	1.196063	.122523	.145149	1.433322	.122370	.143114
3	.986179	.106976	.126354	.886926	.106737	.125686
4	.571639	.093268	.110253	.778632	.092936	.110052
5	.369565	.081129	.097095	.341110	.080612	.096031
6	.204882	.069291	.084135	.203052	.069121	.083231
7	.084881	.058362	.072131	.117400	.058076	.073497
8	.053478	.047825	.061380	.100461	.047626	.062739
9	.005873	.037199	.050885	.026821	.037760	.051062

Table 4. Fit indices for EFA models, PRE and NOW

# Factors	# Parameters	χ^2	*df*	AIC	CFI	RMSEA	SRMR
4 PRE	138	1383.8	186	403257	0.96	0.034	0.02
4 NOW	138	1294.2	186	364866	0.96	0.034	0.02
5 PRE	Could not be computed.						
5 NOW	158	821.6	166	364113	0.98	0.027	0.02
6 PRE	177	634.5	147	402353	0.99	0.025	0.01
6 NOW	177	580.6	147	363774	0.99	0.023	0.01
7 PRE	195	395.6	129	402112	0.99	0.019	0.01
7 NOW	195	366.5	129	363512	0.99	0.018	0.01
8 PRE	212	364.8	112	402026	0.99	0.020	0.01
8 NOW	212	281.1	112	363402	0.99	0.017	0.01

All χ^2 tests of model fit were significant (p<.0001).

Both the structure and pattern matrices were interpreted in making decisions for subscale loadings. Items that loaded on either the PRE or NOW structure matrix with a coefficient >.450 were considered, as were items that loaded on either the PRE or NOW pattern matrix with a coefficient >.300. In comparing the content of items with the strongest coefficients, it became clear that the PRE factor corresponding to the "Enjoyment of engineering" subscale also included the "Aspirations" items, while the NOW coefficients for those items were loaded across two factors, one "Enjoyment" and the other "Aspirations." For this reason, PRE coefficients in the Table 5 "Enjoyment" column are bolded or italicized as half pairs when they are intended to be paired with the NOW coefficients in the "Aspirations" column in the final, compromise model.

All item-factor loadings with at least three out of four (PRE, NOW, pattern, and structure) coefficients larger than the thresholds were chosen for subscales. Only two item-factor loadings with fewer than three above-threshold coefficients were chosen as secondary, cross-loading items on factors: Item 5 "I would like to work with other engineers to solve engineering problems" (which loaded on the "Aspirations" and "Value to me" subscales) and Item 10 "It is important to understand engineering in order to get a good job" (which loaded primarily on the "Value to society" subscale and secondarily on the "Aspirations" subscale). Each of these two item-factor loadings was chosen because of the content of the item, which sensibly cross-loaded, and because some of the below-threshold coefficients were similar in value to the above-threshold coefficients for the same item's primary factor loading.

Table 5. Structure and pattern matrices for PRE/NOW

Item # -Type	Value to me	Enjoyment	Value to society	Male bias	Aspirations (NOW only)	Female bias	School
1-S	**.506/.552**	.470/.345	.454/.418				
1-P	**.378/.404**	.105/.018	.261/.156				
2-S	**.493/.521**		*.465/.461*				
2-P	**.408/.400**		*.385/.298*				
3-S	*.563/.576*	**.658/.675**			.603		.495/.353
3-P	*.332/.359*	**.502/.497**			.077		.016/-.055
4-S				**.758/.808**			
4-P				**.759/.826**			
5-S	*.434/.501*	**.503**/.465			**.559**		
5-P	*.257/.293*	**.280**/.079			**.351**		
6-S	*.534/.557*	.527/.528			.482		**.626/.618**
6-P	*.334/.311*	.012/.228			-.006		**.476/.394**
8-S	.484/.496	**.710/.794**	.481/.358		.661		.550/.439
8-P	.200/.183	**.509/.666**	.010/.000		.050		.025/.042
9-S		.498/.443	.478/.360				**.570/.489**
9-P		.023/.252	.095/.077				**.352/.298**
10-S		*.463*/.261	**.521/.372**		*.374*		.474/.283
10-P		*.091*/-.074	**.335/.239**		*.238*		.112/.073
13-S		**.673/.734**	.522/.394		.650		.634/.491
13-P		**.383/.510**	.009/.042		.173		.287/.154
14-S			**.692/.650**				.478/.310
14-P			**.682/.669**				-.024/-.065
15-S				**-.450/-.399**			
15-P				**-.443/-.359**			
17-S		.496/.340	**.567/.486**				.479/.343
17-P		.160/.223	**.397/.403**				.044/.054
18-S		**.659**/.486	.490/.287		**.665**		
18-P		**.645**/.007	.179/.068		**.688**		
19-S						**.640/.768**	
19-P						**.641/.753**	
20-S						**.929/.827**	
20-P						**.929/.843**	
21-S		.461/.283	**.660/.559**				.555/.436
21-P		.029/-.032	**.557/.430**				.180/.173
22-S		.557/.431	.594/.496		.456		**.723/.687**
22-P		.029/.008	.178/.135		.148		**.585/.538**
23-S		**.478/.373**					
23-P		**.362/.281**					
24-S			**.617/.569**				.487/.342
24-P			**.530/.521**				.111/.042
25-S		.530/.430	*.527/.448*				**.598/.469**
25-P		.198/.167	*.145/.220*				**.381/.225**
26-S		**.765**/.688	.516/.371		**.806**		.607/.469
26-P		**.747**/.152	-.005/-.014		**.683**		.134/.088
28-S				**.734/.757**			
28-P				**.736/.750**			
30-S		**.734**/.657	.461/.346		**.805**		.530/.378
30-P		**.779**/.102	.005/-.031		**.700**		.006/.024

Bold/*italics* marks items in subscales. Coefficients were omitted when none of the set met minimum thresholds.

Because the intended "Gender bias" subscale split across two factors, we decided to drop the "Female bias" factor, as only two items loaded on it, while the "Male bias" factor also captured Item 15, and the wording of its items more strongly corresponded to traditional gender stereotypes for engineering. Table 6 lists the final choices and full text of items with their primary loadings onto subscales, and, where applicable, a secondary loading.

Table 6. Items in final subscales

Item #	Subscale	Cross-Loading	Text of Item from the EIA Questionnaire
8	Enjoyment		Engineering is fun
13	Enjoyment		I am interested when we do engineering in school
23	Enjoyment		Engineering is easy for me
3	Enjoyment	Value to me	I enjoy studying engineering
1	Value to me		It is important for me to understand engineering
2	Value to me	Value to society	Engineering helps me understand today's world
6	School	Value to me	We learn about interesting things when we do engineering in school
9	School		When we do engineering, we use a lot of interesting materials & tools
22	School		We learn about important things when we do engineering in school
25	School	Value to society	I try hard to do well in engineering
14	Value to society		Engineers help make people's lives better
17	Value to society		I know what engineers do for their jobs
21	Value to society		Engineering is useful in helping to solve the problems of everyday life
24	Value to society		Engineering is really important to my country
10	Value to society	Aspirations	It is important to understand engineering in order to get a good job
30	Aspirations		I really want to learn engineering
18	Aspirations		I would enjoy being an engineer when I grow up
26	Aspirations		I would like to learn more about engineering
5	Aspirations	Value to me	I would like to work with other engineers to solve engineering problems
4	Gender bias		Boys are better at engineering than girls
28	Gender bias		Girls have a harder time understanding engineering than boys
15	Gender bias		Girls and boys are equally good at engineering
19	Dropped		Boys have a harder time understanding engineering than girls
20	Dropped		Girls are better at engineering than boys

To assess the relations between factors, we examined the correlation matrices for the PRE and NOW 7-factor extractions (Table 7). None of the factors were excessively correlated, which would be an indication that factors should be combined. Only "Male bias" and "Female bias" had low enough correlations with other factors to be statistically insignificant. "Value to me" was the only factor correlated with the two gender bias subscales on the PRE, a relationship that may require analysis by gender to understand. On the other hand, most of the other variables (except "Aspirations") were significantly and negatively (though only mildly) correlated with the NOW gender bias subscales, indicating that less bias was associated with more positive attitudes of enjoyment of engineering in general and at school, and more positive assessment of the value of engineering to society.

Table 7. Correlations between factors extracted by EFA with MLR estimators PRE/NOW

	Value to me	Enjoyment	Value to society	School	Aspirations	Male bias
Value to me	1.000					
Enjoyment	**.451/.411**	1.000				
Value to society	**.259/.384**	**.643/.373**	1.000			
School	**.382/.411**	**.724/.461**	**.706/.527**	1.000		
Aspirations	X/**.497**	X/**.757**	X/**.431**	X/**.425**	1.000	
Male bias	**.049**/-.047	.000/**-.109**	-.029/**-.138**	**-.051/-.200**	X/-.021	1.000
Female bias	**.061**/-.034	.030/**-.051**	-.021/**-.117**	-.005/**.199**	X/.022	.004/**-.109**

X: No factor corresponding to "Aspirations" was found in the PRE data. **Bold** correlations are significant (p<.05).

Confirmatory Factor Analysis

As with EFA, CFA was conducted on the PRE and NOW datasets in parallel, to ensure that the final model would fit well for both datasets. Using Mplus 7.4, we tested the factor structure specified in Table 6 with the second random half of the full dataset, to cross-validate the factor structure with new data. All CFA models were estimated with MLR, which allows for the estimation of missing values using FIML.

Table 8. Model Information

	Degrees of Freedom	# Free Parameters	#Observations PRE	#Observations NOW	# Missing data patterns PRE	# Missing data patterns NOW
Model 1	188	87	5495	5507	310	294
Model 2	194	81	5495	5507	310	294

We chose to examine and compare two nested models. The models are illustrated in Figure 3, model information is given in Table 8, and model specifications are presented in Table 9. The primary difference between Model 1 and Model 2 is the specification of cross-loading terms. Model 1 includes all of the cross-loading item-factor relationships specified in Table 6, and has 188 degrees of freedom. Model 2 contains no cross-loading items, and therefore has more degrees of freedom: 194. For each latent variable, we chose a marker indicator, an item that had a high pattern coefficient on the corresponding EFA factor and low cross-loading pattern coefficients (see Table 5). Our initial choice of marker indicators was successful for all but one latent variable: Value to me (initial choice EIA_1). In this case, examination of the Modification Indices for the model showed it would be substantially improved by freeing EIA_1 and substituting EIA_2 as the marker indicator.

Table 9. Model specifications

	Model 1 Indicators		Model 2 Indicators	
	Marker (Fixed)	Freely estimated	Marker (Fixed)	Freely estimated
Enjoyment	EIA_8	EIA_3, 13, 23	EIA_8	EIA_3, 13, 23
Value to me	EIA_2	EIA_1, 3, 5, 6	EIA_2	EIA_1
School	EIA_22	EIA_6, 9, 25	EIA_22	EIA_6, 9, 25
Value to society	EIA_14	EIA_2, 10, 17, 21, 24, 25	EIA_14	EIA_10, 17, 21, 24
Aspirations	EIA_30	EIA_5, 10, 18, 26	EIA_30	EIA_5, 18, 26
Gender bias	EIA_4	EIA_4, 15, 28	Variance@1	EIA_4, 15, 28

Overall goodness-of-fit was very good for both models (Table 10). All fit indices met threshold requirements, except the CFI measure of parsimonious fit: values for the PRE models slightly missed the threshold (0.94 < 0.95). CFI values for Model 1 NOW were better than those for Model 2. RMSEA 95% confidence intervals were under the 0.05 threshold.

The AIC measure of comparative fit was smaller for the Model 1 PRE and NOW than for the corresponding values for Model 2, indicating that Model 1 is the better fit to the data. To further compare the models for goodness-of-fit, we calculated the Satorra-Bentler scaled χ^2 difference test (TRd), to compensate for the MLR χ^2 having a different distribution than the standard χ^2 distribution. The value for the PRE models is TRd=243, and the value for the NOW models is TRd=291; positive values indicate that the model with fewer degrees of freedom (Model 1) is the better model. As all indications are that Model 1, with cross-loading indicators, is superior to Model 2 without cross-loading, we proceed to specify Model 1 in the remainder of this paper.

Table 10. Fit indices for nested models, PRE and NOW

Model	χ^2	*df*	AIC	CFI	RMSEA	RMSEA 95% CI	SRMR
1 PRE	1953.6	188	367792	0.94	0.041	0.040 - 0.043	0.060
1 NOW	1294.2	194	328411	0.96	0.033	0.031 - 0.034	0.046
2 PRE	2169.2	188	368076	0.94	0.043	0.041 - 0.045	0.061
2 NOW	1539.5	194	328777	0.95	0.035	0.034 - 0.037	0.049

All χ^2 tests of model fit were significant (p<.0001).

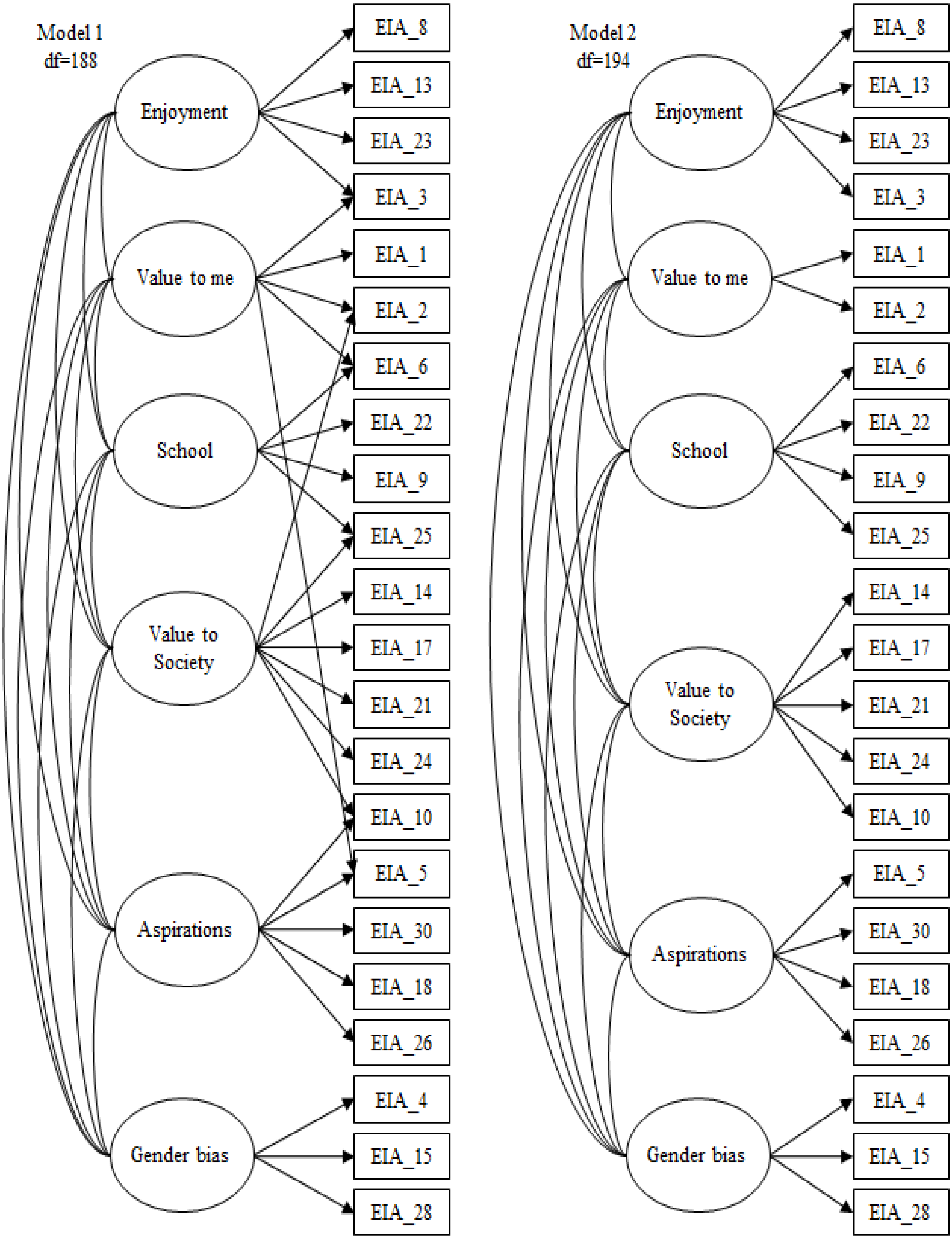


Figure 3. Two nested models to compare

Table 11. Standardized parameter estimates for Model 1: PRE / NOW.

	Parameter			Residual Variance			
Name	Estimate	S.E.	P-Value	Estimate	S.E.	P-Value	R-Square
Enjoyment BY							
EIA_8	0.740/0.804	0.009/0.009	0.000/0.000	0.452/0.354	0.013/0.014	0.000/0.000	0.548/0.646
EIA_3	0.504/0.612	0.043/0.027	0.000/0.000	0.528/0.484	0.012/0.015	0.000/0.000	0.472/0.516
EIA_13	0.716/0.767	0.009/0.010	0.000/0.000	0.488/0.412	0.013/0.015	0.000/0.000	0.512/0.588
EIA_23	0.492/0.375	0.013/0.016	0.000/0.000	0.758/0.859	0.013/0.012	0.000/0.000	0.242/0.141
Value to me BY							
EIA_2	0.447/0.361	0.030/0.035	0.000/0.000	0.676/0.684	0.016/0.017	0.000/0.000	0.324/0.316
EIA_1	0.654/0.715	0.018/0.023	0.000/0.000	0.572/0.489	0.023/0.032	0.000/0.000	0.428/0.511
EIA_3	0.215/0.155	0.046/0.033	0.000/0.000	0.528/0.484	0.012/0.015	0.000/0.000	0.472/0.516
EIA_5	0.341/0.211	0.040/0.029	0.000/0.000	0.682/0.643	0.014/0.014	0.000/0.000	0.318/0.357
EIA_6	0.367/0.150	0.056/0.055	0.000/0.006	0.559/0.504	0.015/0.018	0.000/0.000	0.441/0.496
School BY							
EIA_22	0.701/0.666	0.010/0.014	0.000/0.000	0.509/0.557	0.015/0.018	0.000/0.000	0.491/0.443
EIA_6	0.341/0.588	0.055/0.047	0.000/0.000	0.559/0.504	0.015/0.018	0.000/0.000	0.441/0.496
EIA_9	0.620/0.569	0.011/0.015	0.000/0.000	0.616/0.676	0.014/0.017	0.000/0.000	0.384/0.324
EIA_25	0.501/0.423	0.071/0.042	0.000/0.000	0.628/0.688	0.014/0.017	0.000/0.000	0.372/0.312
Value to society BY							
EIA_14	0.655/0.604	0.011/0.015	0.000/0.000	0.571/0.636	0.015/0.018	0.000/0.000	0.429/0.364
EIA_10	0.423/0.262	0.029/0.022	0.000/0.000	0.683/0.814	0.013/0.012	0.000/0.000	0.317/0.186
EIA_17	0.608/0.511	0.011/0.017	0.000/0.000	0.631/0.739	0.014/0.017	0.000/0.000	0.369/0.261
EIA_21	0.671/0.600	0.011/0.015	0.000/0.000	0.550/0.640	0.015/0.018	0.000/0.000	0.450/0.360
EIA_24	0.606/0.571	0.012/0.015	0.000/0.000	0.633/0.674	0.015/0.017	0.000/0.000	0.367/0.326
EIA_2	0.164/0.263	0.027/0.033	0.000/0.000	0.676/0.684	0.016/0.017	0.000/0.000	0.324/0.316
EIA_25	0.120/0.165	0.074/0.046	0.106/0.000	0.628/0.688	0.014/0.017	0.000/0.000	0.318/0.357
Aspirations BY							
EIA_30	0.728/0.813	0.009/0.009	0.000/0.000	0.471/0.339	0.014/0.014	0.000/0.000	0.529/0.661
EIA_5	0.269/0.452	0.040/0.026	0.000/0.000	0.682/0.643	0.014/0.014	0.000/0.000	0.318/0.357
EIA_10	0.175/0.230	0.029/0.021	0.000/0.000	0.683/0.814	0.013/0.012	0.000/0.000	0.317/0.186
EIA_18	0.665/0.632	0.011/0.011	0.000/0.000	0.557/0.601	0.014/0.014	0.000/0.000	0.443/0.399
EIA_26	0.768/0.818	0.009/0.008	0.000/0.000	0.411/0.331	0.013/0.014	0.000/0.000	0.589/0.669
Gender bias BY							
EIA_4	0.769/0.780	0.017/0.014	0.000/0.000	0.409/0.391	0.026/0.023	0.000/0.000	0.591/0.609
EIA_15	-0.424/-0.408	0.018/0.017	0.000/0.000	0.820/0.834	0.015/0.014	0.000/0.000	0.180/0.166
EIA_28	0.726/0.775	0.017/0.015	0.000/0.000	0.473/0.399	0.025/0.023	0.000/0.000	0.527/0.601

Standardized parameter estimates are listed in Table 11. All parameters are statistically significant (p<.001) except for Value to me by EIA_6 - NOW (p<.01) and Value to society by EIA_25 - PRE (p=.106). The lack of significance of the Value to society by EIA_25 parameter for the PRE data also corresponds with an unusually high standard error for Model 1 (S.E.=0.074). EIA_25 also has a high standard error for its primary loading on the PRE, with the latent variable School (S.E.=.074). The content of EIA_25 is "I try hard to do well in engineering" and it may make sense that children who have just completed a questionnaire about their first experience with engineering might provide unreliable answers to this question. However, we decided not to drop this indicator because EIA_25 still fits well with the NOW data, with a highly significant estimate (p<.001) and much smaller standard error (S.E.=.046), much more in line with the range of standard errors for the rest of the parameter estimates.

For the most part, Model 1 parameters explained substantial item variance: R^2 for the non-cross-loading terms ranged between 0.141 (EIA_23 NOW) and 0.669 (EIA_26 NOW). Most R^2 values ranged between .3 and .5, which corresponds to approximately 30% to 50% of each observed indicator's variance explained by Model 1. Disattenuated correlations between the factors are presented in Table 12. All factors except Gender bias had statistically significant positive correlations with each other (p<.001), indicating that positive interest and attitudes tend to go hand-in-hand. The Gender bias correlations are theoretically interesting in the pattern of changes from PRE to NOW, which indicates that before participating in engineering student gender bias was unrelated to their attitudes and interest, but after participating in engineering students who expressed more positive attitudes also tended to indicate that they feel less biased about gender. Gender bias is measured in the negative, as indicated by the signs of the associated indicators: Items 4 and 28, for which larger values indicate more gender bias (see Table 6), are positively associated with the Gender bias latent factor, while Item 15, “Girls and boys are equally good at engineering,” is negatively associated with it. Therefore a negative correlation of Gender bias with the attitude and interest latent factors indicates that less gender bias is associated with more positive attitudes and interest.

Table 12. Disattenuated correlations between factors Model 1 PRE/NOW

Factor	Enjoyment	Value to me	School	Value to society	Aspirations
Enjoyment	1.00				
Value to me	.791**/.621**	1.00			
School	.899**/.840**	.754**/.720**	1.00		
Value to society	.768**/.584**	.668**/.612**	.887**/.762**	1.00	
Aspirations	.932**/.891**	.702**/.566**	.819**/.734**	.732**/.539**	1.00
Gender bias	.009 /-.105**	-.013 /-.148**	-.060* /-.208**	-.041 /-.180**	-.021 /-.046[t]

[t]*p*<.05; * *p*<.01; ***p*<.001

Table 13 displays the factor determinacies for the refined factor scores derived by Mplus from Model 1 with the random half 2 sample data. All factors exceed the threshold of 0.8, indicating good quality and replicability. More than half of the factor determinacies exceed the preferred threshold of 0.9.

Table 13. Model 1 factor determinacies

	Enjoyment	Value to me	School	Value to society	Aspirations	Gender bias
PRE	0.934	0.865	0.918	0.908	0.908	0.858
NOW	0.939	0.839	0.917	0.873	0.936	0.876

Conclusion

Intentions about careers can be shaped as early as elementary school, when few children have a clear sense of, never mind enthusiasm toward, engineering. Strategies for addressing shortages in the STEM pipeline that target high school or even middle school students may come too late, particularly to tap into the interests of presently underrepresented groups, including women, African Americans, and Latinos/as. Therefore many researchers, policymakers, funders, and educators are working to bring engineering curricula, environments, clubs, and activities to elementary school children, both in and out of school, to address the pressing need for more young people interested in and prepared to pursue further education for STEM careers.

Given high interest in addressing the existing STEM pipeline shortages through interventions with younger children in and out of school intended to positively affect interests and attitudes, an instrument is needed to measure the impact of such programs. This study shows that the EIA questionnaire has strong evidence of content and structural validity. The instrument can be used with students ages 8–11 to measure changes in student enjoyment of engineering, desire to learn engineering, interest in school engineering, aspirations to become an engineer, and attitudes toward the value of engineering to society. It can also be used to measure self-reported changes in the level of student gender bias regarding participation in engineering, and the relationship of gender bias to other engineering attitudes and interests. We expect that researchers and curriculum developers will want to use this instrument to measure changes in student interests and attitudes after participation in engineering activities, programs, and curricula.

Recommendations

The sample for this study was diverse with regard to race, ethnicity, and socioeconomic status as well as geography within the United States, and the EIA instrument likely will perform similarly with like populations. The sample of English learners was about 6%, so use of the measure in samples with a higher proportion of English learners may differ. Translations for bilingual classrooms may be necessary to preserve the characteristics of the questionnaire. No translations of the questionnaire have been tested and the properties of the instrument for use outside the United States have not been investigated.

These results are specific to school-aged children enrolled in grades 3–5, although many items were adapted from questionnaires for older children. Use with younger children in a written format might be challenging. Use of the instrument with older students might be possible, but the factor structure, reliability, and validity exercises should be revisited.

The largest concern in the use of this assessment is the one experienced in the process of development and reported by others, which is that among naive elementary-aged children, there is insufficient knowledge of the concept of engineering to respond meaningfully to questions that use the term. In our circumstances, this was resolved by exposing the students to an engineering curriculum and administering the questionnaire afterward. To assess a "pre" time point, students were asked to reflect back to a specific time before their exposure to the engineering curriculum. An effective method to assess constructs such as desire to learn engineering, perceived value of engineering, and gendered attitudes around engineering in elementary-/middle-school-aged children who have not been exposed to an engineering curriculum is yet to be demonstrated.

Our intention is to use the refined factor scores as outcome variables in future work where we explore the impact of engineering curricula on student interest in, aspirations for, and attitudes toward engineering. Others may similarly use the instrument with children ages 8-11 to gather data on changes in children's interests and attitudes following engineering interventions in formal and informal settings. We recommend, when evaluating the attitudes and interest of students who are new to engineering, that the retrospective post version of the EIA be administered after completion of an engineering intervention. However, in cases where students already have enough engineering experience that they can reasonably interpret the questions, the instrument could be adapted and used as a pre-post survey. In either case, "NOW" subscales should be used as outcome variables, with the corresponding "PRE" subscales as covariates.

Notes

This material is based upon work supported by the National Science Foundation under Grant No. [1220305]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology*, 36(1), 4–12. https://doi.org/10.1016/j.cedpsych.2010.08.001

Appelbaum, P., & Clark, S. (2001). Science! Fun? A critical analysis of design/content/evaluation. *Journal of Curriculum Studies*, 33(5), 583–600.

Cunningham, C. M., & Lachapelle, C. P. (2010, June). *The impact of Engineering is Elementary (EiE) on students' attitudes toward engineering and science*. Paper presented at the American Society for Engineering Education Annual Conference & Exposition, Louisville, KY. Retrieved from https://peer.asee.org/15989

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences*. New York, NY: Routledge.

Bhanji, F., Gottesman, R., de Grave, W., Steinert, Y., & Winer, L. R. (2012). The retrospective pre–post: A practical method to evaluate learning from an educational program. *Academic Emergency Medicine*, 19(2), 189–194.

Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., & Toepperwein, M. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education*, 30(7), 961–977.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Buccheri, G., Gurber, N. A., & Bruhwiler, C. (2011). The impact of gender on interest in science topics and the choice of scientific and technical vocations. *International Journal of Science Education*, 33(1), 159–178. https://doi.org/10.1080/09500693.2010.518643

Burke, R. J. (2007). Women and minorities in STEM: A primer. In R. J. Burke & M. C. Mattis (Eds.), *Women and minorities in science, technology, engineering and mathematics: Upping the numbers* (pp. 3–27). Northampton, MA: Edward Elgar Publishing, Inc. https://doi.org/10.4337/9781847206879.00008

Dillon, W. R. & Goldstein, M. (1984). *Multivariate analysis: methods and applications*. New York, NY: Wiley.

Drechsel, B., Carstensen, C., & Prenzel, M. (2011). The role of content and context in PISA interest scales: A study of the embedded interest items in the PISA 2006 science assessment. *International Journal of Science Education*, 33(1), 73–95. https://doi.org/10.1080/09500693.2010.518646

Duit, R., & Treagust, D. F. (1998). Learning in science: From behaviourism towards social constructivism and beyond. *International Handbook of Science Education*, 1(Part 1), 3–25.

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching*, 25(8), 689–703.

Gibbons, S. J., Hirsch, L. S., Kimmel, H., Rockland, R., & Bloom, J. (2004). *Middle school students' attitudes to and knowledge about engineering*. Presented at the International Conference on Engineering Education, Gainesville, FL.

Gogolin, L., & Swartz, F. (1992). A quantitative and qualitative inquiry into the attitudes toward science of nonscience college students. *Journal of Research in Science Teaching*, 29(5), 487–504.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430.

Häussler, P., & Hoffmann, L. (2002). An intervention study to enhance girls' interest, self-concept, and achievement in physics classes. *Journal of Research in Science Teaching*, 39(9), 870–888.

Hirsch, L. S., Gibbons, S. J., Kimmel, H., Rockland, R., & Bloom, J. (2003). High school students' attitudes to and knowledge about engineering (p. F2A7-12). Presented at the 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE.

IBM Corporation. (2013). IBM SPSS Statistics for Windows (Version 22.0). Armonk, NY: IBM Corporation.

Kelloway, E. K. (2015). *Using Mplus for structural equation modeling: A researcher's guide* (2nd ed.). SAGE Publications, Inc.

Lichtenstein, M. J., Owen, S. V., Blalock, C. L., Liu, Y., Ramirez, K. A., Pruski, L. A., … Toepperwein, M. A. (2008). Psychometric reevaluation of the scientific attitude inventory-revised (SAI-II). *Journal of Research in Science Teaching*, 45(5), 600–616. https://doi.org/10.1002/tea.20244

Lindahl, B. (2007). A longitudinal study of students' attitudes towards science and choice of career. Presented at the NARST Annual Conference, April 15-18, 2007, New Orleans.

Lyons, T. (2006). Different countries, same science classes: Students' experiences of school science in their own words. *International Journal of Science Education*, 28(6), 591–613.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

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. https://doi.org/10.1080/09500690902792385

Miller, P. H., Blessing, J. S., & Schwartz, S. (2006). Gender differences in high-school students' views about science. *International Journal of Science Education*, 28(4), 363–381.

Moore, R. W., & Foy, R. L. H. (1997). The scientific attitude inventory: A revision (SAI II). *Journal of Research in Science Teaching*, 34(4), 327–336.

Moore, R. W., & Sutman, F. X. (1970). The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching*, 7(2), 85–94.

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review*, 84(308), 109–116.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Academies of Sciences, Engineering, and Medicine. (2017). *Building America's skilled technical workforce*. The National Academies Press. https://doi.org/10.17226/23472

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine. (2010). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. https://doi.org/10.17226/12984

National Research Council, & National Academy of Engineering. (2014). C*areer choices of female engineers: A summary of a workshop*. https://doi.org/10.17226/18810

OECD. (2008). *Encouraging students' interest in science and technology studies.* OECD. Retrieved from http://dx.doi.org/10.1787/9789264040892-en

Owen, S. V., Toepperwein, M. A., Marshall, C. E., Lichtenstein, M. J., Blalock, C. L., Liu, Y., … Grimes, K. (2008). Finding pearls: Psychometric reevaluation of the Simpson–Troost Attitude Questionnaire (STAQ). *Science Education*, 92(6), 1076–1095.

Royal Society, The. (2006). *Taking a leading role*. The Royal Society.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, 75(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Sawyer, R. K. (2006). Introduction: The new science of learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 1–16). New York, NY: Cambridge University Press.

Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research*, 39(2), 295.

Simpson, R. D., & Troost, K. M. (1982). Influences on commitment to and learning of science among adolescent students. *Science Education*, 66(5), 763–781.

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science*, 312(5777), 1143–1144. https://doi.org/10.1126/science.1128690

Tytler, R. (2014). Attitudes, identity, and aspirations toward science. In *Handbook of research on science education* (pp. 82–103).

Vaughan, G. M., & Hogg, M. A. (2013). *Social Psychology*. Pearson Higher Education AU.

Venville, G. J., Wallace, J., Rennie, L. J., & Malone, J. A. (2002). Curriculum integration: Eroding the high ground of science as a school subject? *Studies in Science Education*, 37, 43–83.

Weinburgh, M. H., & 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.

Author Information

Cathy P. Lachapelle
Museum of Science, Boston
Engineering is Elementary
1 Science Park, Boston, MA 02114
U.S.A.
Contact e-mail: *clachapelle@mos.org*

Robert T. Brennan
Harvard T. H. Chan School of Public Health
Department of Global Health and Population
Harvard Humanitarian Initiative
14 Story Street
Cambridge, MA 02138
U.S.A.

Appendix

Engineering Interest and Attitudes Assessment

We are interested in learning about your opinions of engineering. Please answer each question honestly. Mark how strongly you agree or disagree after each statement. Thank you very much!		Strongly Agree	Disagree Somewhat	Not Sure	Agree Somewhat	Strongly Agree
1. It is important for me to understand engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
2. Engineering helps me to understand today's world.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
3. I enjoy studying engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
4. Boys are better at engineering than girls.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
5. I would like to work with other engineers to solve engineering problems.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
6. We learn about interesting things when we do engineering in school.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
7. I really want to learn engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
8. Girls are better at engineering than boys.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
9. Engineering is useful in helping to solve the problems of everyday life.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
10. We learn about important things when we do engineering in school.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
11. Engineering is easy for me.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
12. Engineering is fun.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
13. When we do engineering, we use a lot of interesting materials and tools.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
14. It is important to understand engineering in order to get a good job.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
15. Girls have a harder time understanding engineering than boys.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
16. I would like to learn more about engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
17. I am interested when we do engineering in school.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
18. Engineers help make people's lives better.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
19. Girls and boys are equally good at engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
20. I try hard to do well in engineering.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
21. I know what engineers do for their jobs.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
22. I would enjoy being an engineer when I grow up.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
23. Boys have a harder time understanding engineering than girls.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4
24. Engineering is really important for my country.	Last summer, I would have said:	0	1	2	3	4
	Now I would say:	0	1	2	3	4