

Gauging Informal STEM Youth Program Impact: A Conceptual Framework and a Measurement Instrument

Robert H. Tai

*University of Virginia
rht6h@virginia.edu*

Ji Hoon Ryoo

*Yonsei University
jr3gv@virginia.edu*

Claire E. Mitchell

*Montana State University
claire.mitchell@montana.edu*

Xiaoqing Kong

*University of Virginia
xk4wa@virginia.edu*

Angela Skeeles-Worley

*University of Virginia
ads4d@virginia.edu*

John T. Almarode

*James Madison University
almarojt@jmu.edu*

Adam V. Maltese

*Indiana University Bloomington
amaltese@indiana.edu*

Katherine P. Dabney

*Virginia Commonwealth University
kdabney@vcu.edu*



New articles in this journal are licensed under a Creative Commons Attribution 4.0 License. This journal is published by the [University Library System](#), [University of Pittsburgh](#) and is cosponsored by the [University of Pittsburgh Press](#). The Journal of Youth Development is the official peer-reviewed publication of the [National Association of Extension 4-H Youth Development Professionals](#) and the [National AfterSchool Association](#).

Gauging STEM Youth Program Impact

Abstract

STEM education programs are often formulated with a "hands-on activities" focus across a wide array of topics from robotics to rockets to ecology. Traditionally, the impact of these programs is based on surveys of youth on program-specific experiences or the youths' interest and impressions of science in general. In this manuscript, we offer a new approach to analyzing science programming design and youth participant impact. The conceptual framework discussed here concentrates on the organization and analysis of common learning activities and instructional strategies. We establish instrument validity and reliability through an analysis of validity threats and pilot study results. We conclude by using this instrument in an example analysis of a STEM education program.

Key words: activity-based science learning, informal STEM education, out-of-school time science education, informal STEM program impact

Introduction

The aim of this report is to offer a new approach to gauging the impact of STEM-focused youth program activities on youth's active learning engagement. We begin by examining existing measurement tools commonly employed by informal STEM programs. Next, we propose a conceptual framework for an approach to program analysis differing from these existing methods. Then, we offer a new tool for measuring youth active learning engagement, exploring both the validity and reliability of this instrument. Finally, we offer an example of its application and interpretation of program analysis results.

Science-focused activities in both formal and informal education can help spark students' science interest (National Academy of Sciences, 2007). Thus, a better understanding of what aspects of these activities may influence students' preferences can help educators, researchers, and policy makers develop and improve various school-based and out-of-school time programs that purposefully involve customized elements to attract an increasing number of students in the STEM pipeline. A stronger grasp of children's preferences for specific pedagogical strategies can also serve to best match individuals' education selections with career options in the STEM workforce (Carnevale et al., 2010).

Background

Interest in and Attitudes Toward Science

Much effort to engage children in science and engineering has concentrated on content, discovery-related hands-on activities, or a combination of these two elements. Indeed,

Gauging STEM Youth Program Impact

educators often rely on the presentation of fascinating facts and images and/or discovery-related, hands-on activities that involve inquiry-based learning that may include projects and experiments (Swarat et al., 2012). In addition, a robust body of literature suggests students' interests and attitudes towards science are two dominant affective factors that can influence their career expectations and participation in science, technology, engineering, and mathematics (STEM) related fields (Koballa, Jr. & Glynn, 2007; Lent et al., 1997; Luce & Hsi, 2014; Simpson et al., 1994).

Science Interest Development

Students' interest in science is considered an essential indicator of the quality of science education, and a body of literature examines students' interest in science over time (Guzzetti & Bang, 2011; Schiefele, 1991). Interest can be defined as "a content-specific motivational characteristic composed of intrinsic feeling-related and value-related valences" (Schiefele, 1991, p. 299). School-based experiences are considered an important influence on youth science learning interest. In their longitudinal study of 33 high school students, Aschbacher et al. (2010) concluded that students most often attributed their interest in science, engineering, or medicine majors or careers to science teachers or early school-based experiences. This study also revealed the possibility that an individual's interest in science may change over time; findings suggested that some students who showed initial interest in science-related fields later abandoned the STEM pipeline in high school due to their discouraging school experiences (Aschbacher et al., 2010). Several of the interviewees, who were originally interested in science but later lost their interest, mentioned that their teachers discouraged them, or their curriculum contained too few hands-on science activities (Aschbacher et al., 2010). In their investigation of different learning environment elements in science classrooms, Swarat et al. (2012) found that instructional pedagogy that incorporates hands-on activities and allows for engagement with technology can pique youth science interest. These studies identify teachers and school-based activities as influences on students' science interest. However, more work is necessary to understand how instructors and learning activities exert their influence.

Beyond formal settings, research has found that out-of-school time science activities can strongly influence students' interest in science or help them maintain interests they developed in more formal school settings (Aschbacher et al., 2010; Bachman et al., 2008; Chang et al., 2009; Fields, 2009; Johnson, 2011). In her study investigating high school students' perceptions of their participation in a summer astronomy camp, Fields (2009) found that students who took part in the camp gained a deeper understanding of science as well as a stronger projective

Gauging STEM Youth Program Impact

identity in science whereby they began to develop future involvement with science-related activities. These findings have been corroborated by other researchers who have also found that out-of-school time programs can reinforce science identities and improve, particularly for girls, students' interests in science (Barton et al., 2013; Farland-Smith, 2009).

Attitudes Toward Science

Attitudes toward science play an important role in the affective dimension of science learning. Although the construct is not well defined in the literature (Osborne et al., 2003), attitudes can refer to "cognitive and emotional opinions about various aspects of science" (Kind et al., 2007, p. 873). While research supports that students' attitudes towards science and scientists are developed in the formative years between kindergarten and 12th grade, studies have also found that as students mature, they tend to have more negative attitudes towards science (George, 2006; Majumdar et al., 1991). Like the shifting patterns of student interest in science, students' attitudes towards science, particularly girls' attitudes, decline significantly as they progress from elementary schools to middle schools (Barmby et al., 2008).

The reasons for this decline are not well understood, although a line of research has found a relationship between the way science is taught in school and student attitudes towards science (Bhattacharyya & Mead, 2011; George, 2006; Myers & Fouts, 1992; Talton & Simpson, 1986). Work by Barman (1999), Ebenezer & Zoller (1993), Häussler & Hoffmann (2000), and Krajcik et al. (2003) has sought to examine the ways in which instructional practice influences students' perceptions of and attitudes toward science. Results have largely found that there is a compelling relationship between the type of pedagogy employed by teachers (e.g., teacher-centered didactic instruction, collaborative learning, problem-based learning, contextualization of content) and students' science attitude development. Ferreira & Trudel (2012) found that problem-based learning in science can lead to significant positive increases in students' science attitudes. This result is supported by other research (Bhattacharyya & Mead, 2011; Robbins et al., 2005).

Studies employing quantitative methodologies have found that participation in out-of-school time programs with hands-on activities have a positive impact on students' STEM attitudes (Elam et al., 2012; Newell et al., 2015). Specifically, in her study evaluating the attitudes of high school students who participated in robotics competition and those in a comparison sample, Welch (2010) found that participating students held more positive science attitudes in four ways: social implications of science, normality of scientists, attitude towards scientific

Gauging STEM Youth Program Impact

inquiry, and adoption of scientific attitudes. Additional research has offered evidence that students' attitude toward science correlates with achievement in the science classroom (Germann, 1988) and career expectations (Gibson & Chase, 2002). In light of these findings, it appears that developing measures of students' science interest, attitudes, and preferences is the next step.

Table 1. Summary of Existing Measures for Science Interest and Attitude

Name	Measuring
<i>Test of Science-Related Attitudes (TOSRA)</i> (Fraser, 1978)	Social implications of science, normality of scientists, attitude towards scientific inquiry, adoption of scientific attitudes, enjoyment of science lessons, leisure interest in science, and career interest in science.
<i>Scientific Attitude Inventory (SAII)</i> (Moore & Foy, 1997)	Students' perceptions of scientists, becoming scientists, contribution of science to the society, and their interest in science.
<i>Modified Attitude Towards Science Inventory (MATSI)</i> (Weinburgh & Steele, 2000)	Students' attitudes towards science, which include students' uneasiness about learning science, interest in science, attitudes towards the social significance of science, and their beliefs about science teachers.
<i>Colorado Learning Attitudes about Science Survey (CLASS)</i> Biology, Chemistry, and Physics (Adams et al., 2004)	Students' opinions about multiple facets in learning biology, chemistry, and physics respectively: knowledge connection, social significance, problem solving, scientific thinking, and enjoyment in learning these subjects.
<i>Relevance of Science Education (ROSE)</i> (Talisayon et al., 2004)	Multiple aspects of students' opinions about science and science learning: their interest in very specific science related topics, career expectations, within and out of school science experiences, as well as social significance of science and technology.
<i>Survey Items of Situational Interest</i> (Hulleman & Harackiewicz, 2009)	Whether classroom activities will increase ninth and 10 th grade students' motivation in learning science, and asks for students' perceptions of expectancies for success, interest in science, and utility value of science.
<i>S-STEM</i> (Unfried et al., 2015)	Measure's student attitudes towards science, technology, engineering, and mathematics and interest in STEM careers for Grades 4-5 and Grades 6-12.

Existing Measurements for Interest and Attitudes Toward Science

To date, numerous measurements for students' interest in and attitudes towards science learning have been developed. Osborne et al. (2003) summarized five main approaches to measuring students' perceptions of science learning: (a) subject preference studies, asking students to rank their liking of school subjects; (b) attitudes scales, asking students' opinions about statements with Likert-scale response options; (c) interest inventories, asking students to select items of their interest from a list; (d) subject enrollment, collecting data on students' enrollments to various subjects; and (e) qualitative methodologies, collecting qualitative data from individual and group interviews. Among those approaches, attitudes scales were the most used in collecting large-scale data that can also present specific changes in students' interest in or attitudes towards science (see Table 1).

Theoretical Framework

There are several theories of learning and motivation that could be applied to unpack the ways in which students develop preferences for learning activities and experiences. The theoretical insights imparted by rational choice theory and expectancy–value theory are of relevance as they pertain to student learning preferences and choices. These two theories afforded us a model with which to better understand our data.

Rational choice theory (RCT) is a broader model rooted in sociology and behaviorist psychology that provides a useful frame for understanding how an individual develops and/or can change their preferences toward certain activities. RCT helps to explain the idea that all action is fundamentally “rational” in nature and that individuals are continually assessing the potential costs and benefits of any action or decision before they make them (Coleman & Fararo, 1992; Homans, 1961; Scott, 2000). These calculations lead to the development of individual preferences, which can be influenced by many factors such as individual personal habits and commitments, external variables such as availability or existence of alternative options, and assessments of the “positive or negative evaluations individuals attach to possible outcomes of their actions” (Wittek, 2013, p. 688).

Deitrich and List (2013) further elucidate the process of developing preferences and further explain also how they can indeed change or shift over time depending on individual variables and external circumstances. They assert that, according to RCT, preferences are not fixed but rather are based on certain “motivationally salient properties” of alternatives over which

Gauging STEM Youth Program Impact

preferences are held (p. 613). This means that an individual's preferences may shift as different alternatives are presented and/or become—or cease to become—salient. It is these alternative choices, along with an individual's interaction with their social and cultural environments, which continually shape and re-shape their preferences (Loveland, 2003; Sherkat and Wilson, 1995). In this regard, RCT provides a useful frame for us to better understand what aspects of STEM learning activities may influence students' preferences.

RCT alone does not embody intentions and attitudes, nor does it fully explain preferences (Opp, 2019). The closely related theoretical lens of expectancy–value theory (EVT) helps to flesh out how individuals formulate positive or negative associations and/or expectations of future outcomes related to certain learning experiences. EVT is a widely accepted model developed by Eccles, Wigfield, and colleagues (e.g., Eccles et al., 1983; Eccles & Wigfield, 1995; Wigfield, 1994) whereby an individual's expectations of success on a learning task and the value they place on the task are central determinants of their motivation to learn (Choi et al., 2010). These two motivational beliefs—success and value—can influence an individual's motivation and behavior.

Research applying EVT to learning activity preference and performance helps to further shed light on the research aims in our paper. In their examination of components of EVT via random assignment study of 70 mental health patients to assess their performance as they interacted with certain arithmetic learning activities, Choi et al. (2010) found that beliefs of content mastery can predict the degree of performance improvement on challenging cognitive tasks, even to a greater degree than general cognitive ability. These findings are reinforced by the exploratory work of Cooper et al. (2017), who developed a conceptual framework of EVT that is closely aligned with the research aims in this paper concerning learning activity preference. They applied the theoretical lens of EVT to evaluate student perceptions of active learning after participating in active learning in a science course. Interview data collected from 25 first-year biology students who had participated in learning biology content through 40 hours of active learning revealed self-reported changes in the direction of increased engagement and self-efficacy, increased perceived value, and decreased resistance toward active learning. They conclude that by increasing the perception of the value of active learning—by offering a student-centered pedagogy and a variety of different active learning activities, for instance—instructors can help increase student motivation to do well in active learning. Ultimately, the work by Cooper et al. (2017) is also supported by the work of other researchers examining how EVT impacts student perceptions of and preferences for active learning curricula (Schoor &

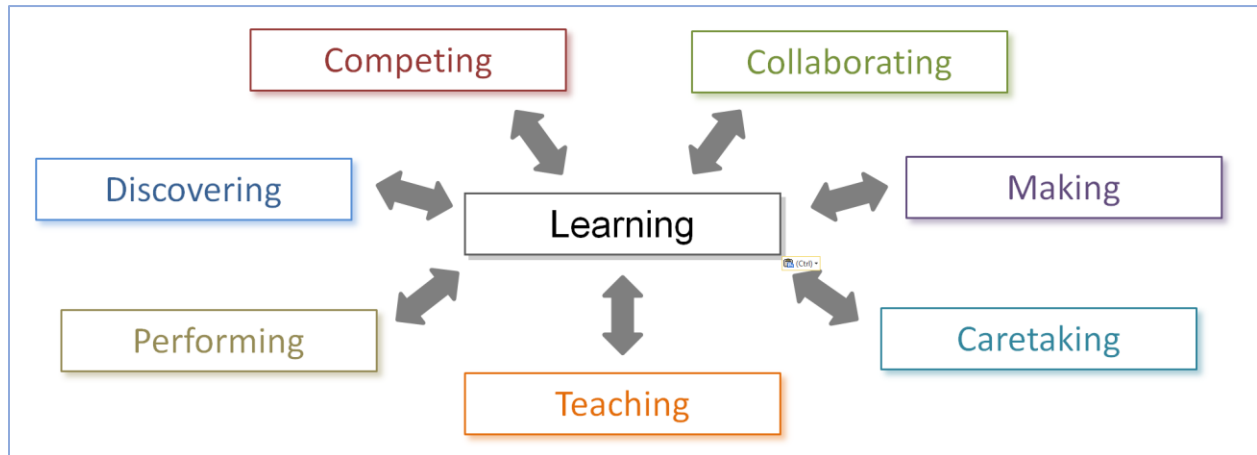
Bannert, 2011; Machemer & Crawford, 2007). These studies suggest that EVT is a valuable lens with which to examine student preferences for different learning modalities and types.

Development of a Conceptual Framework Focused on Active Learning

This theoretical framework distinguishes between two modes of learning: receptive and active. Forms of receptive learning include watching, listening, and reading—receiving information. While youth may ask questions or make comments, the aim of receptive learning is to deliver information. Forms of active learning include making and discovering new things, collaborating, and performing. In this paper, we focus on forms of active learning and how these active learning formats may be used as measures of youth engagement and programmatic impact.

The instruments in Table 1 primarily target students' general attitudes towards science, impressions of the importance of science, specific interests in science, and motivation to learn science. All these measures are important, but the information they provide do not give youth STEM program developers and instructors the type of information necessary for reevaluating program impact on youth STEM engagement and making modifications.

The conceptual framework development began with an examination of a variety of informal STEM programs based in the United States. The programs included robotics competitions, outdoor adventure programs, urban ecology programs, and technology programs, among others. In addition, we examined K-12 formal science curricula to round out STEM education activities. These formal curricula were taken from curriculum guides and texts at the elementary, middle, and high school levels with examples from Project WILD, Project WET, Biological Science Curriculum Study, and National Science Resources Center, etc. From this wide-ranging list of informal programs and formal curricula, we selected examples for a further examination of instructional strategies, which led to the creation of a list of learning activities. As we developed this list, links and similarities in the activities used in the programs and curricula began to emerge. Deconstruction of the activities coupled with the subsequent pilot studies revealed that the instructional strategies appeared as combinations of seven types of active learning: (a) collaborating, (b) competing, (c) making, (d) discovering, (e) performing, (f) caretaking, and (g) teaching/tutoring (see Figure 1). Here we offer a concise discussion of how we define each of these seven active learning modes.

Figure 1. Framework for Observing and Categorizing Instructional Strategies (FOCIS)

Typology of Active Learning

The activity of *collaborating* requires communication, sharing, and comparing thought and ideas among group members working together on a project or task. The form and type of collaboration may vary, but at its core is the requirement that the activity necessitates the cooperation and engagement of two or more individuals. One example is two youth working together to build a mousetrap-propelled model cart. Another example is a group of youth working together on a robotics team. An important characteristic of successful collaborations is clarity among team members of their respective roles and responsibilities. For example, while some youth focus on the mechanics of building the robot, others work on the control software programming, and still other team members study the tasks their robot will be required to do and develop strategies. The degree and duration of communication may vary across different examples.

There are many examples of competing in informal STEM youth programs, which include science fairs, model rocketry, and robotics. *Competing* activities center on the motivating factors that compel participants to seek to win. However, competing does not rely on the setting, but rather on a prearranged set of parameters.

Discovering activities lead participants to learning new things, figuring things out, and problem solving. This type of activity engages participants in learning something new to them. Examples include learning how to use new tools such as a microscope to see and identify the different types of organisms living in a drop of pond water or experimenting with a simple pendulum by

Gauging STEM Youth Program Impact

using different masses, lengths, and release angles to see how changes affect the pendulum's period. In both of these examples, individuals discover new information previously unknown to them.

Making activities apply ideas and materials to constructing an object. Building towers of drinking straws, making bridges from toothpicks, making containers to protect raw eggs to be dropped from different heights, programming a website, and writing code to operate a robot are examples. Making-type activities are strongly associated with discovering-type activities and in many ways, discovering is to making as science is to engineering.

Performing is commonly associated with presentations and audiences but performing at its core involves accountability for an outcome at a specific place and time, meeting a challenge. While presenting to an audience is clearly a form of performance, other forms of performance may not involve an audience. A solitary youth developing her/his skill at starting a campfire by using a bow drill to create an ember is performing. The outcome matters to the youth engaging in the activity. Youth planting gardens to grow vegetables for their families or to donate to food banks is another example.

Caretaking is a type of activity where youth are given responsibility for caring for others, animals, and even objects. Caring for animals, growing a garden, designing a wheelchair ramp, and engaging in trash recycling to help clean up a community park are all forms of caretaking.

Teaching is an activity where youth are engaged in helping others to learn. Teaching involves engaging the mind and attention of another individual or group of individuals. In a teaching activity, the youth in the role of teacher has knowledge or information that they are tasked with passing along to others or has skills he/she helps others to attain.

The Instrument and Its Development

The youth engagement survey is shown in Table 2. This instrument was designed to assess the preferences that youth might have for each of the seven types of active learning. In this paper, we examine validity with respect to the conceptual framework and the survey instrument. These efforts included pilot testing of the survey instrument as well as focus group interviews regarding the conceptual framework and the instrument. Next, we examine the reliability of the survey instrument through a confirmatory factor analysis (CFA) using a large-scale survey of

Gauging STEM Youth Program Impact

7,382 youth in 24 schools across four school districts (urban, suburban, small town, and rural) in two states.

Table 2. Survey Instrument

We want to know how you feel about different activities. (Please check only 1 box for each activity listed below.)	
When I find out that an activity involves . . .	I feel . . . ☹ ☺ 😊
Being in a group	1 2 3 4 5
Being in a competition	1 2 3 4 5
Making or building things	1 2 3 4 5
Discovering and learning new things	1 2 3 4 5
Presenting in front of lots of people	1 2 3 4 5
Taking care of animals	1 2 3 4 5
Helping people learn things	1 2 3 4 5
We want to know what you think about each of the statements below. If you strongly agree, then choose 5. If you strongly disagree, then choose 1.	
<i>(Please circle only 1 number for each statement below)</i>	👉 👈
Working with others is more fun than working alone.	1 2 3 4 5
I like being part of a team.	1 2 3 4 5
I learn better when I am working with others.	1 2 3 4 5
I get excited when I hear there will be a competition.	1 2 3 4 5
I enjoy competing against other people.	1 2 3 4 5
I like to focus on my own goals, rather than competing with others.	1 2 3 4 5
I like figuring out how things work.	1 2 3 4 5
I like taking things apart to see what is inside.	1 2 3 4 5
I like trying different ways to figure things out.	1 2 3 4 5
I like solving problems.	1 2 3 4 5
Helping others to learn things is fun.	1 2 3 4 5
I like teaching things to others.	1 2 3 4 5
Having a pet is a big responsibility, but something I like to do.	1 2 3 4 5
I like to take care of things like plants and aquariums.	1 2 3 4 5
I feel good when people depend on me.	1 2 3 4 5

Gauging STEM Youth Program Impact

Table 2. (continued)

We want to know what you think about each of the statements below. If you strongly agree, then choose 5. If you strongly disagree, then choose 1.	
<i>(Please circle only 1 number for each statement below)</i>	☹ ☺
Performing in front of people is fun.	1 2 3 4 5
I like telling people about my work.	1 2 3 4 5
I like presenting my work to my class.	1 2 3 4 5
I like doing projects where I make things.	1 2 3 4 5
Whenever I can, I make the things I need.	1 2 3 4 5
I like building things.	1 2 3 4 5

Examining Construct, Consequential, and Content Validity

To establish instrument validity, we used focus groups, pilot surveys, and educational policy document analysis. The piloting efforts included 20 pilot surveys administered in 17 out-of-school time program settings and three formal classrooms across Grades 4, 7, and 8. Focus group discussions with youth participants were held in four of the 20 pilot surveys. The focus groups were chosen to include a range of different ages. In addition, focus group discussions with informal educators and teachers were conducted. In the focus groups, we typically began by gathering feedback on the survey questions by asking the participants about their observations about the survey. We discussed with youth their perspectives on the responses they gave to the various questions and with educators their observations about the conceptual framework and its potential use in their educational practice.

Based on the discussion of validity in the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2011, pp. 11-17) and other sources (e.g., Gall et al., 1996; Messick, 1994), we address three facets of validity: *construct validity*, *content validity*, and *consequential validity*.

Construct validity concerns the degree to which the survey instrument measures differences in learning activity preferences. This fundamental characteristic of validity concerns the question of whether individuals acknowledge that there exist different types of learning activities and that they assign different preferences to these types of learning activities. To this end, we carried out a pilot survey and focus group discussion with a fourth-grade class of 17 students. After completing the survey, students were asked about their responses and their opinions

Gauging STEM Youth Program Impact

about the questions. When the researcher asked students in a focus group what they thought the survey was about, a student responded, "It seemed like you [researcher] wanted to know what we like to do." When asked what types of activities they like to do and what activities they did not like to do, student responses included: "I don't like competitions." "I like working in groups." "I really don't like talking in front of the class." In one particular focus group, a student noted that the survey questions were redundant, i.e., the same question appeared to be asked more than once and asked, "Was it because you [researcher] wanted to be sure we were really thinking about our answers?" It appeared from this interaction that this student understood that the survey questions were intended to check for consistency in the students' survey responses. Our analysis of the survey responses from the class revealed that students' responses were self-consistent across the items within each learning activity type. Subsequent focus group discussions and pilot surveys of youth across ages produced comparable outcomes. These findings suggest that the conceptual framework and the survey instrument possess a reasonable degree of construct validity.

Both *consequential* and *content validity* were ascertained by focus group discussions with teachers and informal educators. Consequential validity considers how the conceptual framework and survey instrument may provide insights into program and curriculum development or links with other educational ideas, activities, and policies. The conceptual framework, survey instrument, and survey responses were presented to educators at a number of different venues that included invited talks (e.g., IPN, Leibniz Institute for Science and Mathematics Education at University of Kiel in 2015; NIH 2013 Science Education Conference; Virginia Governor's Schools Annual Conference in 2015) and teacher workshops (e.g., CRISP 2015 Professional Development Workshop at Yale University; MRL 2014 Science Teacher Workshop at UC Santa Barbara). These professional presentations were before audiences of informal educators, teachers, and educational researchers. The invited talks and teacher workshops included a presentation of the conceptual framework, the survey instrument, and the results of surveys carried out over the course of the past 4 years. The survey instrument was distributed to the audiences (including question-and-answer discussion sessions) and the researcher fielded questions from the audience regarding a variety of topics. Discussions during the post-presentation question-and-answer sessions highlighted important intersections with educational policies in the United States. The Partnership for 21st Century Skills (2011) identifies learning and innovation skills that include "creativity and innovation," "critical thinking and problem solving," and "communication and collaboration." These three skills overlap directly with the learning activities of making, discovering, and collaborating, respectively.

Gauging STEM Youth Program Impact

At the outset, this work began with the need for a means of measuring youth engagement in informal STEM education programming that offered information relevant to informal program development and effectiveness. The design of this instrument offers information to out-of-school time program developers based on the responses of youth participants to informal program characteristics specific to each program. By implementing pre-program/post-program surveys and examining changes in the participants' responses, informal program developers will have some measure of how different facets of their programs might be influencing youth participants.

Content validity was examined by engaging with STEM educators through invited talks and workshops. The researchers engaged with informal program providers, schoolteachers, school principals, and educational researchers. The interactions during the presentations and the discussions that followed offered insights into the perspectives of other educators and helped the researchers to refine both the conceptual framework and survey instrument as well as establish content validity of both the conceptual framework and the survey instrument. These discussions led to more expansive definitions of "performing" and "caretaking" than had originally been considered.

Examining Reliability, Discriminant Validity, and Convergent Validity

In this next section, we consider the technical performance of the survey instrument in terms of reliability as well as discriminant validity and convergent validity through the application of CFA. We explore the quantitative performance of the survey instrument with respect to reliability as well as discriminant and convergent validity of the seven-factor conceptual framework through its fit onto the data collected from students across Grades 3 through 12.

Participants

The data in this study were obtained from a survey administered in the fall of 2012 asking for participants' opinions about many statements regarding their preferences for learning activities falling within each of the seven types described above. Four participating school districts spanned the range of urban, suburban, and rural areas. The survey included students from Grades 3 through 12 and had robust enrollments of Black and Hispanic students. At the teacher's discretion, younger students were read the survey questions. All students in attendance within each of the 24 participating schools were given surveys. The total number of participants was 7,382 students with 33.7%, 33.9%, and 29.4% at the elementary, middle, and high school grade levels, respectively. Females constituted 48.9% of participants. Participants

Gauging STEM Youth Program Impact

reported their racial/ethnic demographic as 46.1% White; 16.5% Black; 16.4% Hispanic; 1.9% Asian and Pacific Islander; and 0.5% Native American/American Indian. Additionally, 14.5% reported having multiple racial/ethnic identities, and 4.1% did not specify their racial/ethnic demographic.

Measures

The seven types of active learning were considered as seven latent factors in the analysis (Figure 1). For each of them, we developed three to five statements with 5-point Likert scale response options (*strongly disagree*, *disagree*, *neutral*, *agree*, and *strongly agree*) in the survey. For the questions with only two response options, the responses were assigned values of 2 and 4 on the scale of 1 to 5. Negative statements were reverse-coded to maintain conceptual and statistical consistency with the other variables. Table 3 has a description of the survey items and variable names for the statistical analysis to follow.

Table 3. Descriptions of 28 Variables in Factor Analysis

Factor	Variable name	Corresponding survey item
Collaborating	Col1	I like an activity that involves "Being in a group".
	Col2	Working with others is more fun than working alone.
	Col3	I like being part of a team.
	Col4	I learn better when I am working with others.
Competing	Com1	I like an activity that involves "Being in a competition".
	Com2	I get excited when I hear there will be a competition.
	Com3	I enjoy competing against other people.
	Com4	I like to focus on my own goals, rather than competing with others.
Making	Mak1	I like an activity that involves "Making or building things".
	Mak2	I like doing projects where I make things.
	Mak3	Whenever I can, I make the things I need.
	Mak4	I like building things.
Discovering	Dis1	I like an activity that involves "Discovering and learning new things".
	Dis2	I like figuring out how things work.
	Dis3	I like taking things apart to see what is inside.
	Dis4	I like trying different ways to figure things out.
	Dis5	I like solving problems.

Table 3. (continued)

Factor	Variable name	Corresponding survey item
Performing	Pre1	I like an activity that involves "Presenting in front of lots of people".
	Pre2	Performing in front of other people is fun.
	Pre3	I like telling people about my work.
	Pre4	I like presenting my work to my class.
Caretaking	Car1	I like an activity that involves "Taking care of animals".
	Car2	Having a pet is a big responsibility, but something I like to do.
	Car3	I like to take care of things like plants and aquariums.
Teaching	Tea1	I like an activity that involves "Helping people learn things".
	Tea2	Helping others to learn things is fun for me.
	Tea3	I like teaching things to others.
	Tea4	I feel good when people depend on me.

Data Analysis

We used CFA to investigate the robustness of the measure of the seven latent constructs described in the conceptual framework. In preparation, we examined the data for possible outliers and for multicollinearity among 28 indicators of the seven factors. Cook's distance (D) was used to identify any potential outliers in the data at the significance level of 0.05 in $F_{(k, n-k)}$, where k is the number of variables and n is the sample size; we did not find a multivariate outlier in $F_{(28, 7354)}$. In terms of multicollinearity, we computed variance inflation factors (VIFs) for 28 indicators and applied the convention that only variables with VIF value greater than 10 indicate multicollinearity (Neter et al., 1989). In this study, the VIF values for all 28 variables were between 1.21 and 3.27 with a mean of 2.03, supporting the conclusion that multicollinearity was not a concern.

To begin our CFA, we set our fit parameters according to the effects coding method of scaling (Little et al., 2006). These fit parameters were then used to determine the following criteria for *good* fit for the approximate fit indices: root mean square error of approximation (RMSEA) < .05, comparative fit index (CFI) > .95, and standardized root mean square residual (SRMR) < .08 (Hu & Bentler, 1999); and *adequate* (or *acceptable*) fit: .05 < RMSEA < .08 (Browne & Cudeck, 1993) and .95 > CFI > .90 (Bentler, 1990). According to convention we also reported

Gauging STEM Youth Program Impact

chi-square values, where a value greater than 0.05 indicated that we retain the null hypothesis, which provided evidence that the data adequately fit the hypothesized model.

Results and Discussion

A seven-factor model was fitted to the data using maximum likelihood estimation which included the following latent constructs and corresponding observed variables: collaborating (four indicators), competing (four indicators), making (four indicators), discovering (five indicators), performing (four indicators), caretaking (three indicators), and teaching (four indicators). RMSEA (.043, 90% CI = [.042, .044]) and SRMR (.055) values indicated a good fit, and the CFI (.931) value indicated an acceptable fit. These results serve as evidence that the seven-factor model was well fitted onto the data.

Details for calculating model fit and reliability measures are provided in the Appendix. Table A1 summarizes unstandardized and standardized estimates of parameters including their corresponding significance values. Factor loadings of indicators were greater than 0.45 and significant at the level of .05 in the CFA using a fixed factor model, which indicates the construct validation along with a good fit of the seven-factor model.

To examine the convergent validity and the factor reliability coefficients, we fitted an equivalent CFA model with the marker variable model fixing the first factor loading as 1 onto the data. We used Raykov's rho as a measure of factor reliability (see Appendix for details on computing this statistic). Table A2 provides a summary of the factor reliability measure findings, which show that the collaborating (0.837), competing (0.840), making (0.828), performing (0.848), teaching (0.820), discovering (0.753) and caretaking (0.763) constructs were reliable at a criterion of 0.7 (Nunnally & Bernstein, 1994). These results indicate that all of the seven constructs displayed convergent validity. The factor reliability measures reported here are comparable to the more traditional Cronbach's α and are a more appropriate measure for confirmatory factor analysis (Brown, 2006). The factor correlations shown in Table A2 were all less than .80, further supporting the discriminant validity of the seven factors (Brown, 2006).

Limitations of the Statistical Analysis

There are some important limitations to note in this analysis. First, the sample of 7,382 included students from a variety of schools across four school districts. While we have argued that this sample is representative of the population, it was not obtained through stratified or randomized

Gauging STEM Youth Program Impact

sampling. This said, the sampling did include all students present in each participating school, with negligible numbers declining to participate. Second, use of Likert-type summated rating scales may be somewhat subjective to survey respondents and raises questions of reliability from survey to survey. We have addressed this concern by using multiple items to triangulate responses for each learning activity type. Considering these limitations, the confirmatory factor analysis results show that instrument reliability, discriminant validity, and convergent validity standards have all been met.

Another important limitation is the impact of confounding learning experiences that youth may have outside of structured, informal STEM programs. Some youth have family and friends who routinely engage them in STEM-related activities outside of structured programs. The instrument has no means of gathering this kind of information. Informal STEM educators must rely on alternative means to assess the impact of these factors. However, one important caveat should be kept in mind: Youth with this type of experience and mentorship are rare, and given that the results are viewed in aggregate, the impact of a few youths with these types of experiences will likely have an exceedingly small, if not negligible, impact on the aggregated analysis.

An Example for Practitioners: Applying the Conceptual Framework and Survey Instrument

One of the authors [Tai] uses the instrument to gather information about the impact of his course on elementary science teaching methods each academic year. The analysis of the data from one of these courses is used as an example here. This course enrolled 33 students, 29 of whom provided complete survey data. The four students not included in this analysis were late enrollees. Only the student responses including both pre and post surveys were analyzed. The semester-long course typically met twice weekly for a total of 28 classes, each lasting 75 minutes. The students in the course were a mix of pre-service elementary teacher candidates ($n = 21$) and undergraduate students ($n = 8$) enrolled in other schools in the university who were majoring in disciplines such as biology, urban planning, history, mechanical engineering, psychology, sociology, anthropology, and mathematics. All students were asked to complete the survey instrument on the 1st day of class and again on the last day of class.

We began by surveying the students in the course to determine their active learning preferences. Table 4 includes the pre-course average preference values across the seven different types of active learning. How might a course instructor use the pre-course average

Gauging STEM Youth Program Impact

preference values? The scales for each of the preference scores range from 1 (*highly negative preference*) to 5 (*highly positive preference*), with a score of 3 indicating *no preference*. On the lower end of the preference values, the average initial preference score for competing was 3.23 with a standard deviation of 0.80 indicating that the students had nearly no preference for engaging in competitions on the average. On the high end, the initial preference score for teaching was 4.87 ($sd = 0.37$). This result indicates that activities involving competition should be approached with care, since it appears that many students in the class are either ambivalent or report negative preferences for competing-type activities. To take this one step further, the individual student responses may be examined to look for the specific students who report strong negative preference scores. As a result, an instructor would have a heads-up regarding which students might potentially struggle with a particular assignment. It is important to note that active learning type preferences do not carry with them a negative stigma. Not preferring to compete or not preferring to collaborate is a status in a particular student's preferences which engagement in learning has the potential to change.

The teaching preference score was 4.87—nearly at the top of the scale. The aim is for the course activities to maintain this high preference score while still engaging the students in learning new content.

Table 4. Class Averages and Standard Deviations for the Seven Active Learning Modes

Learning activity type	Pre-course average (<i>SD</i>)	Post-course average (<i>SD</i>)	Difference (Post - Pre)	<i>p</i> -value for paired two-tailed <i>T</i> -test
Collaborating	3.73 (0.83)	4.04 (0.88)	0.31	.002**
Competing	3.23 (0.80)	3.49 (0.99)	0.26	.123
Making	3.91 (0.78)	3.94 (0.90)	0.03	.758
Discovering	4.26 (0.63)	4.24 (0.67)	-0.02	.807
Performing	3.42 (0.86)	3.37 (0.93)	-0.05	.732
Caretaking	4.17 (0.72)	4.22 (0.61)	0.05	.595
Teaching	4.87 (0.37)	4.80 (0.26)	-0.07	.312

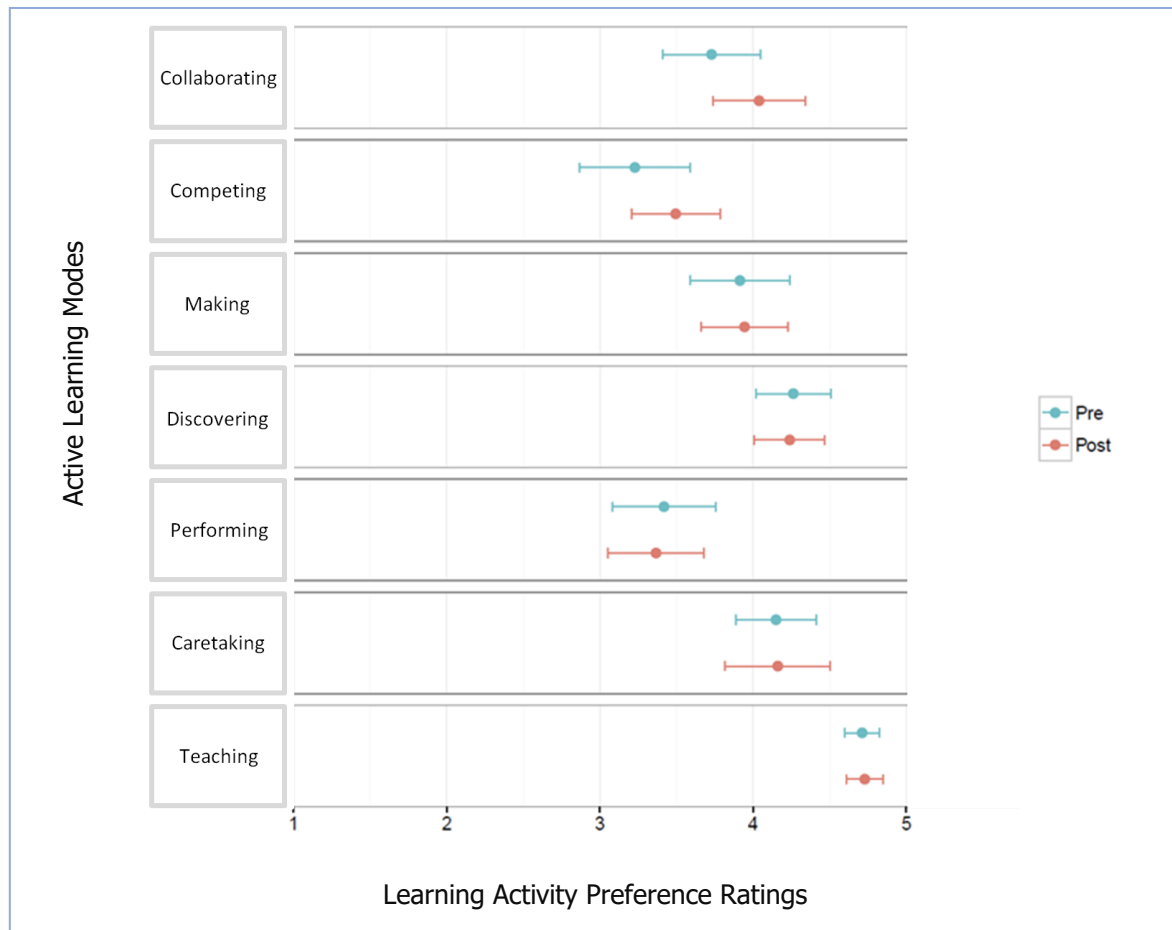
** Significant at $p < 0.01$

Gauging STEM Youth Program Impact

Next, let us turn to the course activities. On the final day of the semester, students were given the same survey again to measure their post-course active learning preferences. The course averages for the preference scores are reported in Table 4 and compared directly in Figure 2.

The conceptual framework discussed earlier in this report is used to examine six course activities shown in Table 5. Not all course activities were included in this discussion. These course activities were designed to be used with college students to engage them in firsthand active learning experiences but may be modified easily for K-6 learners. The students' active learning experiences with these lessons were followed by class discussions aimed at engaging them with the pedagogical choices made by the instructor. Note that all seven active learning types are applied in the course activities in various combinations.

Figure 2. Graph of the Class Means for the Seven Learning Activity Types with Error Bars Displaying Standard Errors



Gauging STEM Youth Program Impact

Table 5. Summary of the Characteristic Learning Activity Types of the Six Course Activities

Learning Activity	Description	Collaborating	Competing	Making	Discovering	Performing	Caretaking	Teaching
Bean Plants	Required students to grow beans from seeds until these seeds mature into plants that produce seeds themselves.				●	●	●	
Slime and Silly Putty	Required students to explore the physical characteristics of polymers created when two types of household glues are mixed with liquid starch.	●		●	●			
Hoverdisk	Required students to pull their teammates who would be riding on a hoverdisk over a predetermined path. A hoverdisk is a miniature hovercraft powered by a battery-operated leaf blower.	●	●			●		
Rube Goldberg Contraption	Required student teams to construct an overly complicated "device" with at least 6 transfers of energy using a combination of simple machines to water a plant.	●	●	●		●		●
Stomp Rockets	Required students working in teams with a given a set of materials to build and launch an air-propelled "rocket."	●		●	●	●		
Egg Launch	A variation on the ubiquitous egg drop competitions, students were required to design and construct a container to house a raw egg to be launched from a catapult.	●	●	●	●	●	●	

Gauging STEM Youth Program Impact

The results for collaborating show a significant positive shift in students learning activity average preference. We can see in Table 5 that collaborating is included in five of six course activities. This result suggests that the course activities had a positive impact on students' average preference for engaging in collaborations, while the results also showed that six of the seven active learning preferences average scores (competing, making, discovering, performing, caretaking, and teaching) were static according to the *t*-test analysis used to compare pre-post course averages also shown in Table 4. To better understand these results, we take a closer look at the average scores themselves. Note that for teaching, discovering, and caretaking, the preference scores were high (i.e., above 4) and they remained high. This result indicates that the course activities were successful in maintaining students' overall positive preferences for teaching, discovering, and caretaking. For making, the average preference score was below 4 and remained below 4 which suggests that more work needs to be done to engage students in making-type activities. The characteristics identified in Table 5 indicate that making was included in four course activities: Slime and Silly Putty, Rube Goldberg, Stomp Rockets, and Egg Launch. Efforts to improve the making-type activities would be concentrated on these four course activities.

The weakest results were for performing and competing. The average preference scores were below 3.5. While the performing score remained nearly unchanged, the competing score average did show some improvement. These results suggest that efforts to improve engagement in performing-type activities should concentrate on the following course activities: Bean Plants, Hoverdisk, Rube Goldberg, Stomp Rockets, and Egg Launch. As for competing, the relevant course activities were: Hoverdisk, Rube Goldberg, and Egg Launch. Note that this analysis reveals that participants' two least preferred activity types (competing and performing) intersect in both the Rube Goldberg and Egg Launch activities. This result indicates that these two activities should be the ones that require the most scrutiny. This analysis resulted in these two course activities being modified with Rube Goldberg no longer engaging students in competing, but rather focusing on performing, while the students were given more time to complete the Egg Launch activity as well as having a class session devoted to testing their initial designs.

Conclusion

In this report, we have described a conceptual framework and survey instrument that can be useful to informal and formal STEM educators in evaluating and understanding the impact that their instructional programs have on the youth science learning engagement. The conceptual

Gauging STEM Youth Program Impact

framework can be used to analyze program composition and structure. Pre-program surveys can identify the types of active learning that participants prefer or have reservations about. This information provides some actionable information for program facilitators and instructors. Youth who express strong positive preferences for active learning types included in a program have a clear path toward engagement and are likely to be highly engaged. On the other hand, youth who have strong reservations (negative preferences) are the ones to watch. They are the ones more likely to become disengaged during program activities they don't want to do. This information allows program facilitators and instructors to focus their attention on youth most likely to fall through the cracks.

After a program has been completed, post-program survey outcomes can be compared with pre-program survey outcomes to offer program facilitators and instructors insight into how their programs may have changed participating youths' preferences. Specific types of active learning may be shown in these pre-post survey comparisons to have a negative impact. This result, when coupled with the program analysis, will provide program facilitators and instructors information about which program activities may require reevaluation and modification.

Often informal STEM program evaluations include general science attitude questions, for example, "Science is helpful in understanding today's world" and "Science is something I enjoy very much" (Weinburgh & Steele, 2000); "I am sure of myself when I do science" and "I know I can do well in science" (Unfried et al., 2015). The responses to these questions offer some general participant impressions, but no cues from participants on where improvements might be made. When asked for suggestions on how to improve programs, many times participants are hesitant, concerned that their responses might be misconstrued as criticisms. The approach here offers a pedagogically focused means of analyzing programmatic impact on participant engagement with enough detail to allow for more focused programmatic reevaluation and improvement.

References

- Adams, W. K., Perkins, K. K., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2004). *The design and validation of the Colorado learning attitudes about science survey*. Physics Education Research Conference. <http://www.colorado.edu/sei/class/>
- Aschbacher, P. R., Li, E., & Roth, E. J. (2010). Is science me? High school students' identities, participation, and aspiration in science, engineering, and medicine. *Journal of Research in Science Teaching*, 47(5), 564-582. <https://doi.org/10.1002/tea.20353>

Gauging STEM Youth Program Impact

- Bachman, N., Bischoff, P. J., Gallagher, H., Labroo, S., & Schaumloffel, J. C. (2008). PR²EPS: Preparation, recruitment, retention and excellence in the physical sciences, including engineering. A Report on the 2004, 2005 and 2006 science summer camps. *Journal of STEM Education: Innovations & Research*, 9(1/2), 30-39.
- Barmby, P., Kind, P. M., & Jones, K. (2008). Examining changing attitudes in secondary school science. *International Journal of Science Education*, 30(8), 1075-1093.
<https://doi.org/10.1080/09500690701344966>
- Barman, C.R. (1999). Students' views about scientists and school science: Engaging K-8 teachers in a national study. *Journal of Science Teacher Education*, 10(1), 43-54.
<http://www.jstor.org/stable/43156207>
- Barton, A. C., Kang, H., Tan, E., O'Neill, T. B., Bautista-Guerra, J., & Brecklin, C. (2013). Crafting a future in science: Tracing middle school girls' identity work overtime and space. *American Educational Research Journal*, 50(1), 37-75. <https://doi.org/10.3102/0002831212458142>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
<https://doi.org/10.1037/0033-2909.107.2.238>
- Bhattacharyya, S., & Mead, T. P. (2011). The influence of science summer camp on African American high school students' career choices. *School Science & Mathematics*, 111(7), 345-353.
<https://doi.org/10.1111/j.1949-8594.2011.00097.x>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Sage.
- Carnevale, A. P., Smith, N., & Strohl, J. (2010). *Help wanted: Projections of jobs and education requirements through 2018*. Georgetown University, Center on Education and the Workforce.
- Chang, S., Yeung, Y., & Cheng, M. H. (2009). Ninth graders' learning interests, life experiences and attitudes towards science and technology. *Journal of Science Education and Technology*, 18, 447-457. <https://doi.org/10.1007/s10956-009-9162-6>
- Choi, J., Fiszdon, J. M., & Medalia, A. (2010). Expectancy-value theory in persistence of learning effects in schizophrenia: Role of task value and perceived competency. *Schizophrenia Bulletin*, 36(5), 957-965. <https://doi.org/10.1093/schbul/sbq078>
- Coleman, J. S., & Fararo, T. J. (Eds.). (1992). *Rational choice theory advocacy and critique*. Sage.
- Cooper, K. M., Ashley, M., & Brownell, S. E. (2017). A Bridge to active learning: A summer bridge program helps students maximize their active-learning experiences and the active-learning experiences of others. *CBE – Life Science Education*, 16(1). <https://doi.org/10.1187/cbe.16-05-0161>
- Deitrich, F., & List, C. (2013). A reason-based theory of rational choice. *Noûs*. 47(1), 104-134.
<https://doi.org/10.1111/j.1468-0068.2011.00840.x>

Gauging STEM Youth Program Impact

- Ebenezer, J.V., & Zoller, U. (1993). Grade 10 students' perceptions of and attitudes toward science teaching and school science. *Journal of Research in Science Teaching*, 30(2), 175-186. <https://doi.org/10.1002/tea.3660300205>
- Eccles J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). W. H. Freeman.
- Eccles, J. S., & Wigfield, A. (1995). In the mind of the achiever: The structure of adolescents' academic achievement related beliefs and self-perceptions. *Personality and Social Psychology Bulletin*, 21, 215-225. <https://doi.org/10.1177/0146167295213003>
- Elam, M. E., Donham, B. L., & Solomon, S. R. (2012). An engineering summer program for underrepresented students from rural school districts. *Journal of STEM Education: Innovations and Research*, 13(2), 35-44.
- Farland-Smith, D. (2009). Exploring middle school girls' science identities: Examining attitudes and perceptions of scientists when working "side-by-side" with scientists. *School Science and Mathematics*, 109(7), 415-427. <https://doi.org/10.1111/j.1949-8594.2009.tb17872.x>
- Ferreira, M.M., & Trudel, A.R. (2012). The impact of problem-based learning on student attitudes toward science, problem-solving skills, and sense of community in the classroom. *Journal of Classroom Interaction*, 47(1), 23-30. <https://www.jstor.org/stable/43858871>
- Fields, D. (2009). What do students gain from a week at science camp? Youth perceptions and the design of an immersive, research-oriented astronomy camp. *International Journal of Science Education*, 31(2), 151-171. <https://doi.org/10.1080/09500690701648291>
- Fraser, B. J. (1978). Development of a test of science-related attitudes. *Science Education*, 62, 509-515. <https://doi.org/10.1002/sce.3730620411>
- Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). Longman Publishing.
- George, R. (2006). A cross-domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, 28(6), 571-589. <https://doi.org/10.1080/09500690500338755>
- 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. <https://doi.org/10.1002/tea.3660250807>
- Gibson, C., & Chase, H. L. (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. *Science Education*, 86(50), 693-705. <https://doi.org/10.1002/sce.10039>

Gauging STEM Youth Program Impact

- Guzzetti, B. J., & Bang, E. (2011). The influence of literacy-based science instruction on adolescents' interest, participation, and achievement in science. *Literacy Research and Instruction, 50*, 44-67. <https://doi.org/10.1080/19388070903447774>
- Häussler, P., & Hoffmann, L. (2000). A curricular frame for physics education: Development, comparison with students' interests, and impact on students' achievement and self-concept. *Science Education, 84*(6), 689-705. [https://doi.org/10.1002/1098-237X\(200011\)84:6<689::AID-SCE1>3.0.CO;2-L](https://doi.org/10.1002/1098-237X(200011)84:6<689::AID-SCE1>3.0.CO;2-L)
- Homans, G. C. (1961). *Social behavior: Its elementary focus*. Harcourt.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Hulleman, C. S., & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*, 1410-1412. <https://doi.org/10.1126/science.1177067>
- Johnson, R. C. (2011). Using summer research to attract pre-college underrepresented students to STEM fields. *Council on Undergraduate Research Quarterly, 31*(3), 7-15.
- Kind, P. M., Jones, K., & Barmby, P. (2007). Developing attitudes toward science measures. *International Journal of Science Education, 29*(7), 871-893. <https://doi.org/10.1080/09500690600909091>
- Koballa, Jr., T. R., & Glynn, S. M. (2007). Attitudinal and motivational constructs in science learning. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research on science education* (pp. 75-102). Lawrence Erlbaum Associates.
- Krajcik, J., Czerniak, C., and Berger, C. (2003). *Teaching children science: A project-based approach* (2nd ed.). McGraw-Hill.
- Lent, R. W., Brown, S. D., & Gore, P. A. (1997). Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy. *Journal of Counseling Psychology, 44*, 307-315. <https://doi.org/10.1037/0022-0167.44.3.307>
- Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling, 13*, 59-72. https://doi.org/10.1207/s15328007sem1301_3
- Loveland, M. T. (2003). Religious switching: Preference development, maintenance, and change. *Journal for the Scientific Study of Religion, 42*(1), 147-157. <https://doi.org/10.1111/1468-5906.00168>
- Luce, M.R., & Hsi, S. (2014). Science-relevant curiosity expression and interest in science: An exploratory study. *Science Education, 99*(1), 70-97. <https://doi.org/10.1002/sce.21144>
- Machemer, P. L., & Crawford, P. (2007). Student perceptions of active learning in a large cross-disciplinary classroom. *Active Learning in Higher Education, 8*(1), 9-30. <https://doi.org/10.1177/1469787407074008>

Gauging STEM Youth Program Impact

- Majumdar, S. K., Rosenfeld, L. M., Rubba, P. A., Miller, E. W., & Schmalz, R. F. (Eds.). (1991). *Science education in the United States: Issues, crises, and priorities* (pp. 92-105). The Pennsylvania Academy of Science.
- Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment, 10*(1), 1-9.
- Moore, R. W., & Foy, R. (1997). The scientific attitude inventory: A revision (SAI II). *Journal of Research in Science Teaching, 34*(4), 327-36. [https://doi.org/10.1002/\(SICI\)1098-2736\(199704\)34:4<327::AID-TEA3>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1098-2736(199704)34:4<327::AID-TEA3>3.0.CO;2-T)
- Myers, R. E., & Fouts, J. T. (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science Teaching, 29*(9), 929-937. <https://doi.org/10.1002/tea.3660290904>
- National Academy of Sciences. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter future*. The National Academies Press.
- Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Irwin.
- Newell, A.D., Zientek, L.R., Tharp, B.Z., Vogt, G.L, & Moreno, N.P. (2015). Students' attitudes toward science as predictors of gains on student content knowledge: Benefits of an after-school program. *School Science and Mathematics, 115*(5), 216-225. <https://doi.org/10.1111/ssm.12125>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Opp, K. (2019). *The rationality of political protest: A comparative analysis of rational choice theory*. Routledge.
- Osborne, J., Simon, S., & Collin, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education, 25*(9), 1049-1079. <https://doi.org/10.1080/0950069032000032199>
- Robbins, M. E., Schoenfisch, M. H., Moore, J. T., & Tolar, D. (2005). An interactive analytical chemistry summer camp for middle school girls. *Journal of Chemical Education, 82*(10), 1486-1488.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*(3 & 4), 299-323. https://doi.org/10.1207/s15326985ep2603&4_5
- Schoor, C., & Bannert, M. (2011). Motivation in a computer-supported collaborative learning scenario and its impact on learning activities and knowledge acquisition. *Learning and Instruction, 21*(4), 560-573. <https://doi.org/10.1016/j.learninstruc.2010.11.002>
- Scott, J. (2000). Rational choice theory. In G. Browning, A. Halcli, & F. Webster (Eds.), *Understanding contemporary society: Theories of the present* (pp. 126-138). Sage.
- Simpson, R. D., Koballa, Jr., T. R., Oliver, J. S., Crawley, F. E. (1994). Research on the affective dimension of science learning. In D. Gabel (Ed.) *Handbook of research on science teaching and learning* (pp. 211-234). Macmillan.

Gauging STEM Youth Program Impact

- Sherkat, W. E., & Wilson, J. (1995). Preferences, constraints, and choices in religious markets: An examination of religious switching and apostasy. *Social Forces*, 73(3), 993-1026.
<https://doi.org/10.2307/2580555>
- Swarat, S., Ortony, A., & Revelle, W. (2012). Activity matters: Understanding student interest in school science. *Journal of Research in Science Teaching*, 49(4), 515-537.
<https://doi.org/10.1002/tea.21010>
- Talisayon, V. M., Mulemwa, J., Corrigan, D., Mehta, J., Jenkins, E., Koulaidis, V., Goel, V., Aikenhead, G., & Ogawa, M. (2004). *Relevance of science education*.
<http://www.ils.uio.no/english/rose/about/rose-brief.html>
- Talton, E. L., & Simpson, R. D. (1986). Relationships of attitude towards self, family, and school with attitude toward science among adolescents. *Science Education*, 70(4), 365-374.
<https://doi.org/10.1002/sci.3730700403>
- Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment*, 33(7) 622-639.
<https://doi.org/10.1177/0734282915571160>
- 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, 87-98. <https://doi.org/10.1615/JWomenMinorScienEng.v6.i1.50>
- Welch, A. G. (2010). Using the TOSRA to assess high school students' attitudes toward science after competing in the FIRST robotics competition: An exploratory study. *Eurasia Journal of Mathematics, Science & Technology Education*, 6(3), 187-197.
<https://doi.org/10.12973/ejmste/75239>
- Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6(1), 49-78. <https://doi.org/10.1007/BF02209024>
- Witteck, R. (2013). Rational choice theory. In *Theory in social and cultural anthropology: An encyclopedia* (pp. 686-689). Sage.

Gauging STEM Youth Program Impact

Appendix

Seven-Factor Model Fit and Reliability Measures

Table A1. Unstandardized and Standardized Parameter Estimates With Their Standard Errors (SE), Error Variances for Indicators, and R-Squared Values Measured in the Seven-Factor Model

Parameter	Unstandardized	SE	Standardized	SE	Error variance	R-squared
Collaborating by						
Col1	1.000		0.773	0.008	0.496	.597
Col2	1.177	0.020	0.799	0.009	0.578	.638
Col3	0.973	0.020	0.742	0.010	0.566	.551
Col4	1.107	0.021	0.699	0.009	0.945	.488
Competing by						
Com1	1.000		0.834	0.006	0.547	.695
Com2	1.163	0.014	0.886	0.006	0.460	.786
Com3	1.150	0.014	0.867	0.006	0.544	.752
Com4	0.406	0.016	0.326	0.012	1.734	.106
Making by						
Mak1	1.000		0.753	0.009	0.568	.567
Mak2	1.045	0.020	0.756	0.009	0.609	.571
Mak3	0.984	0.023	0.627	0.010	1.111	.393
Mak4	1.243	0.018	0.827	0.007	0.532	.684
Discovering by						
Dis1	1.000		0.556	0.011	0.784	.309
Dis2	1.307	0.036	0.648	0.011	0.827	.420
Dis3	1.237	0.045	0.528	0.012	1.390	.279
Dis4	1.494	0.040	0.718	0.009	0.735	.516
Dis5	1.446	0.037	0.632	0.010	1.100	.400
Performing by						
Per1	1.000		0.795	0.007	0.779	.632
Per2	1.028	0.013	0.776	0.007	0.935	.602
Per3	0.731	0.016	0.608	0.010	1.218	.370
Per4	1.115	0.014	0.846	0.006	0.661	.716
Caretaking by						
Car1	1.000		0.809	0.010	0.446	.654

Gauging STEM Youth Program Impact

Car2	0.960	0.018	0.752	0.010	0.598	.565
Car3	0.963	0.026	0.629	0.012	1.196	.396
Teaching by						
Tea1	1.000		0.765	0.008	0.568	.585
Tea2	1.252	0.018	0.846	0.007	0.498	.716
Tea3	1.268	0.019	0.823	0.007	0.615	.677
Tea4	0.558	0.019	0.409	0.013	1.237	.168

Factor Reliability Calculation: Raykov's Rho

The factor reliability was computed using the following formula originally given by Raykov (1997, 2004) with notations in Klein (2011):

$$\hat{\rho} = \frac{\left(\sum_i \hat{\lambda}_i \right)^2 \hat{\phi}}{\left(\sum_i \hat{\lambda}_i \right)^2 \hat{\phi} + \sum_i \hat{\theta}_{ii}}$$

$\sum_i \hat{\lambda}_i$ is the sum of the estimated unstandardized factor loadings among indicators of the same factor (Table A1), $\hat{\phi}$ is the estimated factor variance (Table A2), and $\sum_i \hat{\theta}_{ii}$ is the sum of the unstandardized error variances of those indicators. The factor reliability measure findings are summarized in Table A2.

Table A2. Factor Variances/Covariances, and Reliabilities of the Seven Factors

	Factor variance	Factor reliability	Factor correlation
Collaborating with	0.735	0.837	
Competing			0.191
Making			0.119
Discovering			0.082
Performing			0.178
Caretaking			0.149
Teaching			0.273
Competing with	1.246	0.840	
Making			0.240

Gauging STEM Youth Program Impact

Discovering		0.153
Performing		0.507
Caretaking		0.043
Teaching		0.140
Making with	0.743	0.828
Discovering		0.361
Performing		0.385
Caretaking		0.306
Teaching		0.274
Discovering with	0.351	0.753
Performing		0.311
Caretaking		0.185
Teaching		0.309
Performing with	1.338	0.848
Caretaking		0.231
Teaching		0.517
Caretaking with	0.844	0.763
Teaching		0.341
Teaching	0.800	0.820