

# Validating Common Measures of Self-Efficacy and Career Attitudes within Informal Health Education for Middle and High School Students

Karen Peterman,<sup>1\*</sup> Kelley Withy,<sup>2</sup> and Rachel Boulay<sup>3</sup>

<sup>1</sup>Karen Peterman Consulting, Durham, NC 27707; <sup>2</sup>Department of Complementary and Integrative Medicine and <sup>3</sup>Office of the Dean, John A. Burns School of Medicine, University of Hawaii Honolulu, HI 96813

## ABSTRACT

A common challenge in the evaluation of K–12 science education is identifying valid scales that are an appropriate fit for both a student's age and the educational outcomes of interest. Though many new scales have been validated in recent years, there is much to learn about the appropriate educational contexts and audiences for these measures. This study investigated two such scales, the DEVISE Self-Efficacy for Science scale and the Career Interest Questionnaire (CIQ), within the context of two related health sciences projects. Consistent patterns were found in the reliability of each scale across three age groups (middle school, high school, early college) and within the context of each project. As expected, self-efficacy and career interest, as measured through these scales, were found to be correlated. The pattern of results for CIQ scores was also similar to that reported in other literature. This study provides examples of how practitioners can validate established measures for new and specific contexts and provides some evidence to support the use of the scales studied in health science education contexts.

A common challenge in the evaluation of K–12 science education is identifying valid scales that have a documented history of measuring outcomes of interest within similar learning contexts and with similar target audiences. In recent years, many new scales have been validated to measure common outcomes in science education. For example, the National Science Foundation (NSF) funded the DEVISE project in 2010 to develop common measures that could be used to evaluate nine outcomes within the context of citizen science projects ([www.birds.cornell.edu/citscitoolkit/evaluation/instruments](http://www.birds.cornell.edu/citscitoolkit/evaluation/instruments)). Similarly, the Learning Activation Lab has received funding from both the Gordon and Betty Moore Foundation and the NSF to develop a suite of measures that can be used to evaluate and study science, technology, engineering, and mathematics (STEM) engagement, learning, and innovation for youth ([www.activationlab.org/about](http://www.activationlab.org/about)). Other instruments, such as the STEM Semantics Scale and the Career Interest Questionnaire (CIQ), have evolved to become common measures as subsets of the science education community learn about and begin to use such instruments across projects (Peterman *et al.*, 2016).

Common measures that are used in cross-project analyses provide the potential to propel learning related to science education and evaluation alike. Even so, educators and practitioners often lack both the expertise in measurement theory and the time required to conduct new validation studies to ensure that existing scales function in their context. As common measures become more prevalent, practical examples are needed to guide practitioners through steps that can be taken to determine whether and how it is appropriate to use these measures. The current study is a step in that

Graham F. Hatfull, *Monitoring Editor*

Submitted July 10, 2017; Revised January 16, 2018; Accepted January 24, 2018

CBE Life Sci Educ June 1, 2018 17:ar26

DOI:10.1187/cbe.17-07-0122

\*Address correspondence to: Karen Peterman ([karenpetermanphd@gmail.com](mailto:karenpetermanphd@gmail.com)).

© 2018 K. Peterman *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

direction, providing an example of how to explore the validity of an established scale with a specific context, namely informal health science education projects. For the purposes of the current study, informal health science projects were defined as learning opportunities that focused on health science topics and that took place in out-of-school settings that provided the opportunity for free-choice learning. The particular programs in the current study were designed to provide a range of engaging hands-on experiences related to health sciences, and particularly to careers in the health sciences. The two scales in this study each measured a construct that is of interest to a wide range of educators: science self-efficacy and science career interest. Each scale was explored across three age levels and two informal health science education projects to answer the following research question: To what extent can the DEVISE Self-Efficacy for Science (SES) scale and CIQ be used to gather reliable and valid data for middle and high school students in informal health career programs?

## LITERATURE REVIEW

### Science Self-Efficacy

Self-efficacy has been defined as the strength of one's belief in one's own competence (Bandura, 1997). The construct is an extension of Bandura's (1986) social cognitive theory, which states that self-reflection allows individuals to assess their knowledge, experiences, and thoughts as a means of determining their likelihood of success. These perceptions then lead to action, as people tend to participate in activities that will result in positive experiences. A wide range of both qualitative and quantitative approaches have been used to study the kinds of mastery and vicarious experiences that result in feelings of self-efficacy in academic contexts (see Usher and Pajares, 2008, for a review). More recently, advanced statistical methods have allowed educational researchers to identify profiles in the sources of self-efficacy that predict achievement (Chen and Usher, 2015). Educators have been encouraged to apply these findings by choosing science activities that offer a high likelihood of early success and then increasing the difficulty of tasks in strategic ways to support perceptions of mastery. These studies, and the positive relation between self-efficacy and achievement, make improving self-efficacy a common goal of educational programs and thus a common outcome measured via program evaluation.

Measures of self-efficacy must be domain specific to document the range of task demands that determine ability within that domain (Bandura, 2005). Science self-efficacy has been found to predict a number of behaviors and outcomes, including the science grades of middle school students (Britner and Pajares, 2001), the science abilities of high school students (Jansen *et al.*, 2015), the academic performance of nursing students (Andrew, 1998), persistence in science among college students (Byars-Winston *et al.*, 2010; Hanauer *et al.*, 2016), and earning a bachelor's degree (Larson *et al.*, 2014).

Given the construct's established history in the literature, it is not surprising that educators and evaluators have often targeted science self-efficacy as an outcome of interest. Several multivariate scales have been developed in recent years to measure self-efficacy and other related constructs. The Sources of Science Self-Efficacy scale (Britner and Pajares, 2006), the Science Motivation Questionnaire (Glynn and Koballa, 2006), and

the Self-Efficacy in Technology and Science—Short Form (Lamb *et al.*, 2014) are examples. While these scales offer key advances for research on science self-efficacy, they were not created for evaluation purposes. One exception is the DEVISE SES scale, part of a suite of common measures that were developed originally for citizen science projects. As defined in the user guide (see the Supplemental Material), the scale “measures one's confidence in learning science topics, engaging in scientific activities and more generally in being a scientist.” This definition was considered an ideal fit for the projects included in this study, because they were designed to provide students with a range of hands-on mastery experiences with health sciences, with the hope that self-efficacy would be enhanced over time as students accumulated these experiences. The scale's dual focus on science learning and science skill was also of interest, and the brevity of the scale was considered a strength, given that shorter scales are more appropriate for informal learning environments. To our knowledge, this is the first study to examine the reliability of the SES scale with a student population and outside the context of citizen science.

### Science Career Interest

Career interest and attitudes are other common outcomes of interest for science education projects, particularly those with goals related to workforce development. Research has shown that student dispositions toward science are shaped by middle school (George *et al.*, 1992; Osborne *et al.*, 2003; Maltese and Tai, 2010) and that these attitudes predict motivation to learn science and to pursue science careers (Tai *et al.*, 2006; Maltese and Tai, 2011; Dabney *et al.*, 2012). Many science education efforts, including those featured in this study, were created to foster and sustain positive attitudes toward science careers. Indeed, experiences such as summer science camps (Kong *et al.*, 2014), science reading, science competitions (Dabney *et al.*, 2012), and undergraduate research experiences (Lopatto, 2007; Villarejo *et al.*, 2008) are related to science career interest.

A growing number of valid and reliable measures have been published in recent years to measure students' science career interests, and any one of these has the potential to serve as a common measure for educational programs that target this outcome. The Student Interest in Technology and Science survey (Romine *et al.*, 2014), the Assessment of Interest in Medicine and Science (Romine *et al.*, 2016), and the Educational and Career Interest Scale (Oh *et al.*, 2013) are all examples. One instrument that already has a history as a common measure is the CIQ (Tyler-Wood *et al.*, 2010), the second scale of interest for the current study. The CIQ is used commonly across NSF ITEST projects (Peterman *et al.*, 2016) and has been used in both K–12 and informal learning contexts to measure science career interest, intentions to pursue a science career, and the perceived importance of science careers (Tyler-Wood *et al.*, 2010; Christensen *et al.*, 2014; Peterman *et al.*, 2016; Christensen and Knezek, 2017). Looking at students across existing studies, there seems to be a general increase in CIQ scores with age among students who have elected to participate in supplemental education programs. For example, the middle school students' scores reported in Christensen and Knezek (2017) are lower than those found for high school students in a separate study published by the same authors (Christensen *et al.*, 2014). The scores of these same high school

students were lower than those of the college students (Tyler-Wood *et al.*, 2012).

Promoting and sustaining interest in health science careers are primary goals of the projects in this study. There were no known validated scales that specifically measured interest in health science careers at the time this study began. Given the CIQ's demonstrated reliability across educational contexts and age groups, it seemed to have the greatest potential to serve as a general measure of science career interest across student programs and groups. To our knowledge, this is the first study to examine the reliability of the CIQ within the context of health science education.

### Science Self-Efficacy and Science Career Interest

Though this section began by presenting science self-efficacy in isolation from science career attitudes and interests, research has also found that these constructs are related. The importance of self-efficacy as a construct that informs career interest has been established by and studied most directly through social cognitive career theory (SCCT) (Lent *et al.*, 1994, 2000). Self-efficacy is one of three pillars that support all SCCT models, and it has been cited as the most studied predictor of career interest (Lent *et al.*, 2002; for a recent review of SCCT in relation to measurement and the biological sciences, see Byars-Winston *et al.*, 2016). While a number of studies have focused on self-efficacy and career interest generally, fewer have focused on science self-efficacy and science career choice specifically. Recent studies have established the positive relation between these constructs among middle school students (Fouad *et al.*, 1997; Navarro *et al.*, 2007; Kier *et al.*, 2014; Nugent *et al.*, 2015), for high school students' interest in biology careers (Uitto, 2014), and among those who persist as undergraduate nursing students (Restubog *et al.*, 2010). Given the goals of the programs included in this study and the robust literature that supports the relation between science self-efficacy and science career outcomes, the current study was also designed to explore whether these constructs are related when measured via the SES and CIQ.

## METHOD

### Setting

The data for this study were collected within the context of two informal health science projects that are part of a program called Hawaii Science Career Inspiration (HiSCI). Funded by the National Institutes of Health's Science Education Partnership Award, HiSCI's overall goal is to enhance science education resources for teachers and students and help populate the health science workforce of tomorrow. Two student-based HiSCI programs served as the context for this study. The Institutional Review Board at the University of Hawaii approved this study (approval #2016-30289).

**Teen Health Camp.** Launched in 2010, Teen Health Camp (THC) consists of a 1-day camp that includes a number of hands-on workshops related to health sciences (e.g., learning to suture, applying casts, understanding food science). A portion of each THC focuses on health careers. See Dunn *et al.* (2013) for specifics on the underlying philosophy and a detailed program description. To date, THCs have served more than 1000 students.

For the purposes of this study, THC data were collected from students who attended one of three 2016 camps, all of which were overseen by the Hawaii/Pacific Basin Area Health Education Center. Of the 276 students who attended one of these camps, 264 completed the survey (96%). Two camps were hosted on the island of Hawaii; both were held on a Saturday at a local school. The third was hosted on Oahu, and was included as part of a summer camp program for Boys and Girls Club students.

**Pre-Health Career Corps.** Pre-Health Career Corps (PHCC) is a free year-round program for high school and college students who are interested in pursuing careers in health. Its purpose is to increase awareness of, exposure to, and preparation for health careers. A number of activities are offered to support these goals, and students select the activities that are the best fit for their educational and career interests. Options include healthcare shadowing, medical simulations and demos, campus tours, standardized test prep, peer mentorship, career mentorship, research and volunteer experiences, and and practice with writing personal statements for applications and interviewing. PHCC was launched in 2016 and welcomed more than 350 members in its first year.

For the purposes of this study, PHCC data were collected in one of two ways. Both the SES scale and the CIQ were added to the program's online intake form in Fall 2016; all students who joined the program after September 2016 completed the scales in this way. Students who joined PHCC before this date were also asked to complete the scales in Fall 2016 using a Google form. A total of 182 of the 316 PHCC members (58%) completed the survey.

### Participants

Table 1 presents a demographic description of the study sample overall and by program. The majority of students in the sample were in high school (72%) and identified as female (76%). Students selected all applicable categories to describe their race/ethnicity. Most identified as Asian, followed by approximately one-third who identified as Native Hawaiian/other Pacific Islander and/or white.

The number of students in the THC and PHCC samples were roughly equal. Both programs served a majority of female and high school students. The demographic profiles of students differed with regard to age and race/ethnicity. For example, middle school students were only eligible to be part of the THC program, while early-college students were only eligible to be part of PHCC. The racial/ethnic profile for THC was slightly more balanced across category than that for PHCC.

### Instruments

**DEVISE SES Scale.** The SES scale was developed and validated by the Cornell Lab of Ornithology with NSF funding ([www.birds.cornell.edu/citscitolkit/evaluation/instruments](http://www.birds.cornell.edu/citscitolkit/evaluation/instruments)). Though originally developed for citizen science projects, the general focus of the instrument on science self-efficacy makes it an ideal measure for many STEM education programs.

The scale includes eight items that measure a person's confidence in his or her science ability. The first four items focus on ability to learn science and the last four focus on ability to do science activities. Ratings are made on a five-point Likert scale

**TABLE 1. Demographic description of student participants overall and by program type (N = 360)<sup>a</sup>**

Variable	Overall	THC	PHCC
Gender			
Female	76%	73%	80%
Male	24%	27%	20%
Race/ethnicity <sup>b</sup>			
African American/Black	3%	5%	2%
American Indian/Alaska Native	8%	14%	2%
Asian	63%	50%	77%
Hispanic/Latino	15%	19%	11%
Native Hawaiian/other Pacific Islander	35%	47%	23%
White/Caucasian	37%	46%	28%
Age group			
Middle school	10%	21%	–
High school	72%	79%	64%
Early college	18%	–	36%

<sup>a</sup>The number of students who answered each demographic question varied slightly. Valid percentages were reported to correct for this variability.

<sup>b</sup>Percentages total more than 100%, because students chose all categories with which they identified.

for all items, and results can be used at the subscale or overall instrument level. The validation work for the SES scale was done with an adult sample. The reliability of the overall scale was 0.92 (Porticella *et al.*, personal communication).

**The CIQ.** The CIQ comprises three separate subscales that can be used separately or combined into an overall score. The Interest subscale includes four items that measure students' perceptions of being in an environment that is supportive of science careers. The Intent subscale includes five items that focus on students' intentions to pursue educational opportunities that would lead to a science career. The three items that make up the Importance subscale focus on the perceived importance of science careers overall. All items are rated on a five-point Likert scale, with higher numbers indicative of more positive attitudes. Cronbach's alpha for the CIQ typically ranges from 0.70 to 0.93 across subscales (Christensen *et al.*, 2014). The reliability for the overall 12-item instrument typically approaches or surpasses 0.90 (Tyler-Wood *et al.*, 2010; Peterman *et al.*, 2016).

### Coding and Scoring

Each instrument was scored according to established methods. The developers of the SES scale, for example, provide a scoring guide (see the Supplemental Material). These instructions were followed to clean the data, reverse-code items, and create average scores for each student. The guide states that scores can be created for each subscale (i.e., Learning Science and Doing Science) and that all items can be combined into an overall self-efficacy for science score. All three average scores were created.

The validation of the CIQ included an exploratory factor analysis that confirms use of the subscales as distinct factors (Tyler-Wood *et al.*, 2010). The authors also presented criterion-related validity evidence to substantiate the use of CIQ total scores. Subsequent use of the CIQ in the literature has featured both subscale and total scores (Tyler-Wood *et al.*, 2012; Christensen *et al.*, 2014; Christensen and Knezek, 2017; Peterman *et al.*, 2016). Using this prior work as a guide, four

average scores were created for each student's CIQ results, including one score for each of the three subscales and an overall score. Average scores were created only for students with complete data for each subscale or scale of interest.

The demographic data were used to create grouping variables for the analysis based on both program type (THC or PHCC) and school level (middle school, high school, early college). Students in grades 6–8 were considered middle school students; those in grades 9–12 were considered high school students. College-aged students identified themselves by indicating the college or university that they attend currently.

### RESULTS

This section first presents results to explore the internal consistency reliability of each scale, followed by results related to validity. DeVellis (2016) states, "a scale is internally consistent to the extent that its items are highly correlated" (p. 42). Several statistics were used to explore internal consistency reliability, including interitem correlations and a number of reliability statistics related to Cronbach's alpha. Coefficient omega was also calculated for each scale, along with confidence intervals for each estimate; this statistic is now readily available through the userfriendlyscience package in R (see <https://cran.r-project.org/web/packages/userfriendlyscience/userfriendlyscience.pdf> for package documentation).

**Internal Consistency of the SES Scale.** Cronbach's alpha statistics for the SES are presented in Table 2, by program and age group. With regard to the Learning Science subscale, all interitem correlations were above the conventional cutoff of 0.40 (George and Mallery, 2003), indicating that each item relates to the overall score associated with all other items on the subscale. Cronbach's alpha was above the conventional cutoff of 0.70 for both program types and all three age groups, indicating an acceptable level of internal consistency reliability for the subscale (DeVellis, 2016). Even so, the results related to item 3 were mixed, in that Cronbach's alpha improved slightly for PHCC and for early-college students when this item was removed.

With regard to the Doing Science subscale, item 7 showed mixed results; the interitem correlation indicated that this item was not related to the remaining items on the scale for THC and high school students. Further, Cronbach's alpha improved in four of five cases when this item was removed and was only above the conventional cutoff for all programs and age levels under this condition.

These results were used to inform a second reliability analysis, using coefficient omega. All Learning Science items were included in the analysis. Calculations for the Doing Science subscale were conducted for items 5, 6, and 8 only. Coefficient omega was calculated by age group within each program type. Table 3 presents the results. Omega coefficients were above the conventional cutoff for both subscales and for the overall score of the modified seven-item scale. The lower-bound confidence interval for each program and group was also above the conventional cutoff for 10 of 12 instances explored. The two exceptions were for the Doing Science subscale and both age groups involved in THC; the lower bound for this omega falls into the questionable range (George and Mallery, 2003).

TABLE 2. Descriptive statistics, interitem correlations, and alpha statistics for the DEVISE SES scale, by program type and age<sup>a</sup>

Items by scale	Mean (SD)				Interitem correlation				Alpha (if item deleted)					
	By program type		By age group		By program type		By age group		By program type		By age group			
	THC	PHCC	TCH	PHCC	TCH	PHCC	TCH	PHCC	MS	HS	EC	MS	HS	EC
Learning Science items														
Item 1	3.65 (0.80)	4.07 (0.69)	0.64	0.66	0.79	0.76	3.69 (0.82)	3.82 (0.75)	4.17 (0.68)	0.63	0.64	0.74	0.78	0.77
Item 2	3.48 (0.81)	3.87 (0.79)	0.71	0.60	0.69	0.65	3.58 (1.00)	3.61 (0.78)	4.00 (0.84)	0.75	0.69	0.67	0.69	0.64
Item 3	3.32 (0.96)	3.33 (0.92)	0.55	0.45	0.77	0.78	3.28 (1.06)	3.33 (0.94)	3.32 (0.87)	0.52	0.48	0.47	0.78	0.66
Item 4	3.25 (0.92)	3.84 (0.79)	0.53	0.51	0.78	0.73	3.39 (0.96)	3.46 (0.89)	4.05 (0.70)	0.59	0.54	0.46	0.74	0.77
Doing Science items														
Item 5	4.04 (0.62)	4.27 (0.62)	0.47	0.62	0.64	0.75	3.92 (0.72)	4.16 (0.69)	4.28 (0.58)	0.63	0.52	0.66	0.69	0.77
Item 6	3.72 (0.77)	4.02 (0.70)	0.63	0.72	0.42	0.61	3.73 (0.69)	3.86 (0.77)	4.00 (0.67)	0.64	0.70	0.68	0.61	0.69
Item 7	3.41 (0.93)	3.53 (0.95)	0.21	0.42	0.72	0.79	3.49 (0.93)	3.42 (0.96)	3.63 (0.85)	0.47	0.28	0.43	0.48	0.67
Item 8	3.34 (0.90)	3.89 (0.83)	0.43	0.53	0.56	0.71	3.32 (1.03)	3.56 (0.87)	4.03 (0.82)	0.50	0.48	0.59	0.77	0.80

<sup>a</sup>Shaded items for interitem correlations and for alpha are below the conventional cutoffs of 0.40 and 0.70, respectively. Shaded items for “Alpha (if item deleted)” are those that improve alpha beyond that for the overall scale when removed. MS, middle school; HS, high school; EC, early college.

TABLE 3. Reliability of the DEVISE SES scale, by subscale, program, and age group

Subscale	Omega	95% confidence interval	
		Lower bound	Upper bound
Self-Efficacy for Learning Science subscale			
THC middle school	0.81	0.71	0.91
THC high school	0.79	0.74	0.85
PHCC high school	0.74	0.66	0.81
PHCC early college	0.79	0.7	0.87
Modified Self-Efficacy for Doing Science subscale			
THC middle school	0.74	0.60	0.89
THC high school	0.73	0.65	0.80
PHCC high school	0.80	0.74	0.86
PHCC early college	0.81	0.73	0.89
Modified overall self-efficacy score			
THC middle school	0.85	0.77	0.93
THC high school	0.84	0.80	0.88
PHCC high school	0.84	0.80	0.89
PHCC early college	0.87	0.83	0.92

These results provide evidence to demonstrate the internal consistency of the SES Learning Science scale, a three-item version of the Doing Science subscale, and overall self-efficacy scores for the 11 items that constitute these two subscales. Coefficient omega levels were above the conventional cutoff for each of these scores across both programs and across all three age groups.

**Internal Consistency of the CIQ.** The analysis plan presented above for the SES scale was repeated for the CIQ. Cronbach’s alpha statistics are presented in Table 4. For the Interest and Intent subscales, all interitem correlations and Cronbach’s alpha levels were above their respective conventional cutoff scores, and alpha levels for both scales were highest when all items were included in the scale.

Responses were mixed with regard to the CIQ Importance subscale. The interitem correlations for the three items in this scale were above the conventional cutoff with one exception (middle school students’ scores for item 12). Cronbach’s alpha was above the conventional cutoff in three of five instances, and this mixed pattern of results also seemed related to item 12 on the scale. Removing this item improved alpha levels for PHCC students and for two of the three age groups. Current convention states that at least three items are needed to create a scale. Given that the original Importance subscale includes three items, we elected to remove this scale from the remaining analysis.

Table 5 presents the results from coefficient omega for CIQ Interest and Intent subscales and a modified overall career interest score that combines these two subscales. Coefficient omega was calculated by age group within each program type. Omega coefficients were above the conventional cutoff for both subscales and for the overall score of the modified overall scale, with one exception: the Interest scores of early-college students were below 0.70 and the lower-bound confidence interval for this group was in the poor range.

**TABLE 4. Descriptive statistics, interitem correlations, and alpha statistics for the CIQ, by program type and age<sup>a</sup>**

	Mean (SD)		Interitem correlation				Alpha (if item deleted)							
			By program type		By age group		MS		HS		EC			
	THC	PHCC	TCH	PHCC	TCH	PHCC	MS	HS	MS	HS	MS	HS	EC	
<b>Interest items</b>														
Item 1	3.47 (1.07)	4.48 (0.76)	0.70	0.57	0.84	0.77	3.34 (1.05)	3.90 (1.02)	4.64 (0.65)	0.58	0.73	0.82	0.85	0.70
Item 2	3.45 (0.88)	3.99 (0.98)	0.63	0.54	(0.78)	(0.72)	3.71 (0.98)	4.04 (0.94)	4.63 (0.60)	0.78	0.76	(0.81)	(0.79)	(0.63)
Item 3	3.73 (0.95)	4.48 (0.74)	0.73	0.67	(0.82)	(0.73)	3.47 (1.18)	3.70 (1.10)	4.05 (1.03)	0.63	0.64	(0.72)	(0.78)	(0.63)
Item 4	3.46 (1.07)	4.02 (1.07)	0.64	0.55	(0.77)	(0.68)	3.45 (0.98)	3.69 (0.97)	4.00 (0.89)	0.62	0.64	(0.79)	(0.83)	(0.61)
					(0.81)	(0.74)				0.45	0.45		(0.83)	(0.66)
<b>Intent items</b>					0.94	0.92						0.91	0.94	0.92
Item 5	3.80 (0.93)	4.47 (0.73)	0.81	0.85	(0.92)	(0.90)	3.42 (0.92)	4.17 (0.89)	4.44 (0.69)	0.85	0.84	(0.87)	(0.93)	(0.89)
Item 6	3.74 (0.95)	4.54 (0.71)	0.88	0.81	(0.91)	(0.90)	3.26 (0.89)	4.18 (0.92)	4.55 (0.64)	0.83	0.89	(0.87)	(0.92)	(0.90)
Item 7	3.66 (0.94)	4.39 (0.76)	0.82	0.83	(0.92)	(0.90)	3.24 (0.98)	4.04 (0.91)	4.41 (0.73)	0.74	0.86	(0.89)	(0.93)	(0.90)
Item 8	3.64 (0.98)	4.48 (0.71)	0.86	0.79	(0.91)	(0.91)	3.26 (0.98)	4.02 (0.94)	4.48 (0.64)	0.85	0.86	(0.87)	(0.93)	(0.90)
Item 9	3.74 (0.98)	4.36 (0.76)	0.76	0.72	(0.93)	(0.92)	3.37 (0.97)	4.10 (0.92)	4.25 (0.80)	0.57	0.79	(0.93)	(0.94)	(0.90)
<b>Importance items</b>					0.65	0.81						0.68	0.72	0.85
Item 10	4.02 (0.92)	4.62 (0.92)	0.45	0.69	(0.57)	(0.71)	3.58 (1.03)	4.32 (0.82)	4.78 (0.45)	0.45	0.55	(0.65)	(0.62)	(0.77)
Item 11	4.37 (0.74)	4.71 (0.60)	0.52	0.74	(0.49)	(0.68)	4.26 (0.83)	4.29 (0.79)	4.75 (0.50)	0.68	0.59	(0.37)	(0.58)	(0.71)
Item 12	4.15 (0.79)	4.46 (0.77)	0.43	0.59	(0.60)	(0.84)	3.89 (0.95)	4.29 (0.79)	4.63 (0.55)	0.38	0.49	(0.72)	(0.69)	(0.89)

<sup>a</sup>Shaded items for interitem correlations and for alpha are below the conventional cutoffs, of 0.40 and 0.70, respectively. Shaded items for “Alpha (if item deleted)” are those that improve alpha beyond that for the overall scale when removed. MS, middle school; HS, high school; EC, early college.

**TABLE 5. Reliability of the CIQ, by subscale, program, and age group**

Subscale	Omega	95% Confidence interval	
		Lower bound	Upper bound
<b>Interest subscale</b>			
THC middle school	0.82	0.73	0.91
THC high school	0.84	0.80	0.88
PHCC high school	0.76	0.69	0.83
PHCC early college	0.63	0.48	0.77
<b>Intent subscale</b>			
THC middle school	0.91	0.86	0.96
THC high school	0.94	0.92	0.95
PHCC high school	0.93	0.91	0.95
PHCC early college	0.91	0.88	0.95
<b>Modified overall career attitude score</b>			
THC middle school	0.91	0.87	0.96
THC high school	0.93	0.91	0.95
PHCC high school	0.90	0.87	0.93
PHCC early college	0.84	0.77	0.90

These results provide evidence to demonstrate the internal consistency of the CIQ Interest and Intent subscales, as well as an overall career attitude score created by summing the Interest and Intent items for middle school and high school students. The results also support the use of the Intent subscale with early-college students. Coefficient omega levels were above the conventional cutoff for each of these scores.

Two additional analyses were conducted to explore whether the SES and CIQ performed as expected, based on the existing literature. The first of these was a correlation analysis. The literature on science self-efficacy and science career interest indicates that these constructs are positively related, such that increased self-efficacy in science is positively correlated with and predictive of interest in and pursuit of science careers (Fouad *et al.*, 1997; Restubog *et al.*, 2010; Uitto, 2014; Nugent *et al.*, 2015). As such, a positive correlation between students’ scores on the SES and CIQ would provide additional evidence to support their use.

Mean scores were calculated for HiSCI students based on the reliability results. Overall CIQ scores were calculated differently for THC and PHCC to include only items for the two subscales that were found to be reliable with each group. Correlation analyses were then conducted between SES and CIQ scores across student scores and by program type. Table 6 presents the correlation coefficients. Statistically significant positive correlations were found between all measures for both HiSCI programs.

Correlations between the SES and CIQ for THC students ranged from 0.36 to 0.53, and thus were in the moderate range, with one exception that was in the high range. The relation between scores of PHCC students was more variable overall, with correlations ranging from 0.21 to 0.52. Correlations between CIQ Importance and SES Learning Science were the lowest, in the small to moderate range are presented in Table 6. All other correlations were in the moderate to large range. The correlations between the overall scores for the two scales were in the high moderate range, at 0.49 for both groups.

TABLE 6. Correlation coefficients for SES and CIQ scores

Scale	SES Learning Science		SES modified Doing Science		SES modified overall	
	THC	PHCC	TCH	PHCC	TCH	PHCC
CIQ Interest	0.36*** (n = 176)	—	0.44*** (n = 177)	—	0.41*** (n = 173)	—
CIQ Intent	0.36*** (n = 176)	0.45*** (n = 179)	0.44*** (n = 176)	0.52*** (n = 180)	0.44*** (n = 177)	0.52*** (n = 179)
CIQ Importance	—	0.21** (n = 179)	—	0.37*** (n = 180)	—	0.30*** (n = 179)
CIQ modified overall	0.42*** (n = 173)	0.40*** (n = 179)	0.53*** (n = 174)	0.51*** (n = 180)	0.49*** (n = 170)	0.49*** (n = 179)

\*\* $p < 0.01$ .\*\*\* $p < 0.001$ .

**Validity Evidence.** A second set of analyses compared the scores of HiSCI students with those of similar groups found in the academic literature. Groups were considered similar if they were of the same grade level and had self-identified as being interested in STEM. For example, one-sample  $t$  tests were used to compare the pretest scores of middle school students who reported an interest in science careers (Christensen and Knezek, 2017) with the scores of the self-selected group of students who attended HiSCI's THC. Comparisons focused only on the three CIQ scores that were reliable for HiSCI students in this age group. Note that the overall scores for THC students were calculated based on the modified scale, while those in the literature were for the full scale; this comparison was considered appropriate, given that the  $t$  test is based on mean scores. Mean scores for each group are presented in Table 7, along with the results from the  $t$  tests. THC students had comparable CIQ scores on

each scale: no differences were found between the scores of HiSCI middle school students and those in the comparison sample. The similarity in students' scores provides additional evidence to support the use of the CIQ with middle school students in health science programs like HiSCI.

A similar strategy was used to explore the CIQ for HiSCI high school students who were compared with high school students who self-selected to be part of a residential science and math program (Christensen *et al.*, 2014). Comparisons focused on the three CIQ scores that were reliable for all HiSCI high school students, including the modified overall score. Table 7 also presents the mean and standard deviation for each group and the results from the one-sample  $t$  tests. HiSCI students' scores were similar to those from the comparison group with regard to Intent. HiSCI high school students had significantly lower Interest and overall scores than students in the comparison group.

TABLE 7. One-sample  $t$  tests comparing HiSCI students with groups from the literature

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>p</i>
HiSCI middle school students and comparison group of middle schoolers interested in science careers (Christensen and Knezek, 2017)					
Interest subscale					
HiSCI middle school	38	3.49	0.85	0.68	0.50
Population mean	363	3.40			
Intent subscale					
HiSCI middle school	38	3.31	0.80	-0.92	0.36
Population mean	363	3.43			
Overall career interest					
HiSCI middle school	38	3.39	0.77	-1.19	0.21
Population mean	363	3.54			
HiSCI high school students and comparison group of high schoolers from a residential science/math program (Christensen <i>et al.</i> , 2014)					
Interest subscale					
HiSCI high school	257	3.84	0.84	-6.20	<0.001
Population mean	364	4.16			
Intent subscale					
HiSCI high school	255	4.11	0.83	1.58	0.51
Population mean	364	4.03			
Overall career interest					
HiSCI high school	254	3.99	0.78	-3.68	<0.001
Population mean	364	4.17			
HiSCI Early-college students and comparison group of upper-level MCAT students (Tyler-Wood <i>et al.</i> , 2012)					
Intent subscale					
HiSCI early college	64	4.43	0.61	-2.84	<0.01
Population mean	364	4.64			
Importance subscale					
HiSCI early college	64	4.72	0.44	-0.93	0.36
Population mean	364	4.77			

This pattern of results provides mixed evidence with regard to using the CIQ with high school students.

A final test was conducted to explore the scores of HiSCI early-college students in relation to upper-level science majors who elected to take a Medical College Admission Test (MCAT) course (Tyler-Wood *et al.*, 2012). Two CIQ subscale scores were available and reliable for both groups; overall CIQ scores were not published for the comparison group and thus could not be included in the analysis. The final section of Table 7 presents the results. HiSCI scores were significantly below those of the comparison group in two of three cases tested; HiSCI students reported Importance scores similar to those of the upper-level in the MCAT comparison group sample.

In sum, positive correlations were found between SES and CIQ scores across program type and age group. In addition, HiSCI students had similar scores to those found in the literature for five of the eight comparisons made. Similarities between HiSCI and comparison group scores were found across each of the three age groups investigated. Though the results from the one-sample *t* tests could be stronger, the consistent pattern of results between HiSCI results and those in the literature provide some confidence that the scales are functioning as intended.

## DISCUSSION

This study was designed to explore the validity of two established instruments within a specific context. Though there has been recent support for the development of common measures, practitioners cannot assume that an established measure will work equally well across different age groups and educational contexts. Few examples exist to demonstrate steps that practitioners should take to determine whether those measures are an appropriate fit for their program. This study provides one such approach by using both Cronbach's alpha, the most commonly used reliability statistic, as well as omega estimates that have been used less frequently in the past but are now readily available. Coefficient alpha has been criticized as a conservative estimate of internal consistency that is often misused and over-interpreted (Sijtsma, 2009; DeVellis, 2016). Even so, these same critics note that the statistic still has value, given its history in the literature and the fact that it is readily available in SPSS. DeVellis notes further that coefficient omega is a more accurate calculation of reliability than coefficient alpha, and suggests that it may replace alpha over time, once its own publication history is established. This study makes a contribution in this regard by integrating omega estimates into our reliability analysis.

This study also illustrates how to respond when common measures do and do not perform as expected. Three of the five subscales tested performed as we hoped, across both informal health science programs and the three age groups tested, but two did not. These results make a contribution in two ways. First, they reiterate the fact that educators and evaluators cannot assume that common measures will perform consistently across contexts. Second, they provide examples of the data-driven choices that might be made when scales do not perform as expected. In the case of the SES Doing Science subscale, the data indicated that dropping a particular item would yield the internal consistency needed to use scores for evaluation purposes. The results for CIQ Interest and Importance subscales indicated that these scales can only be used for a subset of our

target audiences. The reliability evidence from this study supports the use of the SES Learning Science, modified SES Doing Science, and CIQ Intent scales within the context of informal health science projects. The CIQ Interest subscale was also found to be an appropriate fit for middle and high school audiences, and the Importance subscale was found to be an appropriate fit for early-college audiences. Additional validation work is needed for these scales to function as common measures across the full range of audiences included in this study.

In addition to the results related to scale reliability, this study compared our data with results in the literature as another way to explore scale functioning. With regard to the CIQ, the pattern of results in the HiSCI data mirrors that from the literature, in that student scores increased from middle to high school and from high school to early college (Tyler-Wood *et al.*, 2012; Christensen *et al.*, 2014; Christensen and Knezek, 2017). Further, five of the eight comparisons indicated no differences in the CIQ scores of HiSCI students compared with similar student groups identified through the literature. The established relationship between science self-efficacy and science career interest in the literature (Wigfield and Eccles, 2000; Lent *et al.*, 2002; Restubog *et al.*, 2010; Uitto, 2014; Nugent *et al.*, 2015) was also replicated with the HiSCI data. The positive relationship between these constructs, as measured by the SES and CIQ, was found across subscale, age group, and program type, providing evidence to support their combined use.

The evidence established in this study is encouraging, though it is clear that more work could be done. Those who are considering either these or other common instruments to evaluate their health science project might begin by conducting think-aloud interviews with their target audience(s) before selecting a scale. We omitted this step based on time constraints and the successful use of these scales in prior evaluation efforts. It is unclear whether think-aloud interviews with youth would have resulted in different choices in the scales used for this study. Though think-aloud interviews are considered a best practice in evaluation, few educators and evaluators include this step in their process for identifying the best scale to use for a given project. With the number of scales currently available to educators and evaluators, taking the time to conduct a small number of think-aloud interviews with a target audience might become standard practice for selecting a scale from a list of possible common measures.

A limitation of the current study is the fact that we did not explore the extent to which either instrument can be used to detect pre–post change in student scores. The literature does include examples of the CIQ as a tool that can detect pre–post change in other contexts (Peterman *et al.*, 2016), and the instructions for the SES indicate that it can be used for such purposes. The current study was intentionally limited to baseline-only measures for two reasons. First, the primary intention of the study was to explore whether each instrument was an appropriate tool for the audiences served by the HiSCI program, with the hope that these measures could be used in later evaluation and research studies to document key information about students' science attitudes and dispositions. It seemed premature to continue collecting data with these tools before reliability or validity had been established.

Second, pre–post change would only be appropriate to measure in one of the two programs used for the current study.

Constructs such as self-efficacy and career interest are appropriate outcomes for longer-term programs like PHCC that offer continuous engagement as part of an intervention, but not so for short-term activities such as THC. The data from the current study have verified the utility of both scales within the context of PHCC and will be integrated into future evaluations of the program, including the extent to which the program results in pre-post change for students.

The past several years have been a productive time in scale development, with both private and federal foundations investing in the creation of new scales that have the potential to be used as common measures across evaluations. There is still much to learn about whether common measures have the potential to detect outcomes across a range of contexts. The current study highlights the need for intentional selection and study of common instruments as they are used in new contexts. Studies like these are necessary to generate learning that will enhance evaluation practice in the short term and life science programming by extension. Moreover, the use of common measures can enable cross-programmatic analysis of different programs measuring similar outcomes, thereby adding to our understanding of what works in different contexts and for different audiences. Evaluators and researchers who consider using common measures across projects need to do so with intentionality to ensure that the measure selected is reliable within the context of the project being evaluated. The current study provides examples for how to begin such work, as well as initial evidence to support the use of some subscales from the common measures tested to evaluate health science projects.

## ACKNOWLEDGMENTS

Funding for this research was provided through the following grants: U.S. National Institutes of Health Grants No. R25GM129175, R25OD020246, P30GM103341, T32HL115505, and P20GM113134.

## REFERENCES

- Andrew, S. (1998). Self-efficacy as a predictor of academic performance in science. *Journal of Advanced Nursing*, *27*, 596–603. doi: 10.1046/j.1365-2648.1998.00550.x
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Upper Saddle River, NJ: Prentice Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2005). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, *5*, 307–337.
- Britner, S. L., & Pajares, F. (2001). Self-efficacy beliefs, motivation, race, and gender in middle school science. *Journal of Women and Minorities in Science and Engineering*, *7*(4), 269–283.
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*, 485–499. doi: 10.1002/tea.20131
- Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: A multiple-groups analysis. *Journal of Counseling Psychology*, *57*(2), 205.
- Byars-Winston, A., Rogers, J., Branchaw, J., Pribbenow, C., Hanke, R., & Pfund, C. (2016). New measures assessing predictors of academic persistence for historically underrepresented racial/ethnic undergraduates in science. *CBE—Life Sciences Education*, *15*(3), ar32.
- Chen, J. A., & Usher, E. L. (2015). Profiles of the sources of science self-efficacy. *Articles 13*. <http://publish.wm.edu/articles/13>
- Christensen, R., & Knezek, G. (2017). Relationship of middle school student STEM interest to career intent. *Journal of Education in Science, Environment and Health*, *3*(1), 1–13.
- Christensen, R., Knezek, G., & Tyler-Wood, T. (2014). Student perceptions of science, technology, engineering and mathematics (STEM) content and careers. *Computers in Human Behavior*, *34*, 173–186.
- Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnert, G., Sadler, P. M., & Hazari, Z. (2012). Out-of-school time science activities and their association with career interest in STEM. *International Journal of Science Education, Part B*, *2*(1), 63–79.
- DeVellis, R. F. (2016). *Scale development*. Newbury Park, CA: Sage.
- Dunn, B. S., Duquez, E., Schiff, T., Lau, N., Malate, A. R., & Withy, K. (2013). Medical school hotline. *Hawai'i Journal of Medicine and Public Health*, *72*(4), 140–142.
- Fouad, N. A., Smith, P. L., & Enochs, L. (1997). Reliability and validity evidence for the middle school self-efficacy scale. *Measurement and Evaluation in Counseling and Development*, *30*(1), 17.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 Update (4th ed.). Boston: Allyn & Bacon.
- George, P., Stevenson, C., Thomason, J., & Beane, J. (1992). *The middle school and beyond*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Glynn, S. M., & Koballa, T. R. (2006). Motivation to learn college science. In Mintzes, J. J., & Leonard, W. H. (Eds.), *Handbook of college science teaching* (pp. 25–32). Arlington, VA: National Science Teachers Association Press.
- Hanauer, D. I., Graham, M. J., & Hatfull, G. F. (2016). A measure of college student persistence in the sciences (PITS). *CBE—Life Sciences Education*, *15*(4), ar54. doi: 10.1187/cbe.15-09-0185
- Jansen, M., Scherer, R., & Schroeders, U. (2015). Students' self-concept and self-efficacy in the sciences: Differential relations to antecedents and educational outcomes. *Contemporary Educational Psychology*, *41*, 13–24.
- 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*(3), 461–481.
- Kong, X., Dabney, K. P., & Tai, R. H. (2014). The association between science summer camps and career interest in science and engineering. *International Journal of Science Education, Part B*, *4*(1), 54–65.
- Lamb, R. L., Vallett, D., & Annetta, L. (2014). Development of a short-form measure of science and technology self-efficacy using Rasch analysis. *Journal of Science Education and Technology*, *23*(5), 641–657.
- Larson, L. M., Pesch, K. M., Bonitz, V. S., Wu, T. F., & Werbel, J. D. (2014). Graduating with a science major: The roles of first-year science interests and educational aspirations. *Journal of Career Assessment*, *22*(3), 479–488.
- Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior*, *45*(1), 79–122.
- Lent, R. W., Brown, S. D., & Hackett, G. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology*, *47*(1), 36.
- Lent, R. W., Brown, S. D., & Hackett, G. (2002). Social cognitive career theory. *Career Choice and Development*, *4*, 255–311.
- Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. *CBE—Life Sciences Education*, *6*(4), 297–306.
- 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.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, *95*(5), 877–907.
- Navarro, R. L., Flores, L. Y., & Worthington, R. L. (2007). Mexican American middle school students' goal intentions in mathematics and science: A test of social cognitive career theory. *Journal of Counseling Psychology*, *54*(3), 320.
- Nugent, G., Barker, B., Welch, G., Grandgenett, N., Wu, C., & Nelson, C. (2015). A model of factors contributing to STEM learning and career orientation. *International Journal of Science Education*, *37*(7), 1067–1088.

- Oh, Y. J., Jia, Y., Lorentson, M., & LaBanca, F. (2013). Development of the educational and career interest scale in science, technology, and mathematics for high school students. *Journal of Science Education and Technology, 22*(5), 780–790.
- Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049–1079.
- Peterman, K., Kermish-Allen, R., Knezek, G., Christensen, R., & Tyler-Wood, T. (2016). Measuring student career interest within the context of technology-enhanced STEM projects: A cross-project comparison study based on the Career Interest Questionnaire. *Journal of Science Education and Technology, 25*(6), 833–845.
- Restubog, S. L. D., Florentino, A. R., & Garcia, P. R. J. M. (2010). The mediating roles of career self-efficacy and career decidedness in the relationship between contextual support and persistence. *Journal of Vocational Behavior, 77*(2), 186–195.
- Romine, W., Sadler, T., Presley, M., & Klosterman, M. (2014). Student Interest in Technology and Science (SITS) survey: Development, validation, and use of a new instrument. *International Journal of Science & Mathematics Education, 12*(2), 261–283.
- Romine, W. L., Miller, M. E., Knese, S. A., & Folk, W. R. (2016). Multilevel assessment of middle school students' interest in the health sciences: Development and validation of a new measurement tool. *CBE—Life Sciences Education, 15*(2), ar21.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107.
- Tai, R. H., Liu, Q. C., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312*, 1143–1144.
- Tyler-Wood, T., Ellison, A., Lim, O., & Periathiruvadi, S. (2012). Bringing Up Girls in Science (BUGS): The effectiveness of an afterschool environmental science program for increasing female students' interest in science careers. *Journal of Science Education and Technology, 21*(1), 46–55.
- 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), 341–363.
- Uitto, A. (2014). Interest, attitudes, and self-efficacy beliefs explaining upper-secondary school students' orientation toward biology-related careers. *International Journal of Science & Mathematics Education, 12*(6), 1425–1444.
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751–796.
- Villarejo, M., Barlow, A. E., Kogan, D., Veazey, B. D., & Sweeney, J. K. (2008). Encouraging minority undergraduates to choose science careers: Career paths survey results. *CBE—Life Sciences Education, 7*(4), 394–409.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81.