Technology Enhanced Outcomes from the NASA Space Science Education Consortium: Impact on Content Knowledge and Attitudes/Dispositions of Learners[1]

Gerald Knezek & Rhonda Christensen
Institute for the Integration of Technology into Teaching & Learning
University of North Texas
December 2020

Introduction

Measures of cognitive and socio-emotional factors important to STEM interest, intent to pursue higher studies, and persistence toward choosing a STEM career – have been refined for two decades. These can be focused more narrowly on the field of space science. This report describes the process of choosing and adapting several measures to conduct research on outcomes of technology-infused learning activities derived through collaborations within the NASA Space Science Education Consortium (NSSEC). A summary of outcomes from four years of research activities, as well as plans for future refinements in research methods, are included.

Supporting Literature

Theories of Learning

The activities created and implemented by the UNT team are largely focused on hands-on, active learning that are relevant to the participants. The activities are based on educational learning strategies such as active, hands-on, relevant, authentic and collaborative activities. Active learning has been shown to improve long-term knowledge retention and deep understanding (Bonwell & Eison, 1991; Christensen & Knezek, 2015; Gallagher, 1997; Akinoglu & Tandogan, 2007). When using the active learning approach, education becomes more personally meaningful and takes advantage of students' natural curiosity. This approach prepares students for the future by having students communicate, collaborate, and try new approaches in finding solutions to real world problems.

Active learning principles are rooted in Dewey's "learning by doing and experiencing" principle (Dewey, 1938). Dewey advocated that a child's schoolwork should have meaning and be engaging as well as have connections to other disciplines and life experiences. In an active learning model, the learner takes more responsibility for his/her own learning under the guidance of a teacher. Characteristics that are included in active learning include:

- relevance to real world applications
- authentic solving of real world problems
- application of prior knowledge and/or experiences to solve new problems

---

[1] This technical report is a prepublication foundation for the more brief *Journal of Computers in Mathematics and Science Teaching* Special Issue article available open source at https://www.learntechlib.org/primary/p/215698/.

- collaboration with others
- integration of subject matters (interdisciplinary) and
- self-directed learning.

Within this context, it is proposed that strategies promoting active learning be defined as instructional activities “involving students in doing things and thinking about what they are doing" (Bonwell & Eison, 1991).

Jonassen, Howland, Moore, and Marra (2003) defined meaningful learning as “occurring when students were actively engaged in making meaning. They broke down this definition into five *interrelated, interactive, and interdependent attributes* with the most meaningful learning activities supporting combinations of these attributes”. Collectively these researchers have established the importance of active, engaged learning in creating learning that is deep and meaningful.

Guiding principles of an active learning paradigm are based on research findings by Griffin (1998) and others that the primary factors enabling learning to occur in informal settings are: a) having a purpose to learn, b) having a choice over learning, and c) having ownership of the learning process in a social context (Griffin, 1998).

The targeted audience for the NASA activities is with middle school aged participants. Middle school is an appropriate age to develop an interest in science that will persist through secondary school, into college and beyond into a career. Providing authentic, active learning experiences contributes to the internalization of learning about science.

Importance of STEM

Choosing a STEM career is a dynamic process that requires both *proficiency* and *interest* in STEM content (Bouvier, 2011; Neathery, 1997). The likelihood of student engagement in learning a specific topic increases when students possess an awareness, positive attitude, and interest in the topic (Jolly, Campbell, & Perlman, 2004).

Enhancing STEM Engagement and Identity. STEM Engagement and Identity have been identified in the literature a important for preserving continuity from interest in STEM, to higher studies in STEM disciplines, to matriculation in a STEM career (citation). Engagement is most closely related to activities that promote initial interest, while Identity evolves along with longer term positive dispositions and produces a sense of belonging (Aschbacher, Ing, & Tsai, 2013). Both engagement and identity can be fostered through formal or informal learning activities, with each having advantages.

Informal learning experiences have the potential to be transformative experiences for learners (Falk & Dierking, 2000; 2013) with impacts extending months and years post experience (Falk & Dierking, 1997; Anderson, Storksdieck, & Spock, 2007). Positive impacts can include acquisition of new skills and content knowledge, increased awareness of STEM, and improved attitudes toward STEM. Informal learning opportunities can be particularly meaningful to children from disadvantaged backgrounds because they most likely have fewer opportunities to participate in these types of activities at home (Hooper-Greenhill, Phillips, & Woodham, 2009). Aschbacher, Ing, and Tsai (2013) found that students who persisted in science, engineering or medical aspirations versus those who dropped out of the pipeline were distinguished by having the opportunity to experience compelling, authentic STEM experiences *outside* of school.

Informal learning is especially well-suited to provide engaging, motivating experiences that can support the development of a STEM identity. In this way, informal learning activities have great potential as transformative learning opportunities and might seed the development of an epistemic frame where the learner develops a professional identity and becomes a member of that domain's community of practice.

The National Research Council has identified six major categories (Learning Strands) in which learners acquire informal science concepts (Bell, Lewenstein, Shouse, & Feda, 2009). These informal science learning strands address: 1) Experiencing excitement, interest, and motivation to learn; 2) Generating, understanding, remembering, and using concepts; 3) Manipulating, testing, exploring, predicting, questioning, and observing; 5) Participating in scientific activities and learning practices with others; and 6) Thinking about themselves as science learners and developing an identity (Sacco, Falk, & Bell, 2014).

Measuring NASA STEM Innovation Activities

Measuring learning in the 21st century is recognized a complex, multidimensional issue involving more than one domain. Learning scientists often study human characteristics that do not belong exclusively to one the traditional domains of psychology: cognitive, affective and psychomotor (ref). The cognitive domain (knowledge, skills, abilities) has traditionally been the focus of outcome measures in STEM disciplines. However in the 21st century, the affective domain has become widely recognized as important as well. In some learning activities germane to space science, such as space walks or maintaining physical fitness during space flights, the psychomotor domain might be very important as well. This paper will focus on the cognitive and affective domains because they are most relevant to broadscale interests in careers in space science.

Cognitive Domain Measures

The academic disciplines blended into the acronym STEM (science, technology, engineering, mathematics) are foundational areas of content knowledge in space science. More specialized areas such as physics and astronomy are mainstream for the field, and these often culminate in even higher specializations such as heliophysics, astrophysics, or astrobiology. For space science both declarative knowledge and procedural knowledge are often important, with the latter relying on step-by-step sequences or algorithms for successful completion of a process. Generally multiple choice, matching or fill-in the blank questions are used to test for declarative knowledge, much like tests for course content in school. Procedural knowledge is often assessed based on successful completion of a sequence of activities required to achieve a goal, complete a mission or a produce a working project. Examples of declarative knowledge assessments used for middle school students by NSSEC are listed in Appendix A. Successfully functioning projects (in robotics competitions or completed NASA missions, for example) are often used as a showcase for how NASA could restructure learning activities.

Socio-Emotional (Non-Cognitive) Measures

Human attributes that learning scientists study outside the realm of cognitive psychology have evolved to have a designation of their own, called non-cognitive variables. Attitudes are one human attribute that have been studied extensively related to technology (Knezek & Christensen, 2008), but there are many others that are now recognized as important as well. The emerging importance of other non-cognitive variables as they relate to teaching and learning with technology, is illustrated by the trend toward studies including many intervening factors that influence whether or not a teacher well trained in technology skills is able to foster the enhancement of 21st Century skills in students. “Noncognitive is used here to refer to variables relating to adjustment, motivation, and student perceptions, rather than the traditional verbal and quantitative (often called cognitive) areas” (Sedlacek, 2011, p. 191). Non-cognitive skills may include self-concept, leadership abilities, creativity, motivation, accurate self-appraisal, empathy, and persistence. Creativity is one example that is spotlighted by others such as Liu (2018) related to creative designers with technology, and Schrier (2018) related to desirable outcomes of game-based learning. Grit (Duckworth & Yeager, 2015; Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013) is another non-cognitive variable that has come to be accepted as very important as an intermediary variable influencing under which conditions technology is effective in enhancing learning. Non-cognitive variables are becoming more valued in the 21st century because they can function in research designs as important intervening or mediating variables that “ … stand between the independent and dependent variables, and […] mediate the effects of the independent variable on the dependent variable (Creswell, 2002, p. 50).”

Many non-cognitive variables are traits deemed desirable in both teachers and students, such as enthusiasm, excitement, and sustained interest. There is some empirical evidence that for these kinds of attributes related to ICT in education, positive or negative valences are transmitted from teacher to student (Christensen, 2002). Such findings have implications for preferred pedagogical style to transmit the motivation to learn from teachers to students, among the alternatives for technological pedagogical approaches to be described in the sections that follow.

Why Measure?

Education projects funded by the U.S. Federal Government often require evaluations to be completed to assess a) process including the number of participants served, number of activities conducted and b) product evaluations that assess impact in one or more psychological domains. These empirical data are used to determine how well the project is achieving its goals and objectives. Research designs are often overlaid on top of evaluation criteria in projects focused on education. For the NSSEC consortium, the BASIK model is used to guide evaluation areas to be addressed. As described in the next paragraph, the constructs represented in the instruments chosen for research on impact are compatible with the BASIK model. The BASIK contextual framework features: **B**ehavior, **A**ttitudes, **S**kills, **I**nterest and **K**nowledge.

These measures are typically given to participants as pre-post measures of what was learned (content) or what dispositions changed as a result of participating in the STEM Innovation activities. The measures we have developed and demonstrated to be effective for NSSEC 1.0 are consistent with the BASIK model. Specifically, we have focused on content knowledge acquisition (K = knowledge), enthusiasm (I = interest), and persistent dispositions (A = attitudes). Since almost all our activities involve hands-on learning, our approach also develops what the BASIK model lists as S = Skills. However these are a byproduct and not directly assessed, but rather indirectly assessed though outcomes such as completion of a product (design and production of a 3D object, for example).

Ways to Assess Learning

Several types of assessment have been used for measuring educational technology activities in the past. Many new types are emerging for 21st Century learners, especially as a result of the affordances provided by new technologies. These forms of assessment include a wide range of methods that vary in expense, invasiveness, and difficulty. These include performance-based, observation, rubric, portfolio, self-assessment, and embedded assessments.

Formative and summative assessments are two traditional categories of assessment that are blurring with the availability of technology enhancements. Formative assessment traditionally has a goal of improving the learning and instruction while summative is in place to judge whether good outcomes have been achieved by use of society's space, time, and money resources. Hickey and Itow (2012) have pointed out that new opportunities exist to assess the abilities of new disruptive and/or innovative technologies

Given that no one form of assessment works for every situation and every learner, a better overview of the depth and breadth of learning might be gained by using more than one type and then combining the information for a more complete evaluation of the learner.

Example Measures Used for Space Science Activities

Multiple choice and short answer questions for space science content.

Well-constructed multiple choice questions have long been regarded as an assessment method that balances the need for efficiency in testing and scoring while still allowing educators to pen questions that tap into higher order cognitive skills in students, such as analysis, synthesis and creativity, according to Bloom's Taxonomy of Educational Objectives in the Cognitive Domain (Bloom et al., 1956). The majority of the content items used in NSSEC research to date have been of this type. Content knowledge assessment might include items such as:

1. An eclipse is defined as an astronomical event that occurs when one celestial object moves _______________ another, partially or fully obscuring it from view.

    A. i**n front of**
    B. next to
    C. in back of
    D. on top of

A short answer question often requires learners to produce knowledge, rather than simply choosing the most correct answer from among those presented. An example of this type of question might be: "Tell how you think the activity in which you participated today increased your knowledge of space science." Short answer questions can often assess richer learning than multiple choice, but short answer are also more difficult to score.

Semantic differentials to measure dispositions toward space science.

Osgood (Osgood, Suci, & Tannenbaum, 1957) is credited with the concept of using adjective pairs as anchors on a 7-point continuum of agreement from 1 to 7. For example, the item might be “To me, Space Science is: Boring __ __ __ __ __ __ __ Interesting.” In this type of instrument, the respondent selects a choice closer to boring or interesting, depending on their perception of the space science. Semantic differential instruments are time efficient and reliable (providing consistent answers). However, this type of assessment may require more instruction to understand how to rate and subjects often mark one end of the spectrum or the other rather than in the middle space. They are commonly used for assessment of socio-emotional variables, where there is no right or wrong answer. One example of a semantic differential used for several NSSEC research studies is listed in Figure 1.

Instructions: Choose one circle between *each* adjective pair to indicate how you feel about the object.

To me, space science is									
1.	Fascinating	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	Ordinary
2.	Appealing	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	Unappealing
3.	Exciting	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	Unexciting
4.	Means nothing	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	Means a lot
5.	Boring	○ 1	○ 2	○ 3	○ 4	○ 5	○ 6	○ 7	Interesting

Figure 1. Semantic differential scale for accessing perceptions of space science.

Likert scales for space science enthusiasm and interest

Likert Scales are among the most common types of items for gathering socio-emotional data, with typical rating choices varying from 1 = strongly disagree to 5 = strongly agree. Some example Likert items related to space science are provided in Figure 2.

Rate each statement on a scale of 1-5, 1=Strongly disagree (SD), 5=Strongly agree (SA)

	SD	D	U	A	SA
	1	2	3	4	5
1. I want to learn more about the moon.	○	○	○	○	○
2. I want to learn more about Mars.	○	○	○	○	○
3. I want to learn more about the sun.	○	○	○	○	○
4. NASA’s Parker Solar Probe mission to the sun will revolutionize our understanding of the sun.	○	○	○	○	○
5. Weather (space weather) that occurs in space can impact my life.	○	○	○	○	○
6. Innovative technologies make learning more engaging.	○	○	○	○	○
7. Innovative technologies help me learn.	○	○	○	○	○
8. I learn better when activities are hands-on.	○	○	○	○	○
9. Using technology to learn gives me more control over my learning.	○	○	○	○	○

Figure 2. Likert items related to interest in space science.

With Likert-type items (as well as semantic differentials), individual items are often very useful for research because they are easy to read and understand, but measurement scales are more powerful for contributing to generalizable and replicable findings. For example, in Figure 2, a researcher might conjecture that items 1, 2 and 3 have a common underlying core as interest in solar system objects. Likewise, items 6, 7 and 9 appear to be about learning with technology. Formal analysis techniques such as factor analysis enable the researcher to determine whether conjectures about scales were correct, after a sizeable set of data has been gathered using items such as those listed in Figure 2. However, the techniques described in the following section allow researchers with smaller data sets such as a single middle school classroom, or participants from two space science summer camps, to confirm whether or not they are justified in combining similar-appearing items into a group to form a scale.

Demographic items

Also adding demographic items like gender, ethnicity, grade level, etc. can allow for additional analysis that can provide feedback to the implementation team on how to best meet the needs of different populations.

Indicators of Validity and Reliability

Validity and reliability are two important concepts for instruments focused on assessment of psychometric attributes related to space science. Especially in socio-emotional areas, a collection of items that form a scale will generally provide more consistent and accurate assessment than one item alone. Validity is concerned with whether the questions being asked of participants are appropriate for (relevant to) what the researcher wants to determine. Reliability has to with consistency of measurement, whether the same opinion or answer would be provided, with respect to a set of items, in the same situation at another time. Reliabilities for Likert and semantic differential scales are commonly estimated by calculating internal consistency reliabilities, where the result is reported as an index from 0 (very low) to 1 (very high). As an example, for the five semantic differential items listed in Figure 1, Cronbach's Alpha for a set of middle school students typically lies in the range of very good to excellent according to guidelines provided by DeVellis (1991) and listed in Table 1. For the set of three items related to sun, moon and Mars in Figure 2, Cronbach's Alpha is typically found to be respectable. For the three technology items in the lower part of Figure 2, Cronbach's Alpha is typically found to be respectable to very good. Readers are referred to conference proceedings for details.

Table 1.

Guidelines for Interpreting Cronbach's Alpha Internal Consistency Reliability for a Psychometric Scale

DeVellis Reliability Guidelines

Below .60	Unacceptable
Between .60 and .65	Undesirable
Between .65 and .70	Minimally acceptable
Between .70 and .80	Respectable
Between .80 and .90	Very good
Much above .90	Excellent (Consider shortening the scale)

(DeVellis, 1991, p. 85)

Note that calculating Cronbach's Alpha only produces estimates of the consistency of a scale of items if they were combined. It is up to the research team to actually produce a scale score for each person (by averaging responses across relevant items) before proceeding with analysis at the scale level.

Adaptation of Feedback Surveys into Research Instruments

Many education outreach and informal science organizations such as the members of the NASA Space Science Education Consortium (NSSEC) distribute feedback surveys at the end of their activities, primarily for formative evaluation purposes. That is, the organizers wish to learn what the participants liked and disliked about the activity just completed, in order to make adjustments before the next group of participants goes through the activity. By making some modifications to a feedback survey and including a research design, it is often possible to expand a feedback survey into a research instrument to create outcome measures that are more robust.

One common modification is to add additional items similar to those already asked, with consistent rating choices for all, so that the reliability of a group of items as a scale can be assessed. This strengthens the instrument itself. Other modifications involve altering administration procedures that result in some basis for comparison of the scores or ratings of the target group after the space science activity has been completed. This can be accomplished by:

a) Having participants complete research instruments before the activity (pre) and after the activity (post); and/or
b) Having a similar group not participating in the activity serve as a comparison against which the anticipated gains in the targeted group can be compared; or
c) Having the instrument administered post test only, but include a retrospective component that asks participants to reflect on their status or state before the activity, and contrast that with ratings after the activity.

Examples of each of these approaches are described in the Findings section of this paper.

Findings Related to Technology Enhanced Space Science Activities

Pre-Post Research

Pre-post assessments have been used in weekend space science camps since 2017. As shown in Table 2, a four-hour Saturday space science camp conducted for sixth grade students in 2019 resulted in significantly ($p < .05$) more positive changes in dispositions toward space science at the time of the post test than prior to camp activities, at the time of the pretest (Christensen et al., 2019). This is based on the semantic differential instrument shown in Figure 1. As shown in the last column of Table 2, the magnitude of the gain for this Saturday camp was effect size = .46, which would be considered moderate according to guidelines by Cohen (1988) and well beyond the effect size = .3 criterion for the point at which the magnitude of the gain becomes educationally meaningful (Bialo & Sivin Kachala, 1996). For this group of informal learners, long term attitudes toward (semantic perceptions of) space science became more positive, when comparing before versus after the weekend space science camp.

Table 2.

Saturday Space Science Camp Pre-Post Changes in Space Science Dispositions Paired Sample t-tests

	Mean	N	Std. Dev.	Sig. (2-tailed)	Effect Size
Space Science Dispositions Pre	6.41	24	.59		
Space Science Dispositions Post	6.67	24	.55	.035	.46

Pre-post research was also used to assess the impact of a similar four-hour Saturday camp in 2017, on knowledge acquisition related to space science. The focus of this camp was solar eclipses, in preparation for the August 2017 Total Solar Eclipse transiting from Oregon to South Carolina across the USA (Christensen, Knezek, Darby, Lepcha, Jiang, Kuo, & Wu, 2017). As shown in Table 3, for the paired pre-post 6th grade students, the gains in content were significant (*p* = .002) and the magnitude of the gains (Cohen's d = 1.12) can be considered educationally meaningful (Bialo & Sivin Kachala, 1996). An effect size of 1.12 represents a very large gain in knowledge about solar eclipses according to the guidelines provided by Cohen (1988) of small = .2, moderate = .5, and large = .8.

Table 3.

Pre-Post ANOVA for Saturday-Camp Participant Eclipse Content Knowledge

	Mean	Std.	N	Sig.	Cohen's

		Deviation			d
Eclipse Score Pre	3.80	1.673	20		1.12
Eclipse Score Post	5.40	1.142	20	.002	

Treatment vs. Comparison Group Research Studies

Post test data were also gathered for the 2017 Saturday space science camp from a comparison group of students who attended the same school and in the same grade level (Christensen, Knezek, Darby, Lepcha, Jiang, Kuo, & Wu, 2017). As shown in Table 4, post test data for the 20 students who attended the Saturday Space Camp were compared to the comparison group who did not attend the camp. The comparison group included 176 students in the same grade level and school. Analysis of variance comparing the treatment group attending the space science camp indicated their average post test score on content knowledge of solar eclipses was significantly (*p* = .0005) higher than the average score the comparison group who did not attend the camp but also completed content knowledge assessments after the camp was completed.

Table 4

Analysis of Variance for Post Test Content Knowledge Score by Attendance vs. Non-Attendance at a Space Science Camp

	N	Mean	Std. Dev.	Sig.
Did not Attend	176	3.77	1.518	
Attended	20	5.40	1.142	
Total	196	3.93	1.563	.0005

The combined findings from Table 3 and Table 4 are illustrated in Figure 3. As shown in Figure 3, the post test scores of the comparison group were very similar to those of the pre-test scores of the treatment group, instilling confidence that both groups would have had similar scores if they had both been tested at the pre-test time. The space science camp attendees, however, showed a significant ($p < .05$) increase in their scores, beginning with an average of 3.80 and increasing to an average of 5.40 questions answered correctly.

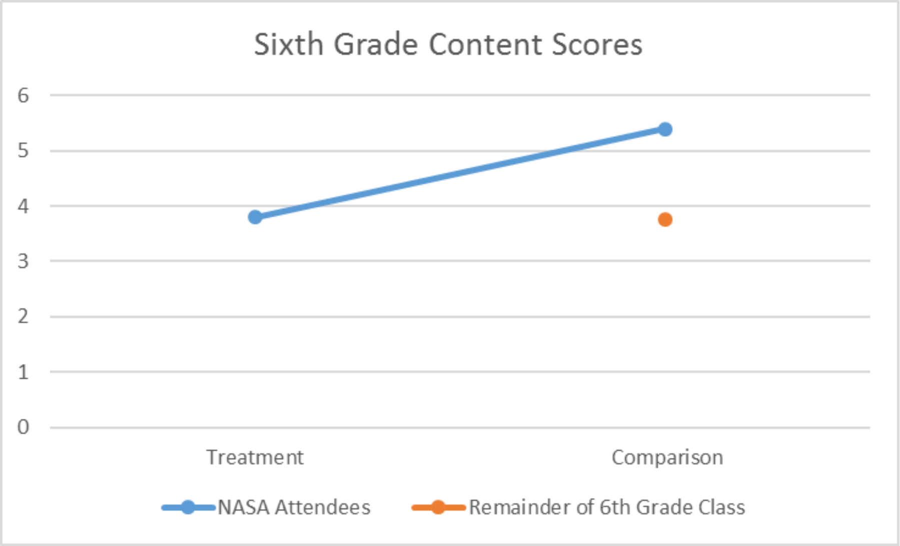

Figure 3. Pre-post data for weekend space science camp attendees vs. comparison group in 2017.

Longitudinal Research

The pre-post analysis for 2017 Saturday space science camp data presented in the previous section was extended by the research team into a longitudinal analysis (Knezek & Christensen, 2018a; 2018b). The 20 students who are listed as the treatment group in Figure 3, were assessed three months after the Saturday space science camp and also after the 2017 Total Solar Eclipse, and their patterns of scores were examined in the context of two other groups who were also scheduled to view the Total Solar Eclipse. Three research questions were the focus of this study (Knezek & Christensen, 2018b):

1. To what extent do students who attended a space science camp gain and retain eclipse-related content knowledge?
2. To what extent do students who attended a space science camp gain and retain enthusiasm for (interest in) a solar eclipse?
3. To what extent do students who attended a space science camp gain and retain positive dispositions toward (long term positive attitudes toward): a) solar eclipses as a space science event, and b) space science in general?

Note that measures of content, disposition and enthusiasm in the cited documents [and the figures for this paper], correspond to Knowledge, Attitude, and Interest in the BASIK framework presented in the literature review section of this paper. The BASIK framework is used for evaluation of the NSSEC program and is therefore incorporated here for cross-walk purposes.

Participants

Included in this research study were data gathered from three middle school sites in 2017, one site that is considered the treatment site and two comparison sites. The participants from the three sites were middle school students from rural, public schools in the vicinity of the university conducting the research.

For Site 1, participants were 20 sixth graders from a rural public school district in the Southwestern US, in which they attended a Saturday half-day event at a nearby university. The 20 students were selected from among approximately 200 sixth graders who wrote short essays about why they wanted to attend a space science camp. The selection committee was composed of a middle school science teacher, technology coordinator, principal and an assistant superintendent. The content portion of the camp began with a four-minute movie explaining the upcoming solar-eclipse, and a brief welcome video from a representative of NASA's STEM Innovation Lab. Students were split into three smaller groups that took part in activities in round-robin fashion. Hands-on activities included: 1) developing a string model of part of the solar system and trying out solar eclipse pinhole cameras as well as NASA-certified solar eclipse glasses; 2) completing 2D and 3D printing of solar eclipse representations and apparatus; 3) interacting with augmented reality space mission apps and virtual reality video segments focusing on space science, and 4) finding their GPS location on NASA Eyes to determine what their view might look like at their home or school on the day of the eclipse. Educators from the student's school accompanied them to the Saturday camp and observed. These activities are described completely in a related publication (Christensen, et al., 2017).

For Site 2, participants included 40 middle school students from a public school district in the Southwestern US who provided baseline data during a brief visit by a university representative who provided solar eclipse glasses, safety information and flyers related to the eclipse, during May 2017. These students were assessed with the same instruments used in Site 1 and Site 3. These students completed the survey instruments once, three months prior to the eclipse event and without any attempt at educational intervention on the topic of space science. These students served as a baseline comparison group for the treatment (intervention) participants at Site 1.

For Site 3, participants included 446 middle school students in the eighth grade of a public school district in the Southwestern US. Students were led by their teachers in discussions of space science and solar eclipse concepts, during the week immediately preceding the solar eclipse on August 21, 2017. These students were assessed with many of the same instruments used in Site 1 and Site 2. They completed the survey instruments toward the end of the week immediately preceding their observation of the eclipse through the viewing glasses led by their teachers. These students served as a second comparison group for the primary treatment participants at Site 1.

Instrumentation

Each participant at Sites 1, 2 and 3 completed a Space Science Survey instrument consisting of demographic items plus: a) a five-adjective, seven-point Semantic Differential scale (example: Boring _ _ _ _ _ _ _ Interesting) with the target of "To me, space science is" (Cronbach's Alpha for these data = .90); b) a five-adjective Semantic Differential scale with the target of "To me, the solar eclipse is" (Alpha = .90); c) a six-item, five-rating point Likert scale (SD to SA) with questions such as "I want to learn more about eclipses." (Alpha = .75); and a seven-item, multiple choice, content knowledge exam with items such as "An eclipse is defined as an astronomical event that occurs when one celestial object moves _______________ another, partially or fully obscuring it from view." The scores include the total number of content items correct.

Complete content items are included in a previous paper (Christensen et al., 2017). Item-level data were combined into scale scores, and these scale scores were used for the analyses presented in this paper. Disposition items related to space science and the solar eclipse are provided in the Appendix.

Trends Across Three Sites

As shown in Table 5 and graphically displayed in Figure 4, the contexts of Sites 1-3 appear to have produced different outcomes with anticipated patterns of measured effects. In particular, the baseline comparison group (Site 2) produced less positive measures than the other two groups in all areas except Enthusiasm for (Interest in) a solar eclipse. This was likely a measure of the excitement produced by the university representative visiting and handing out flyers and NASA-certified solar viewing glasses in advance of the Total Eclipse of 2017, approximately three months before the event. The school at Site 1 is a nearby school district, similar in size and student representation, to Site 2. Site 1 students (treatment), who were selected from 200 of their classmates by school personnel based on essays, were significantly ($p < .05$) higher than the baseline comparison students at Site 2, in their perceptions of Space Science and the Solar Eclipse, even before their space science camp Saturday activity, and well before any follow up discussion of the topic by their teachers. This is likely a measure of Site 1 students' predispositions toward space science, which also resulted in their writing more interesting/meaningful essays and getting selected for the space science camp as one of 20 among 200. Most students at Site 3 completed assessments either the Friday immediately before the Monday eclipse, or on the morning of the event that occurred at 1 pm on the same day. Measured levels of content knowledge and especially perceptions of (dispositions toward) a solar eclipse for these 446 8th grade students were high. This is likely due to the integration of in school with outside of school activities and the proximity to the eclipse event itself. Several additional patterns of findings, with implications for preservice and inservice teacher education are presented in the following sections, based on the site that had pre-post data acquisition.

Table 5.
Descriptive Statistics for Three Space Science Activity Sites

Site	N	Content Knowledge	STD	Space Science Att./Disp.	STD	Solar Eclipse Att./Disp.	STD	Enthusiasm/ Interest	STD
Site 1: Treatment									
Pre Sat. Camp	20	3.55	1.47	6.03	1.52	5.95	1.09	3.22	0.86
Site 1: Post Camp	20	4.60	0.94	5.86	1.41	5.72	1.34	3.81	0.77
Site 1: Follow-Up	20	4.45	1.23	6.48	0.52	6.34	0.62	3.62	1.00
Site 2: Baseline									
Comparison	40	3.37	1.37	4.94	1.45	5.18	1.46	3.70	2.67

Site 3: Baseline	44								
Comparison	6	4.70	1.37	5.30	1.29	5.87	1.15	3.38	0.80

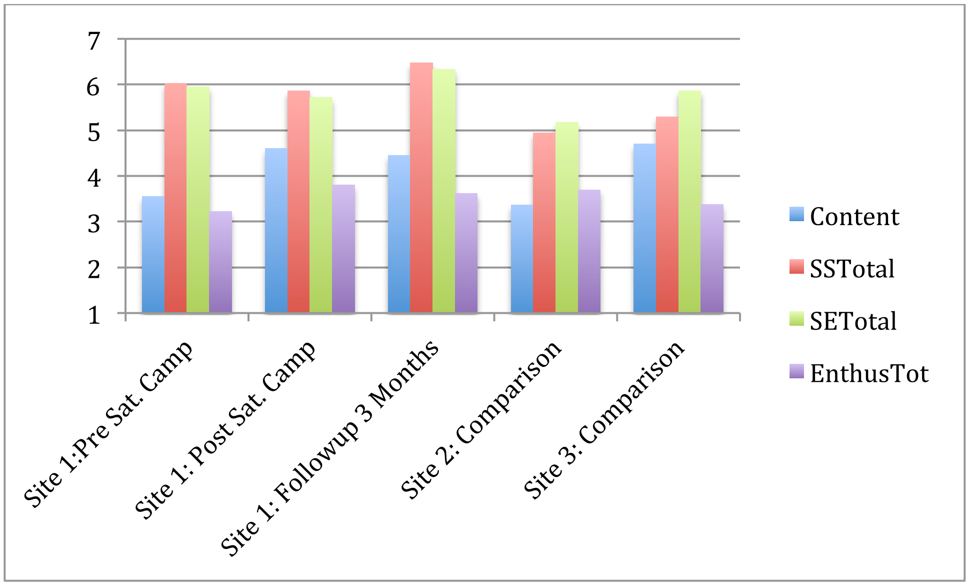

Figure 4. Trends across three sites with assessments at 1-3 time periods.

Longitudinal Findings from Treatment Site Participants

Site 1 included pretest before the Saturday Space Science Camp activity, a post test after the event, and a three month follow up assessment for the same students. Hands-on, technology enhanced, space science learning activities for these middle school students, focused on the topic of a solar eclipse, resulted in large Space Science Content Knowledge gains (p = .002, ES = 1.12), and this knowledge was largely retained three months later (p = .004, ES = .66). These activities also resulted in a significant ($p < .05$) increase in Enthusiasm for a solar eclipse from pretest to post test (p = .03, ES = .54). Enthusiasm waned from posttest time to three months later, to the point where Enthusiasm was still higher than at the time of the pretest, but not significantly higher than pretest or posttest, by the time of the follow up assessment (pretest to follow-up, p = .48 (NS), ES = .25). In addition, these hands-on, technology enhanced, space science learning activities resulted in a significant ($p < .05$) increase in Semantic Perception of a Solar Eclipse from pretest to follow up assessment three months later (p = .03, ES = .67). Semantic Perception of a Solar Eclipse became slightly less positive from pre-camp activities to immediately after, but increased extensively three months later, after which teachers had taken the opportunity to discuss the upcoming eclipse with all students on several occasions. Semantic Perception of Space Science (vs. Solar Eclipse) produced a similar but less strong (NS) profile to that of Perception of a Solar Eclipse. These findings are graphically displayed in Figure 5.

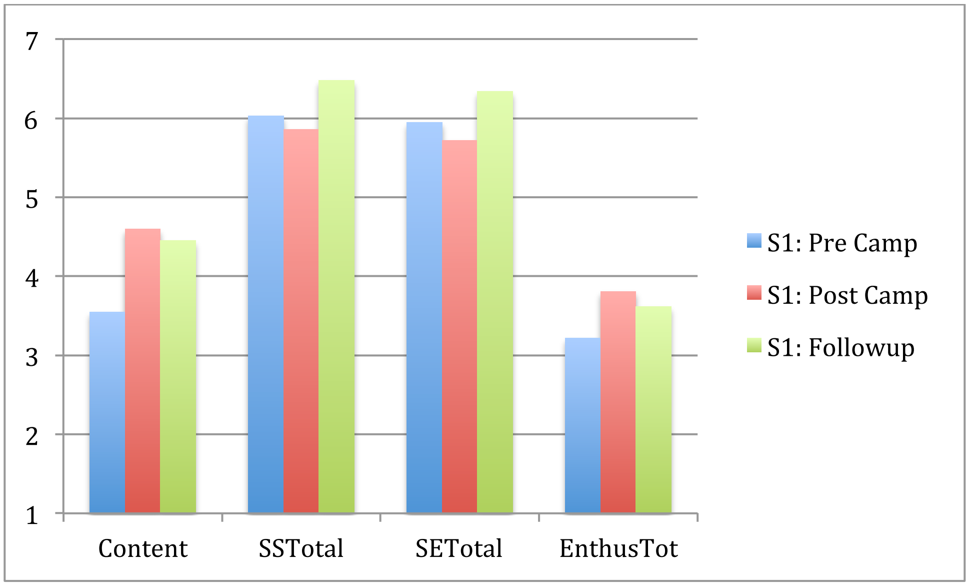

Figure 5. Trends across three time periods for students attending a space science camp.

Summary of Findings

With respect to the original research questions listed for this study, key findings were:

- RQ1. Hands-on, technology enhanced, space science learning activities for middle school students – focused on the topic of a solar eclipse – resulted in large space science content knowledge gains (*p* = .002, ES Site 1 Time 1 to Time 2 = 1.12, ES Time 2 vs. Comparison Site 1 = 1.09) and this knowledge was largely retained three months later (*p* = .004, ES Group 1 Time 1 to Time 3 = .66). See Figure 6 for longitudinal trend.
- RQ2. Hands-on, technology enhanced, space science learning activities for middle school students resulted in a significant (*p* < .05) increase in enthusiasm for a solar eclipse from pretest to post test (Study 1, *p* = .03, ES = .54). Enthusiasm waned from posttest time to three months later, to the point where enthusiasm was still higher than at the time of the pretest, but not significantly higher than pretest or posttest, by the time of the follow up assessment (Study 1 pretest to follow-up, *p* = .48 (NS), ES = .25).
- RQ3. Hands-on, technology enhanced, space science learning activities for middle school students resulted in a significant (*p* < .05) increase in semantic perception of a solar eclipse from pretest to follow up assessment three months later (Study 1, *p* = .03, ES = .67). Semantic perception of a solar eclipse became slightly less positive from pre-camp activities to immediately after, but increased extensively three months later, after which teachers had taken the opportunity to discuss the upcoming eclipse with all students on several occasions. Semantic perception of space science (vs. solar eclipse) produced a similar but less strong (NS) profile to that of perception of a solar eclipse. See Figure 7 for longitudinal trend.

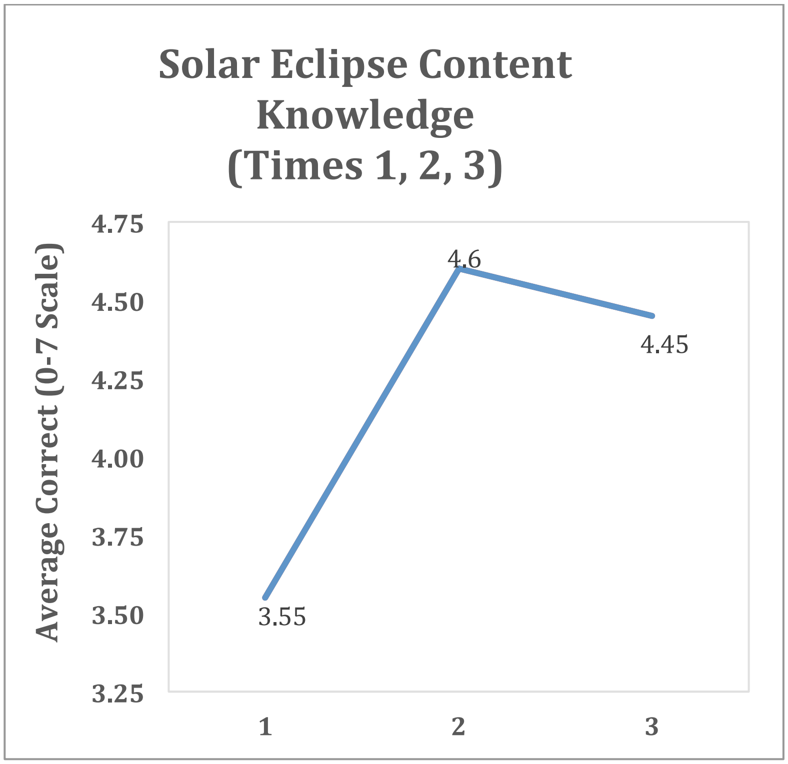

Figure 6. Longitudinal trend for content knowledge retention, T1-T3.

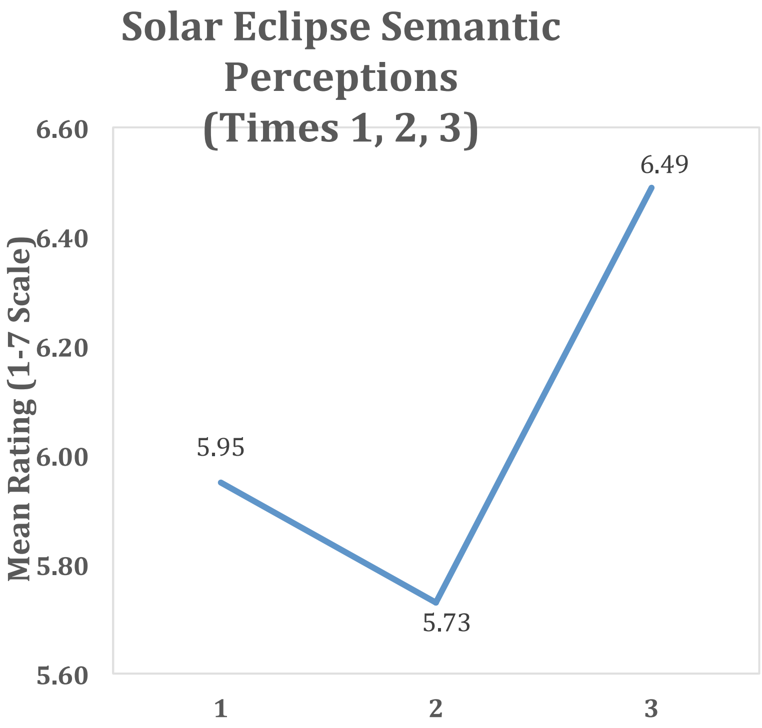

Figure 7. Longitudinal trend for semantic perception of solar eclipse, T1-T3.

Discussion

This longitudinal study focused on technology-enhanced space science opportunities made possible through NASA's educational outreach initiatives in informal learning contexts. In particular, activities and preparation for the 2017 solar eclipse that bisected the entire United States was a focus of learning in both formal and informal environments. The Great American Eclipse of 2017 was NASA's largest education and outreach event in history with 90 million views on the NASA solar eclipse page. Their social media presence also included Twitter, Facebook Live, Snapchat, Instagram, Reddit, Tumblr and NASA TV broadcast which alone had 30 million views. This was the first solar eclipse visible in North America in nearly 40 years and was of interest to formal and informal educational organizations.

The University of North Texas (UNT) was part of a five-year collaborative with NASA Goddard Space Science Center to involve innovative technologies in support of learning about space science. One focus during 2017 was to provide educational learning opportunities about the eclipse, with a larger goal of interesting more students in space science. UNT developed a Saturday camp on campus six months prior to the eclipse to teach middle school participants about the eclipse and the importance of the sun. As the eclipse day neared, more schools and teachers became interested in the opportunity to teach their students about this event. While an event such as the eclipse is not part of the typical middle school curriculum, it is often studied or observed in a more informal learning environment. However, this solar eclipse occurred in these locations during midday on a school day. Therefore, in addition to providing resources and activities for the Saturday camp students and others at their school, resources including solar eclipse viewing glasses as well as safety guidelines were provided via informal means to two additional middle schools to allow their students to participate in the eclipse viewing.

The implications of this study can be summarized as:

- Space science content knowledge gain appears to be rapid and retention is strong as a result of hands-on, technology-enhanced learning activities.
- Enthusiasm for (Interest in) engaging in space science topics accelerates rapidly but tends to wane somewhat over time.
- Dispositions (long term Attitudes) toward a specific space science topic (a solar eclipse), and to a lesser extent the broader field of space science, will increase but need time to grow under nurturing guidance such as that which is often delivered by a teacher in school.
- Informal learning events (weekend or summer camps) combined with formal, ongoing guidance, may produce robust outcomes more lasting than either alone.
- Findings collectively imply that successful implementation of technology-enhanced learning in an informal learning context may require attention to the learner's understanding of the content, enthusiasm for, or interest in the subject, and sustaining long term attitudes or dispositions, at different points in time.

Conclusion of Longitudinal Study

Three middle school sites are compared to assess the measured impact of informal space science education activities on science content knowledge, enthusiasm for an event such as a total solar eclipse, and longer-term dispositions toward space science and a solar eclipse. The space science camp intervention of the treatment group resulted in the greatest overall gains, pre to post, and preserved most of these gains three months later in a follow up assessment. However, analysis of comparison data produced context-enriching findings as well, including that school-based preparation of middle school students for an event such as a solar eclipse, by their teachers, can also produce measureable positive effects. The implications of these findings are that formal learning environments (schools) should seek out opportunities to capitalize on the excitement of informal science events, and integrate these into the traditional school context. The authors propose that the combination could aid in fostering the development of knowledgeable learners at the middle school level with a passion to pursue. Additional research is needed with larger treatment groups to confirm or refute initial findings reported here.

Retrospective Analysis Research

Experimentation with retrospective analysis has been underway with the Saturday camps for middle school students. For example, in February 2020, 20 students attending the 6th grade Saturday space science camp were asked to reply to four retrospective items. As shown in Table 6, all of the ratings when reflecting after the event were higher than perceptions prior. Future research is needed to cross-validate retrospective assessment with more traditional means.

Table 6.

T-Test for Retrospective Items

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	SSinterestBefore1	4.20	20	1.105	.247
	SSinterestAfter1	4.45	20	1.146	.256
Pair 2	SSinterestBefore2	3.15	20	1.387	.310
	SSinterestAfter2	3.45	20	1.356	.303
Pair 3	SSinterestBefore3	4.25	20	1.164	.260
	SSinterestAfter3	4.50	20	1.000	.224
Pair 4	SSinterestBefore4	3.35	20	1.496	.335
	SSinterestAfter4	3.75	20	1.333	.298

Paired Samples Test

		Paired Differences							
		Mean	Std. Dev	Std. Error Mean	95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
					Lower	Upper			
Pair 1	SSinterestBefore1 - SSinterestAfter1	-.250	.550	.123	-.507	.007	-2.032	19	.056
Pair 2	SSinterestBefore2 - SSinterestAfter2	-.300	.733	.164	-.643	.043	-1.831	19	.083
Pair 3	SSinterestBefore3 - SSinterestAfter3	-.250	.851	.190	-.648	.148	-1.314	19	.204
Pair 4	SSinterestBefore4 - SSinterestAfter4	-.400	.598	.134	-.680	-.120	-2.990	19	.008

Acknowledgement

This research was supported in part by the U.S. National Aeronautics and Space Administration (NASA) Grant # NNX16AL63A.

References

Anderson, D., Storksdieck, M., & Spock, M. (2007). Understanding the long-term impacts of museum experiences. In J. Falk, L. Dierking, & S. Foutz (Eds.), *In principle, In practice*. (pp. 197-215).

Aschbacher, P.R., Ing, M., & Tsai, S.M. (2013). Boosting student interest in science. *Kappan Magazine, 95*(2), 47-51.

Bell, P., Lewenstein, B., Shouse, A., & Feder, M. (2009). *Learning science in informal environments: People, places, and pursuits.* Washington: National Academies Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). Taxonomy of educational objectives, handbook I: The cognitive domain. New York: David McKay Co Inc.

Bonwell, C., & Eison, J. (1991). *Active learning: Creating excitement in the classroom.* AEHE-ERIC Higher Education Report No. 1. Washington, D.C.: Jossey-Bass.

Bouvier, S. (2011). *Increasing student interest in science, technology, engineering, and math (STEM): Massachusetts STEM pipeline fund programs using promising practices.* University of Massachusetts Donahue Institute: Hadley, MA.

Bransford, J.D., Brown, A.L., & Cocking, R.R. (Eds.) (2000). How people learn: Brain, mind, experience, and school. Commission on Behavioral and Social Sciences and Education National Research Council. Washington, DC: National Academy Press.

Christensen, R. & Knezek, G**.** (2015). Active learning approaches to integrating technology into middle school science classrooms: Reconceptualizing a middle school science curriculum based on 21st century skills. In X.Ge, D. Ifenthaler, & J.M. Spector (Eds.). *Full Steam Ahead: Emerging Technologies for STEAM*. New York: Springer Academic.

Christensen, R., Knezek, G., Darby, D., Den Lepcha, S., Jiang, B., Kuo, A. & Wu, A. (2017). Outcomes from technology enhanced informal learning activities: Total solar eclipse. In J. Johnston (Ed.), *Proceedings of EdMedia 2017* (pp. 1064-1071). Washington, DC: Association for the Advancement of Computing in Education (AACE). Retrieved January 21, 2018 from https://www.learntechlib.org/p/178499/.

DeVellis, R.F. (1991). *Scale development*. Newbury Park, NJ: Sage Publications.

Dewey, J. (1938). *Experience and education.* A touchstone book. Kappa Delta Pi, New York.

Falk, J.H., & Dierking, L.D. (1997). School field trips: assessing their long-term impact. *Curator 40*(3), 211-218.

Falk, J.H., & Dierking, L.D. (2000). *Learning from museums: Visitor experiences and the making of meaning.* Altamira Press.

Falk, J.H., & Dierking, L.D. (2013). The museum experience revisited. *Left Coast Press.*

Gallagher, S. (1997). Problem-based learning: Where did it come from, what does it do and where is it going? *Journal for Education of the Gifted, 29*(4), 332-362.

Griffin, J. (1998). *School-museum integrated learning experiences in science: A learning journey*. Unpublished PhD thesis, University of Technology, Sydney.

Hooper-Greenhill, E., Phillips, M., & Woodham, A. (2009). Museums, schools and geographies of cultural value. *Cultural Trends, 18*(2), 149 – 183.

Jolly, E., Campbell, P., & Perlman, L. (2004). *Engagement, capacity, and continuity: A trilogy for student success.* Minneapolis, MN: GE Foundation.

Jonassen, D.H., Howland, J.L., Moore, J.L., & Marra, R.M. (2003). *Learning to solve problems with technology: A constructivist perspective.* Merrill Prentice Hall, Upper Saddle River, New Jersey.

Knezek, G., & Christensen, R. (2018a). Capitalizing on informal learning opportunities to enhance formal learning in classroom environments. In E. Langran & J. Borup (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 1920-1926). Washington, D.C., United States: Association for the Advancement of Computing in Education (AACE). Retrieved March 19, 2020 from https://www.learntechlib.org/primary/p/182791/.

Knezek, G., & Christensen, R. (2018b). *Innovative technologies to encourage interest in space science: Impact of solar eclipse activities on students.* Poster presentation to the American Educational Research Association Annual Conference, New, NY.

Neathery, M.F. (1997). Elementary and secondary students' perceptions toward science and the correlation with gender, ethnicity, ability, grade, and science achievement. *Electronic Journal of Science Education* 2(1).

Osgood, C.E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning.* Urbana, IL: University of Illinois Press.

Sacco, K., Falk, J. H., & Bell, J. (2014). Informal science education: Lifelong, life-wide, life-deep. *PLOS Biology, 12*(11), 1-3.