



Addressing the Ceiling Effect when Assessing STEM Out-Of-School Time Experiences

Nancy L. Staus*, Kari O'Connell and Martin Storksdieck

STEM Research Center, Oregon State University, Corvallis, OR, United States

The aim of this paper is to describe an analytical approach for addressing the ceiling effect, a measurement limitation that affects research and evaluation efforts in informal STEM learning projects. The ceiling effect occurs when a large proportion of subjects begin a study with very high scores on the measured variable(s), such that participation in an educational experience cannot yield significant gains among these learners. This effect is widespread in informal science learning due to the self-selective nature of participation in these experiences, such that participants are already interested in and knowledgeable about the content area. When the ceiling effect is present, no conclusions can be drawn regarding the influence of an intervention on participants' learning outcomes which could lead evaluators and funders to underestimate the positive effects of STEM programs. We discuss how the use of person-centered analytic approaches that segment samples in theory driven ways could help address the ceiling effect and provide an illustrative example using data from a recent evaluation of a STEM afterschool program.

OPEN ACCESS

Edited by:

Nancy Longnecker,
University of Otago, New Zealand

Reviewed by:

Jeffrey K. Smith,
University of Otago, New Zealand
Kathryn Holmes,
Western Sydney University, Australia

*Correspondence:

Nancy L. Staus
stausn@oregonstate.edu

Specialty section:

This article was submitted to
STEM Education,
a section of the journal
Frontiers in Education

Received: 02 April 2021

Accepted: 21 June 2021

Published: 12 July 2021

Citation:

Staus NL, O'Connell K and
Storksdieck M (2021) Addressing the
Ceiling Effect when Assessing STEM
Out-Of-School Time Experiences.
Front. Educ. 6:690431.
doi: 10.3389/feduc.2021.690431

Keywords: ceiling effect, informal STEM learning, evaluation, person-centered analysis/approach, out-of-school activities

INTRODUCTION

As concerns arise about the need to increase the number of US STEM professionals in order to remain globally competitive, the pressure to emphasize STEM education particularly for adolescent youth has never been greater. Many educators and researchers recognize an urgent need to identify strategies for developing youth skills, abilities and dispositions in STEM early in life, particularly for underserved youth, to increase the potential for future academic and professional participation in STEM fields (National Research Council, 2010).

Out-of-school time (OST) activities such as afterschool programs, summer camps, and other enrichment programs (e.g., Girl Scouts science clubs) are uniquely situated to address this need with their ability to reach large numbers of young people, including low-income youth and youth of color (Afterschool Alliance, 2014). While schools often focus on delivering STEM content knowledge and science process skills National Research Council (2012a), OST programs emphasize the fostering or development of affective and emotional outcomes, such as STEM interest and identity, that are strongly associated with STEM persistence and increased future academic and professional participation in STEM fields (National Research Council, 2009; Maltese and Tai, 2011; Venville et al., 2013; Maltese et al., 2014; Stets et al., 2017). However, evaluating the success of such programs can be problematic due to the variable, unstructured nature of informal learning environments themselves, as well as the fact that participants often self-select programs based on their prior interests (National Research Council, 2009). Thus, although some studies have documented

significant cognitive and affective gains from participation in out-of-school STEM activities such as science clubs Bevan et al. (2010), Stocklmayer et al. (2010), Young et al. (2017), Allen et al. (2019), many others, particularly smaller programs with fewer participants, have failed to document significant increases for participants as a whole (Brossard, et al., 2005; Falk and Storksdieck, 2005; Judson, 2012).

The most likely reason for this phenomenon is the presence of a measurement limitation called the ceiling effect which can occur when a large proportion of subjects begin a study with very high scores on the measured variable(s), such that participation in an educational experience cannot yield significant gains among these learners (National Research Council, 2009; Judson, 2012) This effect is often attributed to the biased nature of participation. Informal science learning opportunities, including after school programs, are particularly susceptible to this effect due to the fact that participants generally choose to participate because they are already interested in and potentially knowledgeable about the content area. When the ceiling effect is present, no conclusions can be drawn regarding the influence of an intervention for youth *on average*. This effect can hinder efforts to evaluate the success of a program by leading evaluators to underestimate the positive effects on affective or cognitive learning outcomes that are measured with standard instruments.

In this paper we describe how person-centered analytic models could help informal science evaluators and researchers address the ceiling effect while potentially providing a better understanding of the outcomes of participants in ISL programs and other experiences. We refer to person-centered analytic models as approaches to data analysis that distinguish main treatment effects by participant type in meaningful (i.e., hypothesis-driven) ways. Although used frequently in other fields such as educational psychology, sociology, and vocational behavior research, person-centered analyses are still fairly uncommon in informal science education research and evaluation (Denson and Ing, 2014; Spurk et al., 2020). We begin with a short discussion of the ceiling effect in OST programs and the affordances and constraints of person-centered approaches as compared to more traditional variable-centered models for analyzing changes in outcomes over time. To further clarify the methodologies, we then provide an empirical example in which each type of approach is used on the same data set from the authors' evaluation of 27 afterschool STEM programs in Oregon Staus et al. (2018) in which the usefulness of the person-oriented approach and the variable-oriented approach are compared.

LITERATURE REVIEW

Out-Of-School-Time Programs

Out-of-school-time (OST) programs are a type of informal STEM learning opportunity provided to youth outside of regular school hours that include afterschool programs, summer camps, clubs, and competitions (National Research Council, 2009; National Research Council, 2015). OST programs provide expanded content-rich learning opportunities, often engaging students in rigorous, purposeful activities that feature hands-on engagement,

which can help bring STEM to life and inspire inquiry, reasoning, problem-solving, and reflecting on the value of STEM as it relates to children and youth's personal lives (Noam and Shah, 2013; National Research Council, 2015). In addition, OST STEM activities may allow students to meet STEM professionals and learn about STEM careers Fadigan and Hammrich (2004), Bevan and Michalchik (2013), and can help learners to expand their identities as achievers in the context of STEM as they are actively involved in producing scientific knowledge and understanding (Barton and Tan, 2010).

Another key aspect of OST STEM time is that it is generally not associated with tests and assessments, providing a space for children and youth to engage in STEM without fear and anxiety, therefore creating a psychologically safe environment for being oneself in one's engagement with STEM. In fact, it is the non-assessed, learner-driven nature that makes OST engagement ideal for fostering affective outcomes around interest, identity, self-efficacy, and enjoyment (National Research Council, 2009; National Research Council, 2015). Consequently, many OST programs promote a number of noncognitive, socio-emotional learning (SEL) skills such as teamwork, critical thinking and problem-solving (Afterschool Alliance, 2014). Also known as twenty-first century skills, these skills are seen as essential to many employers when hiring for STEM jobs. Thus, participation in OST programs could potentially positively affect youths' later college, career, and life success (National Research Council, 2012b).

Despite the strong potential for OST programs to provide positive benefits to participants, there have been few studies that document significant changes in outcomes for youth as a result of participation in these programs (Dabney, et al., 2012; National Research Council, 2015). One recent study utilized a mixed-methods approach including surveys and observations of over 1,500 youth in 158 STEM-focused afterschool programs to investigate the relationship of program quality on a variety of youth outcomes and found that the majority of youth reported increases in STEM engagement, identity, career interest, career knowledge, and critical thinking (Allen et al., 2019). The largest gains were reported by youth who engaged in longer-term (4 weeks or more) and higher quality programs as measured with the Dimensions of Success (DoS), a common OST program assessment tool.

Similarly, using a meta-analysis of 15 studies examining OST programs for K-12 students, Young et al. (2017) found a small to medium-sized positive effect of OST programs on students' interest in STEM, although the effect was moderated by program focus, grade level, and quality of the research design. For example, programs with both an academic and social focus had a greater positive effect on STEM interest, while exclusively academic programs were less effective at promoting interest in STEM. The authors found no significant effect for programs serving youth in K-5; all other grade spans showed positive effects on STEM interest. Unlike Allen et al. (2019), this study found no effects related to the duration of the programs.

In contrast to the above large-scale research projects, many researchers or evaluators have failed to document significant increases in STEM outcomes for OST program participants as

a whole. In particular, evaluations of single OST programs with fewer participants may have difficulty showing significant changes in STEM outcomes as a result of participating in the program. For example, an evaluation of a collaboration between libraries, zoos and poets designed to use poetry to increase visitors' conservation thinking and language use, found few significant changes in the type or frequency of visitor comments related to conservation themes or in their thinking about conservation concepts (Sickler et al., 2011). Similarly, in an evaluation of 330 gifted high school students participating in science enrichment programs, evaluators found no positive impact on science attitudes after participation in the program (Stake and Mares, 2001). Although mostly serving adults rather than children, several citizen science projects reported similar difficulty in documenting significant positive outcomes for participants (Trumbull et al., 2000; Overdeest et al., 2004; Jordan et al., 2011; Crall et al., 2012; RK and A, Inc., 2016). For example, an evaluation of The Birdhouse Network (TBN), a program in which participants observe and report data on bird nest boxes, revealed no significant change in attitudes toward science or understanding of the scientific process (Brossard et al., 2005). It is likely that there are many more examples that we were unable to access since program evaluations in general and studies that fail to find significant results in particular, often do not get published.

One plausible explanation for the lack of significant results in program-level evaluations like those described above is not that these programs failed to provide benefits to their participants, but that at least in those with significant positive bias in the participants, the presence of a ceiling effect resulted in a lack of significant gains among these learners on average (National Research Council, 2009; Judson, 2012). For example, in the TBN citizen science study mentioned above, participants entered the program with very strong positive attitudes toward the environment such that the questionnaire used to detect changes in attitudes was insensitive for this group (Brossard et al., 2005). As described earlier, the ceiling effect is a common phenomenon in OST programs which often attract learners who elect to participate because they are already interested in and knowledgeable about STEM (Stake and Mares, 2001; National Research Council, 2009). The potential danger of the ceiling effect is that positive outcomes due to participation in the OST program may go undetected when measured by standard measures which could lead to funding challenges or even termination of a program. Therefore, it is critical that program evaluators utilize appropriate analytic approaches that account for the ceiling effect to better understand how OST programs influence learner outcomes.

Analytic Approaches

Historically, the most common analytic methods when evaluating OST programs have involved a pre-post design using surveys administered at the beginning and end of the program to measure changes in knowledge, attitudes, and similar outcomes, presumably as a consequence of the educational experience (Stake and Mares, 2001). The pre-post data are typically analyzed with a variety of variable-centered approaches such as t-tests or ANOVAs to examine changes in outcomes of interest (e.g., content knowledge, attitude toward science) over the course of the program. However, as described above,

the traditional pre-post design may be insufficient for measuring the impact of intervention programs when many participants begin the program with high levels of knowledge and interest in STEM topics and activities. This is because variable-centered analytic models produce group-level statistics like means and correlations that are not easily interpretable at the level of the individual and do not help us understand how and why individuals or groups of similar individuals differ in their learning outcomes over time (Bergman and Lundh, 2015). In other words, if subgroups exist in the population that do show significant changes in outcomes (perhaps because they began the program with lower pre-test scores), these results may be obscured by the use of variable-centered methods.

In contrast, "person-centered" analytic models are predicated on the assumption that populations of learners are heterogeneous, and therefore best studied by searching for patterns shared by subgroups within the larger sample (Block, 1971). Therefore, the focus is on identifying distinct categories or groups of people who share certain attributes (e.g., attitudes, motivation) that may help us understand why their outcomes differ from those in other groups (Magnusson, 2003). Standard statistical techniques include profile, class, and cluster analyses, which are suitable for addressing questions about group differences in patterns of development and associations among variables (Laursen and Hoff, 2006). However, because of the "regression effect" (i.e., regression to the mean) phenomenon in which those who have extremely low pretest values show the greatest increase while those who have extremely high pretest values show the greatest decrease Chernick and Friis (2003), subgroups must be constructed from variables other than the outcome score being measured. In addition, the selected variables that form the groups must have a strong conceptual basis and have the potential to form distinct categories that are meaningful for analyzing outcomes (Spurk et al., 2020). In the case of OST programs, one such variable may be motivation to participate.

Substantial research shows that visitors to informal STEM learning institutions such as museums, science centers and zoos arrive with a variety of typical configurations of interests, goals, and motivations that are strongly associated with learning and visit satisfaction outcomes (Falk, 2009; Packer and Ballantyne, 2002). Moussouri (1997) was one of the first to identify a typology of six categories of visitor motivations including education, social event, and entertainment, two of which (education and entertainment) were associated with greater learning than other motivation categories (Falk et al., 1998).

Packer (2004) expanded on this work in a study of educational leisure experiences including museums and interpretive sites, in which she identified five categories of visitor motivations: 1) passive enjoyment; 2) learning and discovery; 3) personal self-fulfillment; 4) restoration; and 5) social contact; only visitors reporting learning and discovery goals showed significant learning outcomes. Since then, numerous informal STEM learning researchers have used audience segmentation to better understand the STEM outcomes of visitors (e.g., Falk and Storksdieck, 2005; Falk et al., 2007; O'Connell et al., 2020; Storksdieck and Falk, 2020). These studies suggest that learning outcomes differ based on learner goals or motivations, supporting the potential usefulness of this variable for person-centered analyses in informal science research and evaluation, including OST programs for youth.

In the case of OST programs, children also participate for a variety of motivations including interest in STEM, to socialize with friends, to have fun, and because they are compelled by parents. Thus, person-centered approaches could be used to identify subgroups of participants with differing motivations for participating in the program that may affect their identity and learning outcomes. Then variable-centered analyses such as t-tests could be used to examine changes in outcomes for each subpopulation. To help clarify how the person-centered methodologies described above could address the ceiling effect problem, we provide an illustrative example in which each type of approach is used on the same data set from the authors' prior research and the findings from the person-centered approach and the variable-centered approach are compared.

AN ILLUSTRATIVE EXAMPLE--STEM BEYOND SCHOOL PROGRAM

Background

The empirical example we provide for this paper is the STEM Beyond School (SBS) Program, which was designed to better connect youth in under-resourced communities to STEM

Component	Definition
Learner Identity	Youth see themselves as succeeding in learning and working environments emphasizing science.
Belonging and Relatedness	Youth demonstrate persistence, utilize problem-solving skills and seek help when faced with learning challenges, obstacles, and setbacks.
Purpose and Relevance	Youth demonstrate active participation and interest in science learning.
Competency and Self-Efficacy	Youth feel like they belong in the learning environment, can relate to others and to the topics they are learning within the program.
Constructive Coping and Resilience	Youth believe that learning activities and professional work in science are meaningful, important, and worthwhile.
Cognitive Engagement	Youth believe that they have the capability to succeed in learning opportunities and careers that involve science.

FIGURE 1 | Definitions of each STEM component used in the pre- and post-survey.

learning opportunities by creating a supportive infrastructure for community-based STEM OST programs (Staus et al., 2018). Rather than creating new programs, SBS supported existing community-based STEM OST programs to provide high quality STEM experiences to youth across the state of Oregon. The 27

TABLE 1 | Items comprising survey components and corresponding Cronbach's alphas.

Component and items	Cronbach's alpha (pre/post survey)
Learner identity (6 items) 1. I like learning new things. 2. I like to solve complex problems. 3. I like going to my out-of-school activities that involve science. 4. I like figuring things out. 5. I can succeed in situations that involve understanding science. 6. I Would like a job that uses science when I'm an adult.	0.84/.83 —
Constructive coping and resilience (4 items) 1. When I have difficulty learning something, I remind myself that this is important for my future. 2. If I get stuck, I try something different to solve the problem. 3. If I don't understand something in science, I ask for help. 4. If a problem in science is really difficult, I just work harder.	0.84/.81 —
Cognitive engagement (3 items) 1. I find topics related to science interesting. 2. I enjoy learning new things in science. 3. I Try hard to do well in science.	0.89/.81 —
Belonging and relatedness (4 items) 1. I feel like I am a part of this program. 2. I feel respected in this program. 3. I feel comfortable in this program. 4. I Feel like I can be myself in this program.	0.92/.86 —
Purpose and relevance (4 items) 1. Science is important for my future. 2. Learning science teaches me valuable skills. 3. Science helps people solve problems to make the world a better place. 4. Science helps people understand the world.	0.89/.86 —
Competency and self-efficacy (3 items) 1. I am good at science. 2. I can help others understand science. 3. I am good at solving challenges that involve science.	0.91/.87 —

TABLE 2 | Comparison of pre- and post-survey outcome scores for six affective constructs related to STEM learner identity and motivational resilience ($n = 172$).

STEM outcomes	Pre-survey		Post-survey		t-value	p-value	Cohen's <i>d</i>
	Mean	SD	Mean	SD			
Learner identity	3.84	0.87	3.75	0.90	1.50	0.136	0.11
Cognitive engagement	4.10	1.00	3.94	0.97	1.99	0.047	0.16
Resilience	3.92	0.85	3.89	0.88	0.49	0.626	0.03
Belonging	4.12	1.04	4.12	0.99	0.01	0.991	0.00
Relevance	4.04	0.97	4.09	0.91	0.56	0.577	0.05
Self-efficacy	3.55	1.12	3.51	1.14	0.58	0.566	0.04

Note: Outcomes coded on a five-point scale from 1 = "Strongly disagree" to 5 = "Strongly agree."

participating programs took place predominantly off-school grounds, served youth in grades 3 through 8, and provided a minimum of five different highly relevant STEM experiences located in their communities. The community-based programs were required to provide at least 50 h of learning connected to the interests of their youth that followed the SBS 4 Core Programming Principles (student driven, students as do'ers and designers, students apply learning in new situations, relevant to students and community-based). For comparison, elementary students in Oregon receive 1.9 h per week of science instruction (Blank, 2012). SBS was therefore a targeted investment towards dramatically increasing meaningful STEM experiences for underserved youth while also advancing the capacity of program providers to design and deliver high quality STEM activities for youth that center around learning in and from the community.

SBS requires programs to intentionally engage historically underserved youth, specifically youth from communities of color and low-income communities as well as youth with disabilities and those who are English-language learners. With a grant requirement of engaging at least 70% participation amongst these groups, programs were challenged and inspired to rethink their traditional ways of reaching out, recruiting, and retaining those students.

To ensure long-term benefits for youth, SBS provided capacity building support to the community-based programs in the form of educator professional development, program design guidance, a community of practice for participating providers, support from a Regional Coordinator, and equipment. Educators working directly with youth participated in high quality, high dose (70 h for new providers and 40 h for returning providers) professional development connected directly to their specific needs. Professional development categories included essential attributes in program quality, best practices in STEM learning environments, fostering STEM Identity, and connecting to the community. Rather than providing one-size-fits-all workshops, the program assessed the needs of the educators and then leveraged expertise from across the state to address specific training or coaching needs. This approach created a community- and peer-based "just-in-time" professional learning experience that allowed educators to modify their programming in real time.

Methods and Findings

Like many of the studies discussed earlier, our evaluation of the SBS Program used a pre-post survey design to measure changes in youth outcomes over the course of the OST experience. The survey was

developed in conjunction with the Portland Metro STEM Partnership's Common Measures project which was designed to address the limitation of current measurement tools and evaluation methodologies in K-12 STEM education (Saxton et al., 2014). The resulting STEM Common Measurement System includes constructs that span from student learning to teacher practice to professional development to school-level variables. For the purposes of the SBS Program evaluation, we chose six of the student learning constructs related to learner identity and motivational resilience in STEM-related activities as our outcome measures (Figure 1). The original Student Affective Survey Saxton, et al. (2014) was modified by revisiting its research base and examining additional research (e.g., Cole, 2012). Scales were shortened based on results from a reliability analysis of the included scales of the pre survey in year 1 of the SBS program, and in response to concerns about length and readability from program provider feedback, which led to a redesign of the post survey for the final measure (O'Connell et al., 2017). The final measure consisted of 24 items with three to six items per STEM component, which were slightly modified from the original to be suitable for OST programs rather than classroom environments (see Table 1 for component items and alphas). In addition to these learning outcomes, the pre-survey included demographic items (e.g., gender, age) and an open-ended question to assess youth motivation for participating ("please tell us about the main reason that you are participating in this program"). The answers to this motivation question fell into three categories: 1) interest in STEM topics and activities; 2) wanted to do something fun; 3) compelled by parents or guardians.

Of the 361 youth who participated in the SBS pre-survey in year 3, 148 also completed a post-survey enabling us to examine changes in outcomes associated with SBS programming activities. Here we present the findings in two ways: a variable-centered approach examining mean changes in outcomes for the sample as a whole, and a person-centered approach in which we identify unique motivation-related subgroups of individuals and examine changes in outcomes for each subgroup. We then discuss the usefulness of the person-oriented approach and the variable-oriented approach for addressing the issue of the ceiling effect in ISL research and evaluation projects.

Variable-Centered Analysis

We conducted paired t-tests to examine overall changes in outcomes over the course of the SBS Program and found no significant changes for five of the six outcomes (Table 2). Although there was a significant decline in cognitive engagement, the effect size was

TABLE 3 | Comparison of changes in pre- and post-survey STEM outcomes by subgroup.

STEM outcomes	Pre-survey		Post-survey		t-value	p-value	Cohen's d
	Mean	SD	Mean	SD			
Learner identity							
Interest	4.11	0.68	3.97	0.86	1.83	0.071	0.18
Fun	3.75	0.81	3.58	1.00	1.16	0.253	0.19
Compelled	3.41	0.98	3.45	0.85	0.33	0.748	0.04
Cognitive engagement							
Interest	4.41	0.69	4.17	0.82	2.80	0.006	0.32
Fun	4.05	1.07	3.68	1.19	1.65	0.108	0.33
Compelled	3.70	1.19	3.68	1.06	0.18	0.862	0.02
Resilience							
Interest	4.17	0.67	4.03	0.85	1.74	0.085	0.18
Fun	3.84	0.96	3.81	0.90	0.12	0.903	0.03
Compelled	3.56	0.97	3.66	1.02	0.52	0.609	0.10
Belonging							
Interest	4.43	0.79	4.31	0.82	1.22	0.227	0.15
Fun	4.51	0.53	3.98	1.01	2.48	0.019	0.66
Compelled	3.65	1.25	3.83	1.11	0.78	0.441	0.15
Relevance							
Interest	4.38	0.67	4.25	0.78	1.68	0.097	0.18
Fun	3.91	0.95	3.85	1.15	0.29	0.775	0.06
Compelled	3.77	1.15	3.90	0.99	0.65	0.520	0.12
Self-efficacy							
Interest	3.80	0.98	3.78	0.94	0.12	0.902	0.02
Fun	3.56	1.14	3.24	1.36	1.68	0.104	0.26
Compelled	3.16	1.24	3.16	1.21	0.00	1.000	0.00

Note: Items in index were coded on a five-point scale from 1 = "Strongly disagree" to 5 = "Strongly agree." Interest (n = 84), Fun (n = 32), Compelled (n = 32).

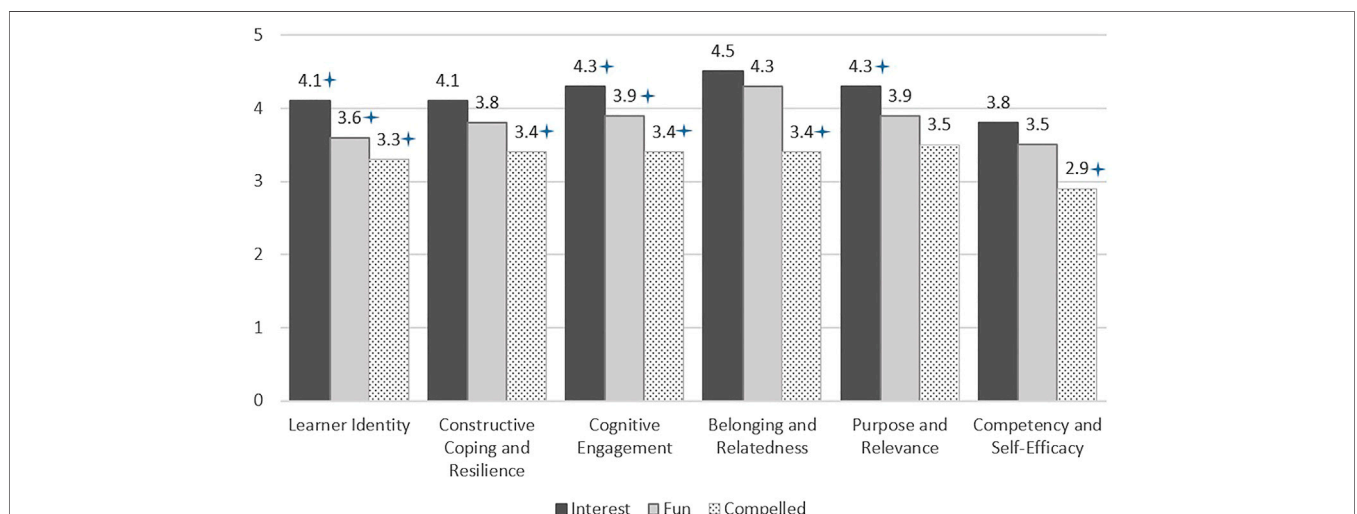


FIGURE 2 | Mean scores for all youth who participated in the pre-survey by motivation class; means with an asterisk are different at the $p < 0.05$ level. All constructs were measured on a scale of 1 (Strongly disagree) to 5 (Strongly agree). Note: $n = 202$ for Interest; $n = 89$ for Fun; $n = 70$ for Compelled.

small ($d = 0.16$). In other words, this analysis indicated that, on average, youth who participated in SBS maintained their STEM identity and motivational resilience over the course of the program but did not show the increases in outcomes that SBS providers desired. An examination of the pre-survey scores indicated that youth on average were already at the higher end of the scale, suggesting that the lack of significant changes in outcomes may be due to the ceiling effect.

Person-Centered Analysis

In order to address the ceiling effect in our data, we segmented youth into unique subgroups based on self-reported pre-survey motivation classes (See **Figure 2**): interested in STEM (Interest), wanted to have fun (Fun), or compelled by parents (Compelled). As described above, theory suggests that youth in these motivation classes may experience different learning outcomes from the same educational intervention. Youth in the Interest subgroup made up 56% of the sample ($n = 202$) and reported significantly greater feelings of learner identity, cognitive engagement, and relevance than youth in the other motivation classes in the pre-survey. The Fun subgroup included 25% of the sample ($n = 89$) and reported similar levels of resilience, belongingness, and self-efficacy as interested youth, similar relevance as Compelled youth, but significantly different learner identity and cognitive engagement than youth in the other subgroups. Finally, Compelled youth comprised 19% of the sample ($n = 70$) and reported significantly lower scores than youth in other subgroups on all outcome measures except relevance.

We then conducted paired t-tests for the 148 youth who completed both a pre- and post-survey. Results indicated only two significant ($p < 0.05$) changes over time: Interested youth reported a significant decrease in cognitive engagement with a moderate effect size ($d = 0.32$), and youth in the Fun subgroup reported a decrease in feelings of belonging with a large effect size ($d = 0.66$) (**Table 3**). None of the subgroups reported significant increases in any of the outcome measures at the end of the program.

DISCUSSION

The above example showed how using person-centered approaches in the evaluation of OST programs has the potential to address the ceiling effect. By segmenting the sample in a theory-driven way, we created three subgroups based on motivation to participate, two of which (i.e., Fun, Compelled) reported low enough pre-survey scores to potentially indicate increases in outcomes as a result of the OST program. In our example, neither the variable-centered nor person-centered approach revealed significant positive changes in outcomes as a result of participating in the program. However, the person-centered approach provided the opportunity to identify such changes for different subgroups of participants. For example, if an OST program led to increased outcome scores for less STEM-motivated youth, such a finding could provide important evidence to funders about the efficacy of OST programs thus promoting longevity of successful STEM-focused youth programs.

Even in the absence of significant changes in STEM outcomes, person-centered approaches provide a more nuanced view of the youth and why they participated which is valuable information that program providers can use to inform future improvements to the program. In the case of SBS, knowing that almost half of youth participated for reasons other than interest in STEM could lead to the development of more effective educational strategies that provide a range of activities designed to engage youth in each motivational category, rather than relying on one-size-fits all programming strategies. Indeed, a recent longitudinal study of youth STEM learning pathways highlighted the importance of customizing STEM resources in the larger learning ecosystem based on the differing interests and motivations of youth in the community (Shaby, et al., 2021). For example, one youth with a strong interest in computer programming eventually lost interest because the content of the OST program he attended did not keep pace with his growing interest in learning new coding languages. While it is unclear why youth outcomes remained largely unchanged after participation in SBS, it is possible that the programming was unable to adequately serve youth with a diversity of interests and motivations for participating.

It is also possible that in addition to the ceiling effect, the study may have suffered from another common measurement challenge associated with traditional pre-post designs known as response shift bias in which participants' comparison standard for measured items (e.g., competency and self-efficacy) differs between pre- and post-assessments (Howard and Dailey, 1979). In other words, program participants may overestimate their knowledge and ability at the beginning of an intervention, while post survey scores may reflect more accurate assessments based on comparisons to others in the program or simply a better understanding of the constructs themselves. Either way, a response shift may exacerbate the ceiling effect and seriously hamper the assessment of true change over time for many respondents (Oort, 2005). One potential remedy to address response shift bias is the use of retrospective pre-post (RPP) designs to simultaneously collect pre- and post-assessment data at the end of a program (Howard, Ralph, et al., 1979). This design provides a consistent frame of reference within and across respondents allowing real change results to be detected from an educational intervention. A growing body of evidence supports the use of the RPP design as a valuable tool to evaluate the impact of educational programs on a variety of outcomes (Little et al., 2020).

Ultimately, to avoid ceiling effects, assessment instruments must be designed to measure outcomes in such a way that participants with a strong affinity for STEM are not already at the high end of the scale when they begin the program. This includes choosing to measure constructs that are not theoretically limited in scale. For example, psychological constructs such as interest have a finite number of phases—once a learner has reached the highest level of individual interest, they will be unable to indicate an increase due to participation in an educational program (Hidi and Renninger, 2006). In contrast, measuring a learner's change in content knowledge may be less limited. Thus, although there is a strong call to use standard, published or previously validated measures in evaluations Noam

and Shah (2013), Saxton et al. (2014), instead of ad-hoc measures adjusted to the nature of a program or the characteristics of the target audience, this may increase the prevalence of the ceiling effect in programs with high positive selection bias if measures are not designed to detect changes over time at the upper end of the distribution.

While it may not be possible to avoid measurement issues such as the ceiling effect altogether in assessments of OST STEM programs, evaluators should be aware of the methodologies and analytic approaches that could be used to address them more effectively. In particular, person-centered approaches that allow the segmentation of participants into motivation-related or other theory-driven subgroups, perhaps in conjunction with retrospective pre-post-survey designs, should be considered at the outset of program evaluations whenever possible.

REFERENCES

- Afterschool Alliance (2014). *America after 3pm: Afterschool Programs in Demand*. Washington, D.C. Available at: <http://www.afterschoolalliance.org/AA3PM/>.
- Allen, P. J., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Little, T. D., et al. (2019). From Quality to Outcomes: a National Study of Afterschool STEM Programming. *IJ STEM Ed.* 6 (37), 1–21. doi:10.1186/s40594-019-0191-2
- Barton, A. C., and Tan, E. (2010). We Be Burnin'! Agency, Identity, and Science Learning. *J. Learn. Sci.* 19 (2), 187–229. doi:10.1080/10508400903530044
- Bergman, L. R., and Lundh, L.-G. (2015). Introduction: The Person-Oriented Approach: Roots and Roads to the Future. *J. Person-Oriented Res.* 1 (1-2), 1–6. doi:10.17505/jpor.2015.01
- Bevan, B., Dillon, J., Hein, G. E., Macdonald, M., Michalchik, V., Miller, D., et al. (2010). *Making Science Matter: Collaborations between Informal Science Education Organizations and Schools*. Washington, D.C.: Center for Advancement of Informal Science Education.
- Bevan, B., and Michalchik, V. (2013). Where it Gets Interesting: Competing Models of STEM Learning after School. *Afterschool Matters*. Retrieved from: <http://eric.ed.gov/?id=EJ1003837> (March 26, 2021).
- Blank, R. K. (2012). What Is the Impact of Decline in Science Instructional Time in Elementary School? Paper prepared for the Noyce Foundation. Available at: <http://www.csss-science.org/downloads/NAEPElemScienceData.pdf> (March 26, 2021).
- Block, J. (1971). *Lives through Time*. New York: Bancroft Books.
- Brossard, D., Lewenstein, B., and Bonney, R. (2005). Scientific Knowledge and Attitude Change: The Impact of a Citizen Science Project. *Int. J. Sci. Edu.* 27 (9), 1099–1121. doi:10.1080/09500690500069483
- Chernick, M. R., and Friis, R. H. (2003). *Introductory Biostatistics for the Health Sciences*. New Jersey, United States, Wiley-Interscience. doi:10.1002/0471458716
- Cole, S. (2012). The Development of Science Identity: An Evaluation of Youth Development Programs at the Museum of Science and Industry. Doctoral dissertation. Chicago: Loyola University, Loyola eCommons.
- Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J., and Waller, D. M. (2012). The Impacts of an Invasive Species Citizen Science Training Program on Participant Attitudes, Behavior, and Science Literacy. *Public Underst. Sci.* 22 (6), 745–764. doi:10.1177/0963662511434894
- Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnet, G., Sadler, P. M., et al. (2012). Out-of-school Time Science Activities and Their Association with Career Interest in STEM. *Int. J. Sci. Educ. B* 2 (1), 63–79. doi:10.1080/21548455.2011.629455
- Denson, N., and Ing, M. (2014). Latent Class Analysis in Higher Education: An Illustrative Example of Pluralistic Orientation. *Res. High Educ.* 55, 508–526. doi:10.1007/s11162-013-9324-5
- Fadigan, K. A., and Hamrich, P. L. (2004). A Longitudinal Study of the Educational and Career Trajectories of Female Participants of an Urban

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The data are shared with partner organizations whose permission we would need to share publicly. Requests to access these datasets should be directed to stausn@oregonstate.edu.

AUTHOR CONTRIBUTIONS

NS, KO, and MS contributed to conception and design of the paper. NS organized the information and wrote the first draft of the manuscript. NS and KO wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

- Informal Science Education Program. *J. Res. Sci. Teach.* 41 (8), 835–860. doi:10.1002/tea.20026
- Falk, J. H. (2009). *Identity and the Museum Visitor Experience*. Walnut Creek, CA: Left Coast Press.
- Falk, J. H., Moussouri, T., and Coulson, D. (1998). The Effect of Visitors' Agendas on Museum Learning. *Curator* 41 (2), 107–120.
- Falk, J. H., Reinhard, E. M., Vernon, C. L., Bronnenkant, K., Deans, N. L., and Heimlich, J. E. (2007). *Why Zoos and Aquariums Matter: Assessing the Impact of a Visit to a Zoo or Aquarium*. Silver Spring, MD: American Association of Zoos & Aquariums.
- Falk, J., and Storksdieck, M. (2005). Using the Contextual Model of Learning to Understand Visitor Learning from a Science center Exhibition. *Sci. Ed.* 89, 744–778. doi:10.1002/sce.20078
- Hidi, S., and Renninger, K. A. (2006). The Four-phase Model of Interest Development. *Educ. Psychol.* 41 (2), 111–127. doi:10.1207/s15326985sep4102_4b
- Howard, G. S., and Dailey, P. R. (1979). Response-shift Bias: A Source of Contamination of Self-Report Measures. *J. Appl. Psychol.* 64, 144–150. doi:10.1037/0021-9010.64.2.144
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., and Gerber, S. K. (1979). Internal Invalidity in Pretest-Posttest Self-Report Evaluations and a Re-evaluation of Retrospective Pretests. *Appl. Psychol. Meas.* 3, 1–23. doi:10.1177/014662167900300101
- Jordan, R. C., Gray, S. A., Howe, D. V., Brooks, W. R., and Ehrenfeld, J. G. (2011). Knowledge Gain and Behavioral Change in Citizen-Science Programs. *Conservation Biol.* 25, 1148–1154. doi:10.1111/j.1523-1739.2011.01745.x
- Judson, E. (2012). Learning about Bones at a Science Museum: Examining the Alternate Hypotheses of Ceiling Effect and Prior Knowledge. *Instr. Sci.* 40, 957–973. doi:10.1007/s11251-011-9201-6
- Laursen, B. P., and Hoff, E. (2006). Person-centered and Variable-Centered Approaches to Longitudinal Data. *Merrill-Palmer Q.* 52 (3), 377–389. doi:10.1353/mpq.2006.0029
- Little, T. D., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Allen, P. J., et al. (2020). The Retrospective Pretest-Posttest Design Redux: On its Validity as an Alternative to Traditional Pretest-Posttest Measurement. *Int. J. Behav. Dev.* 44 (2), 175–183. doi:10.1177/0165025419877973
- Magnusson, D. (2003). "The Person Approach: Concepts, Measurement Models, and Research Strategy," in *New Directions for Child and Adolescent Development. Person-Centered Approaches to Studying Development in Context*. Editors S. C. Peck and R. W. Roeser (San Francisco: Jossey-Bass), 2003, 3–23. doi:10.1002/cd.79
- Maltese, A. V., Melki, C. S., and Wiebke, H. L. (2014). The Nature of Experiences Responsible for the Generation and Maintenance of Interest in STEM. *Sci. Ed.* 98, 937–962. doi:10.1002/sce.21132
- Maltese, A. V., and Tai, R. H. (2011). Pipeline Persistence: Examining the Association of Educational Experiences with Earned Degrees in STEM Among U.S. Students. *Sci. Ed.* 95, 877–907. doi:10.1002/sce.20441
- Moussouri, T. (1997). *Family Agendas and Family Learning in Hands-On Museums*. Doctoral dissertation. England: University of Leicester.

- National Research Council (2009). *Learning Science in Informal Environments*. Washington, DC: National Academies Press.
- National Research Council (2010). *Surrounded by Science*. Washington, DC: National Academies Press.
- National Research Council (2012a). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press. doi:10.17226/13165
- National Research Council (2012b). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press. doi:10.17226/13398
- National Research Council (2015). *Identifying and Supporting Productive STEM Programs in Out-Of-School Settings*. Washington, DC: National Academies Press.
- Noam, G. G., and Shah, A. M. (2013). Game-changers and the Assessment Predicament in Afterschool Science. Retrieved from the PEAR institute: Partnerships in education and resilience website: http://www.pearweb.org/research/pdfs/Noam%26Shah_Science_Assessment_Report.pdf. (March 26, 2021) doi:10.4324/9780203773406
- O'Connell, K. B., Keys, B., Storksdieck, M., and Rosin, M. (2020). Context Matters: Using Art-Based Science Experiences to Broaden Participation beyond the Choir. *Int. J. Sci. Educ. Part B* 10 (2), 166–185. doi:10.1080/21548455.2020.1727587
- O'Connell, K., Keys, B., Storksdieck, M., and Staus, N. (2017). *Taking Stock Of the STEM beyond School Project: Accomplishments And Challenges*. Technical Report. Corvallis, OR: Center for Research on Lifelong STEM Learning.
- Oort, F. J. (2005). Using Structural Equation Modeling to Detect Response Shifts and True Change. *Qual. Life Res.* 14, 587–598. doi:10.1007/s11136-004-0830-y
- Overdeest, C., Orr, C. H., and Stepenuck, K. (2004). Volunteer Stream Monitoring and Local Participation in Natural Resource Issues. *Hum. Ecol. Rev.* 11 (2), 177–185.
- Packer, J., and Ballantyne, R. (2002). Motivational Factors and the Visitor Experience: A Comparison of Three Sites. *Curator* 45 (3), 183–198. doi:10.1111/j.2151-6952.2002.tb00055.x
- Packer, J. (2004). Motivational Factors and the Experience of Learning in Educational Leisure Settings. Unpublished Ph.D. thesis. Brisbane: Centre for Innovation in Education, Queensland University of Technology. doi:10.2118/87563-ms
- RK&A, Inc. (2016). *Citizen Science Project: Summative Evaluation*. Technical Report. Washington, DC: Randi Korn & Associates, Inc.
- Saxton, E., Burns, R., Holveck, S., Kelley, S., Prince, D., Rigelman, N., et al. (2014). A Common Measurement System for K-12 STEM Education: Adopting an Educational Evaluation Methodology that Elevates Theoretical Foundations and Systems Thinking. *Stud. Educ. Eval.* 40, 18–35. doi:10.1016/j.stueduc.2013.11.005
- Shaby, N., Staus, N., Dierking, L. D., and Falk, J. H. (2021). Pathways of Interest and Participation: How STEM-interested Youth Navigate a Learning Ecosystem. *Sci. Edu.* 105 (4), 628–652. doi:10.1002/sce.21621
- Sickler, J., Johnson, E., Figueiredo, C., and Fraser, J. (2011). *Language Of Conservation Replication: Summative Evaluation of Poetry in Zoos*. Technical Report. Edgewater, MD: Institute for Learning Innovation.
- Spurk, D., Hirschi, A., Wang, M., Valero, D., and Kauffeld, S. (2020). Latent Profile Analysis: A Review and “How to” Guide of its Application within Vocational Behavior Research. *J. Vocational Behav.* 120, 103445. doi:10.1016/j.jvb.2020.103445
- Stake, J. E., and Mares, K. R. (2001). Science Enrichment Programs for Gifted High School Girls and Boys: Predictors of Program Impact on Science Confidence and Motivation. *J. Res. Sci. Teach.* 38 (10), 1065–1088. doi:10.1002/tea.10001
- Staus, N. L., O'Connell, K., and Storksdieck, M. (2018). *STEM beyond School Year 2: Accomplishments and Challenges*. Technical Report. Corvallis, OR: Center for Research on Lifelong STEM Learning. Available at: <https://stem.oregonstate.edu/sites/stem.oregonstate.edu/files/esteme/SBS-evaluation-report-year2.pdf>.
- Stets, J. E., Brenner, P. S., Burke, P. J., and Serpe, R. T. (2017). The Science Identity and Entering a Science Occupation. *Soc. Sci. Res.* 64, 1–14. doi:10.1016/j.ssresearch.2016.10.016
- Stockmayer, S. M., Rennie, L. J., and Gilbert, J. K. (2010). The Roles of the Formal and Informal Sectors in the Provision of Effective Science Education. *Stud. Sci. Edu.* 46 (1), 1–44. doi:10.1080/03057260903562284
- Storksdieck, M., and Falk, J. H. (2020). Valuing Free-Choice Learning in National parks. *Parks Stewardship Forum* 36 (2), 271–280. doi:10.5070/p536248272 Available at: <https://escholarship.org/uc/item/2z94016m>.
- Trumbull, D. J., Bonney, R., Bascom, D., and Cabral, A. (2000). Thinking Scientifically during Participation in a Citizen-Science Project. *Sci. Ed.* 84, 265–275. doi:10.1002/(sici)1098-237x(200003)84:2<265::aid-sce7>3.0.co;2-5
- Venville, G., Rennie, L., Hanbury, C., and Longnecker, N. (2013). Scientists Reflect on Why They Chose to Study Science. *Res. Sci. Educ.* 43 (6), 2207–2233. doi:10.1007/s11165-013-9352-3
- Young, J. R., Ortiz, N. A., and Young, J. L. (2017). STEMulating Interest: A Meta-Analysis of the Effects of Out-Of-School Time on Student STEM Interest. *Int. J. Edu. Math. Sci. Tech.* 5 (1), 62–74. doi:10.18404/ijemst.61149

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Staus, O'Connell and Storksdieck. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.