

Evaluating University-Based Summer STEM Programs: Challenges, Successes, and Lessons Learned

Christopher J. Cappelli¹, Katherine L. Boice¹, and Meltem Alemdar¹

¹Center for Education Integrating Science, Mathematics, and Computing, Georgia Institute of Technology

Keywords: STEM, Out-Of-School Time Activities, Summer Camps, Evaluation

Publication Date: July 23, 2019

DOI: <https://doi.org/10.15695/jstem/v2i1.13>

ABSTRACT: As interest increases in promoting STEM education in America, summer STEM programs are a promising option for increasing student engagement, interest, and knowledge of STEM. However, STEM programs pose challenges for evaluation, especially programs that serve a large number of students and address a wide range of STEM topics. This paper describes how a team of researchers and practitioners collaborated to design and implement an evaluation of a series of STEM summer programs held at a large, public university. The programs varied in the STEM topics they covered and the age of participants. This created challenges for evaluating a series of programs of such scope and variety. This paper will further describe the programs and the methods used to evaluate them. Illustrative results of the evaluation will be shared, in addition to lessons learned from our evaluation in the hopes that this paper can serve as a resource for those looking for a feasible way to evaluate large, diverse programs.

INTRODUCTION

Science, technology, engineering, and mathematics (STEM) education is commonly viewed as a promising response to the national need to build America's modern workforce (National Center on Education and the Economy, 2008; National Science Foundation & National Science Board, 2015). Summer STEM programs offer a unique opportunity to meet this demand, allowing students to engage deeply with multiple STEM topics. However, there are challenges to evaluating the impact of such programs on students because of the need to design evaluations appropriate for program duration, size, and scope, in addition to numerous logistical factors. Program duration is an important factor for measuring the intended outcomes of a program. Summer STEM programs often have a limited time frame, restricting the outcomes that can be effectively measured and the data collection tools appropriate for measuring these outcomes (Allen and Peterman, 2019; Wilkerson and Haden, 2014). An additional challenge of evaluating summer STEM programs is the need to tailor evaluations to the size and scope of the program. Evaluations of summer STEM programs that focus on only one program or one STEM topic often face chal-

lenges related to small sample sizes or a narrow disciplinary scope (Binns et al., 2016; Ciston et al., 2010; Yilmaz et al., 2010). Conversely, programs that focus on multiple STEM topics or integrated STEM may face challenges designing measures that capture student outcomes across many STEM topics. This is further complicated given that many programs are interested in understanding student outcomes that are not easily defined and measured, such as interest and motivation (Allen and Peterman, 2019).

When numerous summer programs are held simultaneously across a university, designing and managing an effective evaluation becomes challenging. There is literature describing the evaluations of such programs; however, those conducting such evaluations report that their work is limited by a number of factors, including relying solely on quantitative survey data, using measures with unknown validity and reliability information, or having limited time to conduct an evaluation (Conrad, 2018; Crombie et al., 2003; Zoldosova and Prokop, 2006). A survey of over 300 camp professionals from across the United States revealed additional logistical concerns surrounding evaluation of summer programs, in-

cluding finding time to administer evaluations, interpreting and using evaluation findings, and promoting a culture that supports evaluation (Wilson, 2017).

In spite of these challenges, designing quality studies of summer STEM programs is crucial for successfully determining program impact. In a meta-analysis of 15 STEM out-of-school time (OST) programs, Young and colleagues (2017) reviewed the construct, internal, external, and statistical validity of 15 studies to assess overall design quality. The authors found that research design quality was a significant moderator of the effect of OST STEM programs on student interest in STEM. This finding indicates the importance of designing methodologically sound evaluations to successfully determine the impact of programs on their participants. Additionally, the identification of stakeholders, an all-encompassing term for those individuals, groups, or organizations who are served by or have an interest in an evaluation, at the beginning of the evaluation process has been identified as an important step to avoid conflict and to enhance the quality of evaluation results (American Evaluation Association, 2018a; Mertens and Wilson, 2012). By identifying key stakeholders at the beginning of a program, as well as continued stakeholder involvement throughout an evolving program, evaluations can be designed and implemented to better understand the intricacies of the unique program context from multiple perspectives, and therefore, collect meaningful data while increasing the likelihood that the evaluation results will be used to guide future program activities (Patton, 2011).

This paper describes how a team of researchers and practitioners collaborated to design and implement an evaluation of a series of STEM summer programs held at a large, public university. The programs varied in the STEM topics they covered and the age of participants, resulting in many challenges regarding the design and implementation of an evaluation for a series of programs of such scope and variety. This paper will provide a brief description of the summer programs, and present the methods used to measure the constructs of confidence in STEM knowledge, attitudes towards STEM, and intent to persist in STEM. Additionally, a series of illustrative results of this evaluation will be shared. Lastly, the lessons learned from our evaluation will be described, with the hope that this paper can serve as a resource for those looking for a feasible way to evaluate large, complex summer learning programs.

Out-of-School Time Learning. There is some evidence in the literature to suggest that summer camp programs related to STEM may be a useful way to promote STEM among K-12 students. For example, participants in STEM summer programs reported increased positive attitudes towards STEM after participating in camps that were one to two weeks long (Crombie et al., 2003; Elam et al., 2012; Jordan and Sundberg, 2004; Mohr-Schroeder et al., 2014; Nugent et

al., 2010). Summer STEM programs have also been shown to increase students' confidence in their understanding of STEM content (Crombie et al., 2003; Nugent et al., 2010) or STEM laboratory techniques (Knox et al., 2003). Summer STEM program participation has also been linked to students' intent to continue taking classes in STEM fields or to pursue a STEM career (Binns et al., 2016; Crombie et al., 2003; Dabney et al., 2011; Kong et al., 2014; Mohr-Schroeder et al., 2014; Yilmaz et al., 2010). For example, a study of summer programs conducted across five states found that students who participated in summer science camps were twice as likely to report that they were interested in a career in science and engineering than participants who had not previously participated in summer science camps (Kong et al., 2014).

Even long after students participate in summer programs, there are positive effects on students' interest in STEM careers (Bischoff et al., 2008; Dabney et al., 2011; Knox et al., 2003). Knox and colleagues (2003) found that almost a year and a half after students participated in a 2- to 4-week-long summer science program, a majority of participants reported that attending the summer program largely contributed to their interest in a science career. One study of almost 7,000 college students across the U.S. found that OST experiences dating as far back as middle school impact a student's likelihood of pursuing a STEM career in college (Dabney et al., 2011). The authors' study focused on surveying college students in introductory English courses to compare those who intended to pursue STEM careers and those who intended to pursue non-STEM careers. Participants who reported participating in science clubs, competitions, or camps in middle or high school were 1.5 times more likely to report that they were interested in a STEM career in college. This finding suggests that OST science activities in middle and high school are an important mechanism for increasing students' interest in pursuing STEM careers.

Evaluation of Summer Programs. Despite the many challenges of evaluating summer STEM programs, there are effective practices for designing evaluations suggested in the literature. Wilkerson and Haden (2014) developed a framework for designing evaluations of STEM OST programs that balance program duration with possible methodological approaches and outcomes. The authors suggest that evaluations of programs with short durations, such as week-long summer STEM programs, should focus on short-term outcomes (such as awareness, interest, attitudes, and program-specific knowledge) and intermediate outcomes (such as continued participation in STEM programs, STEM self-efficacy, and persistence in STEM through courses or future degrees). Academic outcomes and attainment of a STEM career are more appropriate outcomes for programs with durations greater than 60 hours.

Additionally, Wilkerson and Haden (2014) suggest that program duration should also inform the data collection tools and methods used by evaluators. Specifically, shorter programs should be evaluated using brief and efficient data collection tools, such as short surveys. Conversely, longer programs can use a wider variety of data collection tools, including focus groups, longitudinal surveys, or student journal reflections.

The Summer STEM Program. The summer programs evaluated here are a series of one-week long summer programs offered by a Center for Education (Center) that is housed on the campus of a public university in a large southeastern metropolitan region. The Center specializes in STEM education outreach and educational evaluation and research, with a broad mission to advocate for and lead systematic changes to increase STEM interest and achievement for all students in the local and regional community. As such, the Center began offering summer programming in STEM to local students in 1991. Since then, the programs have grown from serving approximately 100 students each summer to serving almost 700, with students traveling from across the United States and abroad to attend.

The summer programs support one of the Center's goals to inspire STEM enrichment and outreach for students. Accordingly, the programs are designed to provide high-quality academic and hands-on STEM enrichment programs, develop partnerships with the larger university and local communities, and expose students to leading-edge research and STEM careers. These efforts are supported by the Center's seat at a large, public, technical university, where student summer program attendees are provided a unique opportunity to experience state-of-the-art facilities, such as teaching spaces and research laboratories. Additionally, because the summer programming is entirely housed on the university campus, it allows students to experience life on a college campus for a week, interacting with current university students and faculty, having lunch in the student center, and experiencing research laboratories by working and touring. The university's metropolitan location and connections with local industry partners also provides students with opportunities to take off-campus field trips to see real-world implementation of the STEM topics they were learning about and meet professionals working in STEM fields. All of these efforts are intended to create an inspiring STEM enrichment experience for students in support of the Center's goals.

During summer 2018, 696 3rd through 12th grade students participated in 30 individual STEM programs, each attended by a maximum of 40 students. These 30 summer program sessions included nineteen unique topics, with each session covering one or more STEM topics. The foci of the camps were diverse, ranging from the analysis and forecasting of hazardous weather, to the physics of roller coasters,

to the psychology of attention. Each program session lasted for one week and provided participants with approximately 30 hours of instructional time. In order to conduct 30 program sessions over seven weeks, the summer programs leadership employed a diverse staff. Specifically, three main staff members were tasked with logistics and coordination of all programs. These staff members worked directly with 11 people from throughout the university who did additional planning and coordination for specific programs. Additionally, 42 faculty, graduate students, undergraduate students, and local teachers were hired across the 30 program sessions as summer program instructors. Five undergraduate students who were summer interns at the Center served as counselors. In addition to those employed by the summer program leadership, the evaluation team also consisted of four evaluators, tasked with evaluation oversight, data collection, data analysis, and reporting. Specifically, the team consisted of two evaluation interns, a research associate, and a senior research scientist who oversaw the evaluation planning and implementation.

Importantly, each of the individuals described in this section, including the summer program students, staff members, and program instructors, were all identified as key program stakeholders. In other words, the individuals involved in both the implementation of the summer program sessions and those individuals who participated in the sessions were identified as having a legitimate interest in the results of this evaluation (American Evaluation Association, 2018b), and are therefore important informants for the purposes of program evaluation.

EVALUATION METHODS

The evaluation of the summer programs supports another of the Center's goals: to advance STEM education through crucial research and impactful evaluation. In addition to using the findings to advance the field of STEM education, the purpose of this evaluation was to provide leaders of the summer programs with formative and summative results that could be used to better understand participant satisfaction and program impact on attending students. Following Wilkerson and Haden's (2014) framework for effective practices for the evaluation of STEM programs, the methods adopted for this evaluation were intentionally chosen to address the challenges of evaluating a large number of unique, one-week summer programs taking place consecutively across a university campus.

Guiding Evaluation Questions. To provide effective formative feedback and summative impact information to program stakeholders, it was necessary to design evaluation questions that guide the evaluation methods while ensuring the data answers questions that are of interest to program

stakeholders, namely whether the programs are meeting program-level and Center-level goals. Thus, the guiding questions for this evaluation were as follows:

1. What is the overall program impact on participants' confidence in STEM content knowledge, attitudes towards STEM, and intent to persist in STEM?
2. To what extent are participants satisfied with their program experience?

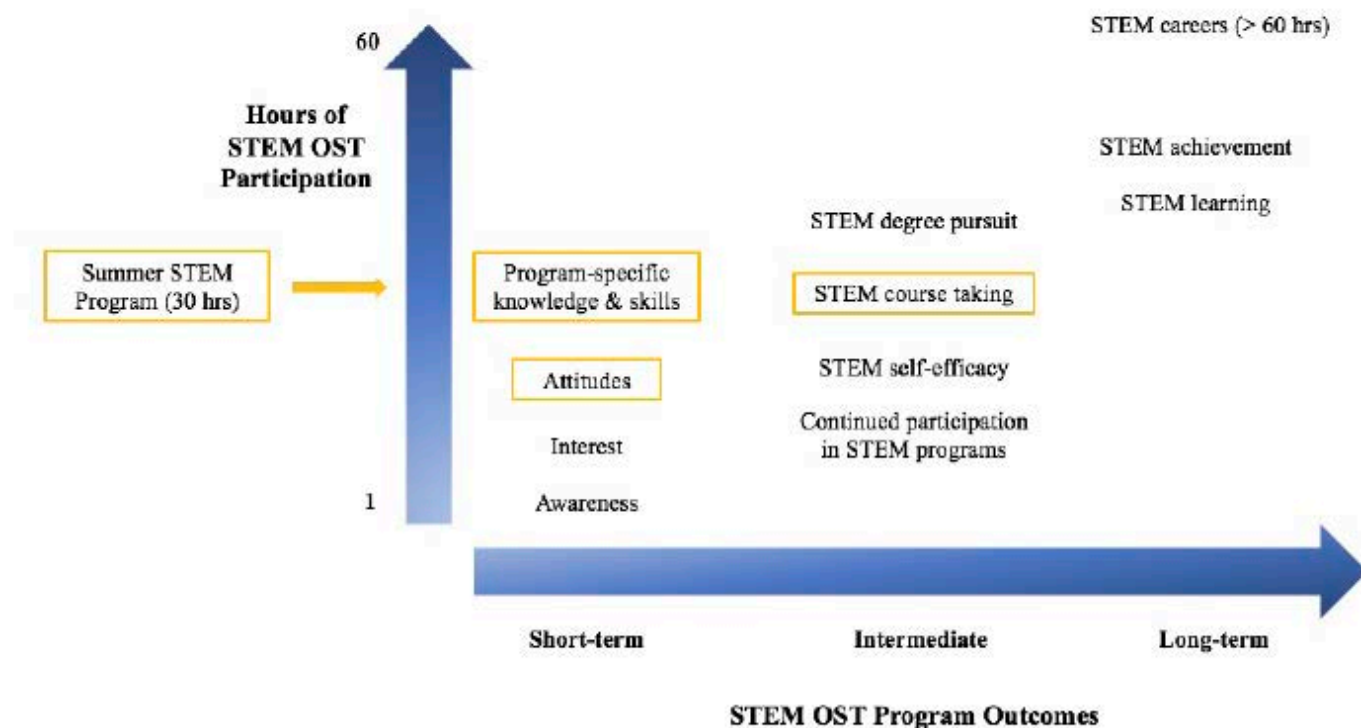
Summer Program Selection. All 30 programs that took place during the summer of 2018 were included in the evaluation. However, 15 programs were selected for a more extensive evaluation that aimed to investigate changes in student perceptions over the course of the week. In collaboration with program stakeholders, these sessions were chosen because they were more established programs that had been implemented for multiple years, and therefore, program stakeholders were specifically interested in the impact of these programs on attendees.

Data Sources. Although the evaluation included all 30 summer programs, the data sources described here are specifically for the 15 programs where a more in-depth evaluation was implemented. This is to provide an illustrative view of methods used to design evaluation instruments that can be easily adapted to any similar summer programs. While these methods are specifically applicable to the 15 programs that were selected for a more extensive evaluation, note that

all remaining programs were also evaluated using at least a post-survey, which is described below.

Survey development. The surveys used in this evaluation were designed to assess whether programs were meeting program-level goals related to specific STEM content, as well as Center-level goals related to creating an inspiring STEM enrichment program. Thus, the survey was designed to include items that were tailored to specific program topics, as well as items that were consistent across all programs. Figure 1 illustrates how the outcomes measured in the survey align with Wilkerson and Haden's (2014) framework for STEM OST program evaluation, using short-term and intermediate outcomes suitable for a program with a 30-hour duration.

In order to accommodate the unique program topics and specific age groups of the camps that were evaluated, the evaluation team created four foundation surveys: pre- and post- surveys appropriate for lower and upper grades. Each of these "foundation" surveys was easy to adapt to each unique program context, while still measuring the constructs of interest. This method allowed for the efficient creation of pre- and post-surveys that were tailored to each of the 15 unique programs that were evaluated. Additionally, when developing the survey, the evaluation team was mindful of the challenge of creating an instrument to measure the same outcomes across a range of participant ages. Programs that included students in lower grades (6th grade or younger) received an adapted version of the survey with grade-appropriate language. For example, while upper-grade surveys



Note. Summer STEM Program outcomes measured in 2018 are highlighted in yellow.

Figure 1. Relationship between program duration and outcomes (adapted from Wilkerson & Haden, 2014).

may ask, “*I plan to take a lot of math classes in college*”, a lower-grade survey instead asks, “*I would like to take more math classes in school than my school makes me.*” Here, upper-grades students are asked about their plans for college, while lower-grades students are asked about their plans in school, given that their college career is well into the future as compared to upper-grades students. By adapting the survey items in this way, the context of the item is more appropriate for upper- and lower-grades respectively. Some research suggests that older children tend to use the middle option of a five-point scale more often than younger students, and younger students may avoid answering a survey item using a middle option in a five-point scale, likely feeling that it is necessary to give a more concrete response (Mellor and Moore, 2014). Therefore, response options were altered, such that all 5-point Likert scales were reduced to 4-point scales for the lower-grades surveys. The complete pre- and post-foundation surveys for the upper grades can be seen in Appendices A and B.

A description of the surveys developed for the High School App & Game program is provided below to illustrate how the surveys were designed to capture data relevant to both program-level and Center-level goals. The App & Game program allowed students to gain hands-on experience creating mobile and computer platforms using an online development software. In doing this, students are exposed to basic elements of user interface design, touch controls, graphics editing, and app/game design.

Confidence in STEM content knowledge. In order to measure whether program-specific goals were being met, learning goals for each program were adapted to create three to seven items assessing participants’ confidence in their understanding of the STEM content specific to the program. The participation of program instructors, an identified group of key stakeholders, was crucial for the development of survey items to assess confidence in STEM content knowledge. As the program learning goals were developed by program instructors to fit the unique topic and skills developed in each program, communication with them was necessary to create survey items that aligned to their specific program context. In the case of the High School App & Game program, learning goals provided by the program instructors were used to create three content knowledge items. For example, one program-specific learning goal was “students will learn what an ‘event’ is in terms of computer programming.” Thus, one of the items assessing confidence in STEM content knowledge on the App & Game survey was, “*How confident do you feel describing what the term ‘event’ means in computer programming?*” Other program learning goals were broader. For instance, another goal was “students will learn to create a basic digital app or game.” This goal was adapted into the item, “*How confident do you feel creating a basic digital app or game?*” Participants rated their confidence on these

items using a 4-point Likert scale ranging from “not all confident” to “very confident.”

Attitudes. An identified key stakeholder for this evaluation, the summer program coordinators and staff, indicated an interest in assessing whether programs were meeting the Center’s goal of providing inspiring STEM enrichment; therefore, pre- and post-surveys included items assessing students’ attitudes towards STEM. Existing, validated measures were adapted to assess the impact of the summer programs on students’ attitudes towards STEM. Thirteen questions on the survey were adapted from the validated *Student Attitudes toward STEM Survey-Middle and High School Students and Student Attitudes toward STEM Survey-Upper Elementary School Students* (Friday Institute for Educational Innovation, 2012a, 2012b). These surveys included constructs assessing students’ attitudes towards science, attitudes towards math, and attitudes towards engineering/technology, such as their perceptions of their own abilities and their interest in choosing a career in that field. For example, an item assessing STEM attitudes from this survey is, “*I know I can do well in science.*” Participants rated their agreement on these statements using a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.” Constructs related to student attitudes towards STEM were found to have good internal consistency in upper and lower grades in the existing literature (Cronbach’s alphas greater than 0.83 and 0.89, respectively; Unfried et al., 2015). Select items were chosen from each of these constructs to create a briefer collection of items related to STEM attitudes. Because items were pulled from multiple constructs, STEM attitudes were analyzed as individual items, rather than as one construct. Additionally, some item statements were adapted as necessary to make items appropriate for lower grade students.

Intent to persist. In addition to student attitudes, evaluation stakeholders indicated an interest in assessing a program’s ability to meet the Center’s goal of providing inspiring STEM enrichment. Therefore, nine items on the pre- and post-surveys were based on a previously validated construct measuring students’ “intent to persist” (Alemdar and Lingle, 2013). The developers of the original instrument define intent to persist as, “students’ commitment to study hard, to take more courses in high school, and their intention to use what they learn in their future careers” (Alemdar and Lingle, 2013, p. 2). Specifically, items assessed students’ intent to persist in science, mathematics, and technology. Again, participants used a 5-point Likert scale ranging from “strongly disagree” to “strongly agree” to rate their agreement with statements. For example, students were asked to rate the extent to which they agreed with the statement, “*I am committed to study hard in my science classes.*” When validated with middle school students, this construct demonstrated good reliability (Cronbach’s alpha = 0.86). As opposed to the previously described items for measuring student at-

titudes that were analyzed as individual survey items, the items that were used to measure Intent to Persist in STEM were analyzed as a construct. As previously described, the survey items were adapted for the lower grades, where the language needed to be understandable for elementary-aged students.

Items specific to pre- or post-survey. In addition to assessing students' confidence in STEM content knowledge, attitudes towards STEM, and intent to persist in STEM, the pre-survey included demographic questions and items assessing students' motivation for attending the summer program. For example, a survey item measuring students' motivation for program attendance was, "*I am attending this program because my parents want me to.*" Participants answered "yes" or "no" to each item.

The post-survey included additional items assessing general satisfaction with the program and instructors. For example, a survey item used in combination with others to assess general program satisfaction was, "*I enjoyed the activities we did this week.*" Participants rated their agreement on a 5-point Likert scale from "strongly disagree" to "strongly agree." Participants were also asked to provide suggestions on the post-survey that could be used to improve the program for future students.

Survey Implementation. In order to effectively use and distribute the surveys to students in each of the summer programs, it was necessary to coordinate logistics with summer program staff. For those programs that received both a pre- and a post-survey, summer program instructors were asked to set aside 20 minutes prior to the start of any program-related instruction, as well as 20 minutes following the completion of any program-related activities on the first and last days of the program, respectively. Therefore, those summer programs in which both a pre- and post-survey were distributed were required to set aside a maximum of 40 minutes for evaluation activities. Alternatively, instructors of programs selected to receive only a post-survey were only asked to set aside a maximum of 20 minutes following all program instruction on the last day of the program for students to participate in the evaluation.

Additionally, given that at least one summer program intern was present in each program, it was determined that the most effective process for survey distribution would be to utilize summer program interns. Therefore, the evaluation team worked to train summer program interns on proper and ethical survey distribution procedures. The primary purpose of this training was to introduce the summer program interns to student assent procedures, and ensure that the evaluation activities remained within the bounds of the approved data collection protocol. Additionally, all summer program interns completed a research ethics and compliance training on human subjects' research to ensure that they were aware

of appropriate procedures for conducting research, particularly with vulnerable populations, such as children.

Human Subjects Approval Processes. An important component of any evaluation that involves working directly with minors is the process of receiving Institutional Review Board (IRB) approval. At the Institution where the summer programs took place, it was necessary to start the IRB process months prior to the start of the summer programs. This provided the evaluation team with the time necessary to make edits to the evaluation methods and implementation plans as needed based on feedback from the human subjects committee, as well as provide the approvals necessary to build in necessary consent procedures with summer program logistics. For example, parents must sign consent forms in order to use the data collected from their children, the evaluation participants, in the resulting analyses. It was requested by the program staff that these processes align with their registration logistics and with morning drop-off of students on the first day of each program. To ensure that these processes were in place, it was necessary to have an approved IRB well ahead of the first day of summer programs. Additionally, everybody involved in the evaluation was required to have a Collaborative Institutional Training Initiative (CITI) certification. Therefore, beginning the IRB process well ahead of time provided the time necessary for all personnel to have the required certifications necessary to implement the approved evaluation activities.

Data Analysis. Quantitative survey data were analyzed using descriptive statistics and significance testing of means when appropriate. Specifically, frequencies, means, standard deviations, and paired-samples t-tests were conducted to compare the difference in student responses on items that were identical on the pre-survey and post-survey. Additionally, effect sizes were calculated using Cohen's *d* to assess practical significance. While significance testing, such as the paired samples t-tests used in this paper, is important for assessing statistical significance, this does not necessarily imply that the results are important in practice (Kirk, 1996). Therefore, the effect size (i.e., Cohen's *d*) is used to test for practical significance, and is operationally defined throughout this report as .20 having "small" practical significance, .50 as "medium" practical significance, and those values of .80 as "large" practical significance (Cohen, 1988). Only descriptive statistical analyses were conducted for items that were unique to either the pre- or post-survey to provide general information on trends in participants' responses. Qualitative data provided in responses to open-ended survey items were analyzed using a general inductive approach (Thomas, 2006). These data were summarized in order to provide richer information on students' perceptions of the summer program in which they participated.

EVALUATION FINDINGS

Data from all programs that received a pre- and post-survey were compiled and analyzed to give a general sense of whether programs were meeting program- and Center-level goals. Where appropriate, results are presented separately for participants in grades 7 through 12 and 3 through 6. Individual reports were also generated for each program so that program stakeholders could investigate the progress of specific programs towards their goals.

Below are findings from summer programs that were evaluated using a pre- and post-survey to illustrate how formative and summative data can be used for better understanding program impact. The evaluation findings presented here are meant to demonstrate the kind of data that can result from an evaluation of the type described in this paper. They are also meant to serve as an example for how one might present evaluation findings of a similar nature. Additionally, the results presented here are followed by a description of the ways in which data from this evaluation was shared with stakeholders, providing an illustrative example of stakeholder involvement in the evaluation process.

Demographics. The final sample of participants for whom pre- and post-survey data are available consisted of 244 students. Some students participated in multiple programs over the course of the summer, and therefore, completed surveys for each program that they participated in. Thus, for the purpose of this evaluation, students who participated in multiple programs during the summer were counted as multiple participants. Of the 235 participants who provided information on their gender, 69% reported that they were male and 31% reported that they were female. The majority of participants self-reported their ethnicity as “African American” (29%), while slightly less participants indicated their ethnicity as “White” (28%). The remaining 43% of participants self-reported their ethnicity as “Asian/Pacific Islander”, “Multi-racial”, “Hispanic/Latino”, or “Other”. Additionally, participants were asked to indicate their motivation for attending the program on the pre-survey. The majority of participants (86.86%) reported that they were attending the program because they wanted to learn more about science and math.

Confidence in STEM Content Knowledge. Items assessing participants’ confidence in program-related STEM content were developed based on the unique learning goals of each program. For this reason, findings related to participants’ confidence in their knowledge of STEM content were presented to program leadership in separate reports for each program. The findings shown here demonstrate the change in average confidence on all items assessing confidence from pre-survey to post-survey for each program (Figure 2). Data from program sessions that were repeated across multiple weeks with the same age group were compiled for

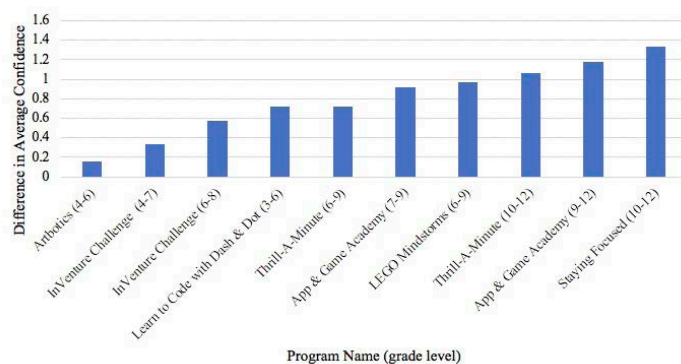


Figure 2. Change in average confidence in knowledge of STEM content after completing a summer camp session.

these analyses. One program did not submit learning goals to be adapted for use in the surveys and is not included in the figure. In all cases, participants’ average confidence with STEM content increased from pre-survey to post-survey. The increase in average confidence was generally greater for programs with upper-grade participants than programs with lower-grade participants.

Attitudes. Results show that participants’ positive attitudes towards STEM increased for students in upper grades. Paired-sample t-tests of data from upper-grade participants revealed a statistically significant mean increase on the majority of items related to STEM attitudes. However, the estimated effect sizes on these items indicated that not all of the mean increases that were statistically significant were also practically significant. There were four items for which the Cohen’s *d* measure of estimated effect size indicated a small but practically significant increase in average agreement

Table 1. Participants’ Changes in Attitudes Towards STEM, Grades 7-12

Item	n	Mean (SD)		Difference	Cohen’s <i>d</i>
		Pre	Post		
I am good at building or fixing things.	205	3.84 (0.93)	4.05 (0.87)	0.21***	0.23
Engineering activities are hard for me.^a	197	3.70 (0.94)	3.91 (0.96)	0.22***	0.23
I feel good about myself when I do engineering activities.	203	3.84 (0.95)	4.09 (0.92)	0.25***	0.26
When I am older I think I will choose a job that uses math.	204	3.60 (1.24)	3.84 (1.15)	0.25***	0.21

^a Note. Difference = post-survey mean – pre-survey mean. Thus, a positive difference indicates an increase in degree of agreement from pre-survey to post-survey.

^aItem was reverse coded.

p* < .05, *p* < .01, ****p* < .001

from pre-survey to post-survey (Table 1). Three of these four items were related to engineering, indicating that the programs that specifically incorporated engineering-related activities have an especially important impact on students' attitudes towards engineering. For students in lower grades, there were no statistically or practically significant changes in STEM attitudes. However, for the majority of items, the post-test scores increased as compared to the pre-test scores. This lack of significant findings could be due to the fact that participants initially reported positive attitudes towards STEM on the pre-survey, resulting in little room for growth following a week-long summer program.

In addition to the finding that the majority of students attended the program specifically to learn more about math and science, this pattern of results suggests that most participants had positive attitudes towards science and math before attending a program. Not only did students begin the programs with positive attitudes towards STEM, their attitudes towards STEM were, on average, either maintained (lower grades) or increased (upper grades) by the end of the programs.

Intent to Persist. Responses to questions assessing students' intent to persist in STEM were analyzed as a construct representing participants' interest in continuing to take STEM courses or pursue a STEM career. There was a statistically significant, but not practically significant, increase in upper-grade participants' intent to persist in STEM. A paired-samples t-test revealed that there was a statistically significant mean increase in students' intent to persist in STEM from pre-survey ($M = 4.01$, $SD = 0.62$) to post-survey ($M = 4.10$, $SD = 0.65$), $t(203) = 3.35$, $p = .001$, $d = 0.14$, 95% CI [0.04, 0.14]. However, the estimated effect size indicates that this increase has little practical significance.

In contrast, there was a statistically and practically significant increase in lower-grade participants' intent to persist in STEM. A paired-samples t-test revealed that there was a statistically significant mean increase in students' intent to persist in STEM from pre-survey ($M = 3.20$, $SD = 0.48$) to post-survey ($M = 3.34$, $SD = 0.49$), $t(38) = 2.63$, $p = .012$, $d = 0.29$, 95% CI [0.03, 0.25]. While the practical significance of this finding is small, these results indicate that participating in a summer program had a positive impact on lower-grade students' intent to continue pursuing STEM courses or careers.

Another demonstration of participants' intent to persist in STEM comes from responses to the pre-survey indicating students' motivation for attending the summer program. Over half of the participants (55.79%) reported that they were interested in attending the university where the programs were hosted for college. Because of the university's focus on and reputation for STEM, this may indicate an interest in persisting in STEM prior to participating in the programs.

Satisfaction

Satisfaction with instructors. On average, participants in upper grades agreed with all statements assessing their perceptions of program instructors, indicating their overall satisfaction with instructors. Participants in lower grades, on average, agreed that they understood their teachers' explanations of activities and disagreed that they did not understand what their teachers were talking about during their summer program. This suggests overall satisfaction with program instructors.

Overall satisfaction. Overall, participants of all ages reported satisfaction with their program experience. Both age groups also agreed, on average, that they enjoyed the activities and that they had fun during the program they attended. Additionally, upper-grade participants were asked to rate their agreement with statements assessing the level of difficulty of program content and their engagement in class discussion throughout their program. Participants, on average, neither agreed nor disagreed with both of these items. Approximately 97% of lower-grade students reported that they would recommend the summer program to their friends and 94% of upper-grade students reported that they would recommend the summer program to their friends. Taken together, these findings reflect generally positive perceptions of the program experience.

Presentation of Results to Stakeholders. The identification and involvement of stakeholders at the beginning of the program and throughout the program helps evaluators in decisions regarding the presentation of evaluation findings to key stakeholders, as well as the eventual use of those findings to both inform the continued evolution of a program and in describing the impact of the program. In this program evaluation, it was determined that the presentation of formative and summative evaluation findings would be primarily to those stakeholders who implement the program, from the teachers in the classroom to the program staff employed by the Center. Therefore, results were presented in two formats: formative and summative results to individual reports for each summer program and an overall report of summative findings from across all summer programs.

For program instructors, individual reports were provided that included findings from their specific program. Findings were presented in a manner similar to the illustrative results previously discussed, such that the results could be used to inform future design of the summer program and make changes in the curriculum or delivery of information as needed. For example, if instructors noticed that students lacked confidence in completing an activity integral to the program's learning goals, the instructor could make changes as necessary for future iterations of the program curriculum. Additionally, data were provided on student satisfaction with various aspects of the program, such as ability to participate

in discussion and time dedicated to hands-on learning. Such data could again be used by instructors to make changes to their program curriculum as necessary.

For the Center-level program coordinators, a larger report was provided that discussed findings from across the summer programs. While formative data was provided that was similar to the previously discussed reports provided to program instructors, the data provided was also summative in that statistical analyses were conducted to examine program impact across all programs on student attitudes, intent to persist, and overall satisfaction with their program experience. Again, these constructs aligned with Center-level goals for the summer programs, and therefore, the presentation of findings in this manner directly relates to the information requested by Center-level staff through previous communication with them as key stakeholders and users of the evaluation. Center-level program staff received a copy of the report, but a presentation of results was also prepared to encourage staff to ask questions as needed, as well as to discuss key findings among each other and beginning the process of planning ways to move forward with summer programs for the following summer. Therefore, for program coordinators and other Center-level staff, the evaluation provided information necessary to move forward with the summer programs as a whole, ensuring that it is known what changes may need to take place to bring the Center closer to meeting its overall goals hoped to be achieved through the offering of summer programs to students.

SUMMARY

With the challenges of evaluating summer programs, and the heightened challenge of evaluating a large, diverse series of programs, this evaluation demonstrates that with planning and preparation, evaluations of this nature can be effectively implemented and presented for use by program stakeholders. Our findings address the guiding evaluation questions that were established at the onset of this evaluation. Participants' average confidence in their knowledge of STEM content increased for all programs for which pre- and post-survey data were available. This increase was greatest for programs with participants in the upper grades. There were small but significant increases in upper-grade participants' average agreement with statements related to their attitudes towards STEM, particularly engineering, after participating in a program. There was a small but significant increase in lower-grade participants' intent to persist in STEM after participating in a program. The increase in upper-grade students' intent to persist was statistically, but not practically, significant. In addition to these outcomes, results indicated that participants were generally satisfied with their instructors and their program experience.

Overall, participants in the summer program showed

increased confidence in their STEM content knowledge, attitudes towards STEM, and intent to persist in STEM. Additionally, students were satisfied with their program experience. Demographic data also suggests that the summer program is supporting the Center's broader mission of bringing STEM to a diverse group of students. Taken together, these findings suggest that the summer programs are meeting program-level goals related to STEM content, as well as the Center-level goal of offering inspiring STEM enrichment for students.

Lessons Learned. With 700 students, 61 program staff, and 30 program sessions held simultaneously across seven weeks, there were many challenges to designing and implementing this evaluation. With preparation and close collaboration at every stage of the implementation process, this evaluation was able to provide useful and meaningful data to program leadership. As summer STEM programs continue to be a popular form of supplementing in-school attempts to foster STEM learning, it will continue to be important to assess whether these programs are achieving their desired outcomes. A description of the methods and lessons learned from conducting evaluations within this complex out-of-school time summer learning context are lacking in the literature, and therefore, those hoping to begin evaluations in such contexts must rely on trial and error over a period of time to adapt the proper methods and logistics needed for a successful and effective evaluation. Similar programs to the one described here may benefit from the lessons learned during this evaluation process.

Time. Given the limited amount of instructional time available in a week-long summer program, designing measures and methods that were efficient was essential to the success of this evaluation. Surveys needed to be easily administered and capable of collecting a large quantity of data in an efficient manner, oftentimes simultaneously across the university campus. The survey that was used took only 15 to 20 minutes to complete, meaning that, if taken on Monday and Friday, the evaluation took only 30 to 40 minutes of program time each week. The pre- and post-surveys were also designed in such a way that their preparation for each program each week could be done efficiently. Specifically, by designing foundation surveys that could be easily modified to fit the content and age group of the programs minimized the time the evaluation team needed to spend creating surveys.

Logistics. To ease the logistical burden of this evaluation, it was essential to plan how evaluation tasks would be divided among program staff. One strategy for managing logistics during this evaluation was to have summer program counselors assisting with evaluation activities. Because there was one counselor assigned to each program, they could ensure that consent, assent, and survey administration took place

each week for each session. This allowed for a more seamless integration of evaluation activities during the week and required less time transitioning between program activities and evaluation activities. It also distributed the work of the evaluation across a number of program staff. Counselors were trained in ethical data collection practices and survey administration in order to prepare them to execute this role and build their confidence conducting evaluation activities. With almost 700 students attending the programs in Summer 2018, this collaboration was crucial to ensure that the logistics of this evaluation were not overwhelming for the program or evaluation staff.

Communication with program stakeholders. In addition to the survey design, some survey items specifically relied on communication with program stakeholders, including program instructors, counselors, and program coordinators. For example, it was necessary to communicate with instructors well ahead of each program to design survey items related to student confidence that aligned with the learning objectives within each program. Therefore, the evaluation team needed to have well-established communication channels with program coordinators to then contact program instructors, who then relayed their instructional goals. Additionally, because counselors played an important role in both consent procedures and survey administration, it was necessary to have an open line of communication with counselors and program coordinators to effectively address any questions or challenges that arose when evaluators were not present. To ensure that this line of communication remained open, a mutual trust was developed between evaluators and summer program interns. From the evaluator's perspective, trust in program interns was important due to their active role in ensuring IRB procedures were followed. From the program interns' perspective, trust in evaluators was necessary to ensure that any questions that arose while they worked on evaluation activities in the field were brought back without fear of repercussions or conflict due to their actions. An open-line of communication, using both e-mail and face-to-face contact, helped to establish this trust. As evaluators, we consistently spoke with program interns, ensuring them that we were both participants in a mutually beneficial relationship. While they were tasked with helping ensure evaluation activities were conducted appropriately, evaluators worked to ensure evaluation activities were well-integrated with necessary program logistics, such that the evaluation added little work to their already full plate. This continued communication built an understanding of each other's roles, and evaluators consistently adjusted protocols as necessary and within the constraints of the IRB to ensure program intern's concerns were heard and respected. Establishing communication and trust between evaluators and program interns proved to be vital to the success of this evaluation.

Interpreting and using evaluation findings. An anti-

ipated challenge of this evaluation was making sure the data that were collected could be meaningfully interpreted and used by program leadership to make decisions. Therefore, it was crucial to establish guiding evaluation questions, in collaboration with program leadership, from the onset. This ensured that the evaluation designed using those questions would be both practical and useful for program stakeholders. Specifically, evaluation questions and data collection activities were guided by the discussions between evaluator's and program stakeholders. The surveys were designed such that the evaluation would provide meaningful data to program leadership by aligning survey constructs and items with the overall Center goals for the summer programs. Data collection and interpretation was then situated in the context of these guiding questions. Findings were presented to program leadership in a manner that would allow them to guide program decisions for future years using data-informed decisions. Following the presentation of findings to program leadership, both formal and informal discussions were held with evaluator's regarding the overall findings. The evaluation findings have proven especially useful given changes in program leadership. While new employees are often burdened by learning to run such a large program in a university context without any prior knowledge of the outcomes in previous years, the evaluation reports have provided them with a starting place from which they have the opportunity to make data-driven decisions regarding program logistics and curriculum moving into a new iteration of the summer programs. Additionally, new program leadership has been encouraged to contact the evaluators regarding any questions or concerns they might have regarding the findings in the report, reestablishing the communication channels and trust-building necessary for another successful evaluation.

Cultivating a culture that supports evaluation. A key component of evaluability, a commonly used term to describe the readiness of an organization or program for engaging in an evaluation, is the existing organizational support for an evaluation (Hare and Guetterman, 2014). Again, stakeholder involvement through a series of meetings concerning the proposed evaluation was used to ensure that from the evaluator's perspective, the program leadership was supportive of evaluation activities. In the case of this evaluation, this was successfully accomplished through the close collaboration between those conducting the evaluation and program staff. The evaluation team was in close contact with the program staff and frequently reminded them of their important role within the evaluation. For example, in these meetings, key conversations revolved around the mission of the Center, the program's goals, and ways in which these goals would align with the proposed evaluation methods. Through such discussions, the evaluation team was able to successfully cultivate a culture supportive of evaluation by ensuring that the evaluation data would be of use to program leadership in

ways that were beneficial from their perspective. By explicitly relating the evaluation methods to the program goals and suggestions from program leadership, the evaluation team created a sense of ownership over the evaluation and improved the culture surrounding the evaluation, as evidenced by the ready acceptance of summer program staff to fill integral roles in the successful implementation of evaluation activities, such as the previously described consent procedures and data collection activities.

Improving the evaluation. The challenges that the evaluation team faced during this evaluation will also be used to improve upon the evaluation for future years. For example, in future years, it is hoped that the evaluation will be scaled-up to include the use of pre/post-surveys in all programs, rather than a selection of programs. Additionally, the evaluation team will continue to address outcomes that are important for program leadership, while being mindful of lingering empirical questions in the field. For example, in future years, it will be important to focus on summer programs that specifically serve underrepresented groups (Binns et al., 2016) and to assess pedagogical strategies used in summer programs and their differential impacts on participants (i.e., Conrad, 2018). Lastly, consistent annual data collection will result in data from multiple years of summer programs. This will provide important insights into which programs are consistently achieving the desired results and which may need more fine tuning. By being mindful of the challenges and successes of this evaluation, future evaluations will be tailored to better serve the students, programs, and the field more broadly.

Evaluation data are an important source of information that can help programs make better decisions and be mindful of where they are directing their resources. This is especially important for programs that are offered at a high cost to participants and universities, such as summer STEM programs. In this case, it is especially important to use evaluation data to ensure that programs are using their resources in ways that promote desired outcomes for students. There is no denying that evaluations are an investment of time and resources. This was evident in the case of the evaluation described here, as is demonstrated by the many program staff who were involved in planning and implementing this evaluation. The time and energy required to conduct an evaluation should not be minimized, but as has been presented here, there are ways of designing evaluations that balance the interest in gathering data with the feasibility of the data collection process and utility of the collected data.

ASSOCIATED CONTENT

Appendices mentioned in this manuscript can be found uploaded to the same webpage as this manuscript.

AUTHOR INFORMATION

Corresponding Author

Christopher J. Cappelli. Center for Education Integrating Science, Mathematics, and Computing. Georgia Institute of Technology. 817 W. Peachtree St. NW, Suite 300. Atlanta, GA 30308. (404) 385-4152. chris.cappelli@ceismc.gatech.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

FUNDING SOURCES

This project was funded by Internal Center Funds

ACKNOWLEDGMENTS

We would like to thank the many Center faculty and staff that contributed to this evaluation. Among those, we would like to thank Mr. Christopher Thompson, Ms. Katherine Guzdial, and Mr. James Surgent for their efforts and support in integrating evaluation activities with their summer programs. Additionally, we would like to thank Mr. Tien Tran and Ms. Rabiya Abdus-Salaam for their support with data collection. We would also like to thank the participating students and their families.

ABBREVIATIONS

OST: Out-of-school Time; STEM: Science, Technology, Engineering, Math;

REFERENCES

- Alemdar, M., and Lingle, J. (2013). Middle school STEM student survey. Developed for National Science foundation (NSF) Project: Advanced Manufacturing and Prototyping Integrated to Unlock Potential (AMP-IT-UP) (NSF #1238089).
- Allen, S., and Peterman, K. (2019). Evaluating informal STEM education: Issues and challenges in context. *New Directions for Evaluation*, 2019(161), 17-33.
- American Camp Association (2011). Year-round education public policy position. Retrieved from https://www.acacamps.org/sites/default/files/page_documents/publicpolicy/Position-Statement-Year-Round-Education.pdf
- American Evaluation Association (2018a). AEA competencies. Retrieved from <https://www.eval.org/page/competencies>
- American Evaluation Association (2018b). American evaluation association guiding principles for evaluators. Retrieved from <https://www.eval.org/p/cm/ld/fid=51>

- Binns, I. C., Polly, D., Conrad, J., and Algozzine, B. (2016). Student perceptions of a summer ventures in science and mathematics camp experience. *School Science and Mathematics*, 116(8), 420-429.
- Bischoff, P. J., Castendyk, D., Gallagher, H., Schaumlöffel, J., and Labroo, S. (2008). A science summer camp as an effective way to recruit high school students to major in the physical sciences and science education. *International Journal of Environmental and Science Education*, 3(3), 131-141.
- Ciston, S., Worsdall, E., and Swenson, J. (2010). Summer technology and engineering program: An outreach program benefiting women engineering students from middle school through graduate school. Retrieved from <https://peer.asee.org/16378>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conrad, J., Polly, D., Binns, I., and Algozzine, B. (2018). Student perceptions of a summer robotics camp experience. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 91(3), 131-139.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., and Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227-268.
- Crombie, G., Walsh, J. P., and Trinneer, A. (2003). Positive effects of science and technology summer camps on confidence, values, and future intentions. *Canadian Journal of Counselling*, 37(4), 256-269.
- Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnert, G., Sadler, P. M., and 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.
- Elam, M. E., Donham, B. L., and 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.
- Friday Institute for Educational Innovation (2012a). *Student attitudes toward STEM survey- middle and high school students*, Raleigh, NC: Author.
- Friday Institute for Educational Innovation (2012b). *Student attitudes toward STEM survey- upper elementary school students*, Raleigh, NC: Author.
- Hare, J., and Guetterman, T. (2014). Evaluability assessment: Clarifying organizational support and data availability. *Journal of MultiDisciplinary Evaluation*, 10(23), 9-25.
- Jordan, W., and Sundberg, C. (2004). Introducing engineering to teenagers through a summer camps program. Retrieved from <https://peer.asee.org/13274>
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56(5), 746-759.
- Knox, K. L., Moynihan, J. A., and Markowitz, D. G. (2003). Evaluation of short-term impact of a high school summer science program on students' perceived knowledge and skills. *Journal of Science Education and Technology*, 12(4), 471-478.
- Kong, X., Dabney, K. P., and 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.
- Mellor, D., and Moore, K.A. (2014). The use of Likert scales with children. *Journal of Pediatric Psychology*, 39(3), 369-379.
- Mohr-Schroeder, M. J., Jackson, C., Miller, M., Walcott, B., Little, D. L., Speler, L., Schooler, W., and Schroeder, D. C. (2014). Developing middle school students' interests in STEM via summer learning experiences: See blue STEM camp. *School Science and Mathematics*, 114(6), 291-301.
- National Center on Education and the Economy (US). (2008). *Tough choices or tough times: The report of the new commission on the skills of the American workforce*. New Commission on the Skills of the American Workforce. John Wiley & Sons.
- National Science Foundation and National Science Board. (2015). *Revisiting the STEM workforce: A companion to science and engineering indicators 2014*. Washington, DC: Author. Retrieved from <http://www.nsf.gov/publications/2015/nsb201510.pdf>
- Nugent, G., Barker, B., Grandgenett, N., and Adamchuk, V. I. (2010). Impact of robotics and geospatial technology interventions on youth STEM learning and attitudes. *Journal of Research on Technology in Education*, 42(4), 391-408.
- Patton, M. Q. (2011). *Essentials of utilization focused evaluation*. Thousand Oaks, CA: Sage.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237-246.
- Unfried, A., Faber, M., Stanhope, D. S., and 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.
- Wilkerson, S. B., and Haden, C. M. (2014). Effective practices for evaluating STEM out-of-school time programs. *After-school Matters*, 19, 10-19.
- Wilson, C. (2017). 2017 Emerging issues survey data and analyses. Retrieved from https://www.acacamps.org/sites/default/files/resource_library/2017-Emerging-Issues-Full-Report.pdf
- Yilmaz, M., Ren, J., Custer, S., and Coleman, J. (2010). Hands-on summer camp to attract K-12 students to engineering fields. *IEEE Transactions on Education*, 53(1), 144-151.
- Young, J., Ortiz, N., and Young, J. (2017). STEMulating interest: A meta-analysis of the effects of out-of-school time on student STEM interest. *International Journal of Education in Mathematics, Science and Technology*, 5(1), 62-74.
- Zoldosova, K., and Prokop, P. (2006). Education in the field influences children's ideas and interest toward science. *Journal of Science Education and Technology*, 15(3-4), 304-313.