

STEM Learning Ecosystems: Building from Theory Toward a Common Evidence Base

Patricia J. Allen, Zoe Brown, Gil G. Noam

Abstract: An innovative system-building initiative known as the STEM Learning Ecosystems Community of Practice (SLECoP) is transforming U.S. STEM education through cross-sector partnerships between schools, afterschool and summer programs, libraries, museums, and businesses, among others. Although logic models exist to describe how SLEs can make positive contributions toward youth STEM learning in theory, it is unknown how individual SLEs are motivated or equipped to collect the evidence needed to demonstrate their value or abilities to solve the problems they were formed to address. The present study describes the results of a 34-item qualitative survey—completed by leaders of 37 SLEs from four U.S. regions—designed to understand where SLEs are in their evaluation planning, implementing, and capacity-building processes. We found that most SLEs were championed by the extended education sector, and all were highly motivated to conduct evaluation and assessment. Most communities reported a willingness to create a shared vision around data collection, which will help researchers and practitioners track, understand, and improve STEM quality and outcomes in and out of school.

Keywords: STEM Learning Ecosystems, common measures, evaluation, assessment,

STEM Learning Ecosystems: Building from Theory Toward a Common Evidence Base

In everyday life, science, technology, engineering, and mathematics—the subjects collectively known as STEM—capture our sense of wonder. Excitement for STEM sweeps the media during major astronomical events like an eclipse or major breakthroughs like 3-D printed human organs. Yet, inspiring that sense of excitement about STEM among young people in formal educational settings—with the hopes of developing fluency in STEM, building STEM skills, and making STEM majors and careers attractive—is a significant challenge in most countries (OECD, 2010, 2019).

International research has consistently found declining attitudes toward STEM between childhood and adolescence (Potvin & Hasni, 2014a), with fewer students electing to pursue university majors and careers in STEM areas over time (National Science Foundation, 2010). In the U.S., as well as many other industrialized countries, there are also significant

concerns about adolescent mathematics and science literacy and performance, with many students unable to achieve a baseline level of proficiency (OECD, 2019). STEM interest, motivation, and performance are connected to college and career readiness, and evidence from many countries suggests that these outcomes are diminished, at least in part, by a “negative experience of STEM at school” (Joyce & Dzoga, 2012). An international review of STEM outcomes concluded that “...the perception that students have of science might in fact be weakened or held back by the perception they have of ‘school science’...” (Potvin & Hasni, 2014b, p. 99). These concerns have led to a more holistic approach to STEM education, which can be seen through the formation of systems, including the STEM Learning Ecosystems Community of Practice (SLECoP). The SLECoP is changing the belief that STEM learning belongs to one institution by creating interconnected systems to provide diverse STEM learning opportunities (Traphagen & Traill, 2014). However, shared measures and an evidence base is necessary to show the value of such systems to solve the problems they were formed to address (Grack Nelson, Goeke, Auster, Peterman, & Lussenhop, 2019).

The present study examines how the SLECoP embeds evaluation and assessment approaches into its strategies and explores the role that the extended education sector plays in this effort. We begin with a brief review of extended STEM education in the U.S. and the SLECoP. After presenting our research questions and methodology, we summarize key results from a national survey designed to understand where SLEs are in their evaluative planning, implementing, and capacity-building processes. Our conclusions focus on how the extended education sector can be a driving force in the creation of a common evidence base that can track, understand, and improve STEM quality and youth outcomes.

Extended Education and STEM Learning Ecosystems

The importance of educational opportunities occurring outside of the formal school day has increased dramatically in the U.S. over the last decade due to shifting priorities and policies (Afterschool Alliance, 2015). These extended education contexts—which are referred to in the U.S. as out-of-school time (OST) programs—include extracurricular activities at all-day schools, afterschool activities, youth clubs, museum and library programs, and so on. OST STEM learning experiences are attended voluntarily and allow hands-on engagement with a variety of STEM activities in a fun way that sparks curiosity and excitement (Afterschool Alliance, 2015). Considering the different international approaches to extended education, STEM-focused OST programs in the U.S. are characterized by a “hybrid approach” that falls somewhere between free play—reminiscent of programs in countries like Finland and Sweden where children often direct their own leisure time activities in afterschool settings under the supervision of adults—and academic “cram schools”—similar to structured and rigorous programs found in Japan, Taiwan, and Korea that focus on academic achievement to reinforce learning from the traditional school day (Noam & Triggs, 2019).

Providing quality opportunities to explore STEM content outside of formal school settings removes the academic pressure and fear of failure that can contribute to STEM disengagement, even among bright and motivated students (Potvin & Hasni, 2014a). It also supports positive youth development—including fostering quality relationships with peers and adults among other social skills—by offering a safe place for children to learn and play

when their primary caregivers are at work or otherwise unavailable (Noam & Triggs, 2019). There is growing evidence that participation in high-quality, STEM-focused OST programs can positively change youth attitudes related to STEM engagement, identity, career interest, and career knowledge (Allen et al., 2019; Chittum, Jones, Akalin, & Schram, 2017; Dabney et al., 2012; Sahin, Ayar, & Adiguzel, 2013; Wulf et al., 2010; Young, Ortiz, & Young, 2017).

Education researchers and practitioners are now searching for ways to expand the availability of high-quality OST STEM programming to be more academically supportive without mirroring the school day or “cram schools.” The STEM Learning Ecosystems Community of Practice (SLECoP) Initiative, founded in 2015 by the STEM Funder’s Network, aims to meet this need for high-quality, inspiring STEM learning opportunities by developing meaningful cross-sector partnerships (STEM Learning Ecosystems, 2020). The SLECoP engages pre-K–12 schools and school districts, afterschool and summer programs, colleges and universities, libraries, museums, businesses, and home environments in cities, states, and regions across the U.S. (Traill & Traphagen, 2015). The initiative’s broad aim is to deepen STEM learning among children and youth, build capacity among educators, provide professional development and assessment tools, and create communities of practice to share experiences and promote best practices (STEM Learning Ecosystems, 2020). A recent case study described how implementation of the SLECoP’s strategies has strengthened partnerships between the extended education sector (e.g., OST programs) and many other community sectors to create a “surround sound of STEM” that provides more educational and workforce opportunities and pathways into STEM (Allen, Lewis-Warner, & Noam, 2020).

The logic behind this collaborative approach is that a successful and sustainable STEM learning ecosystem (SLE) will cultivate high levels of interest and motivation that can play a significant role in building STEM skills and career aspirations (Maltese & Tai, 2011, Traphagen & Traill, 2015). STEM careers are linked to social and economic mobility of individuals, families, and communities, and having a skilled STEM workforce supports international global competitiveness (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 2007). For this reason and others, participation in the SLECoP was recently identified as a key priority in the U.S. government’s five-year federal STEM strategic plan (National Science & Technology Council, 2018). There are now 89 SLEs across 35 U.S. states, and new SLEs are beginning to launch internationally (i.e., Canada, Israel, Kenya, and Mexico). As the initiative scales, there is a need for focused research and evaluation of SLECoP efforts to understand what strategies are changing, and ideally improving, youth outcomes in STEM.

Study Aims and Hypotheses

Using a qualitative survey, our first study aim was to understand whether SLE leaders are implementing evaluation and assessment in a purposeful, systematic way. Toward this goal, we mapped the assessment landscape (i.e., the methods, types, and systems of data collection) to characterize evaluation efforts across communities. Our second study aim was to explore SLE leaders’ motivations for evaluation and assessment—to understand if communities were being driven by top-down requirement or bottom-up choice—and to identify

any obstacles related to evaluation and assessment. Our third study aim was to discover how SLEs prioritized the use of assessments that are common across communities (i.e., a national vision) compared with the use of assessments that address specific needs of their local community (i.e., a local vision). The survey culminated with this question because a common vision around data is an essential ingredient for the success of complex collective impact initiatives like the SLECoP: “To truly evaluate their effectiveness, collective impact leaders need to see the bigger picture...rather than attempting to isolate the effects and impact of a single intervention, collective impact partners should assess the progress and impact of the changemaking process as a whole...” (Parkhurst & Preskill, 2014, p. 17).

Given that the SLE leaders had shared experiences and received common guidance as part of the SLECoP, we hypothesized that they would report similarly high levels of interest and motivation in evaluation and assessment but have different opportunities due to different local conditions. We also expected that SLEs would be at different stages of implementation; as SLEs mature, they will have stronger partnerships, resources, and funding to support evaluation and assessment efforts in their community. Lastly, we predicted that most SLEs would be willing to adopt common assessments to measure the SLECoP’s collective impact—and that some may already be using shared measures—but that some questions and issues are locally-based. We planned to examine the variety of reasons for and against common measures based on local conditions.

Method

This section describes the participants, measures, procedure, and data analyses used to examine evaluation and assessment strategies among SLE leaders from all 37 communities that joined Cohort 1 of the SLECoP, which began in 2015.

Participants

Most SLE leaders (65.7%) reported being directors/executive directors of extended education programs (e.g., city’s zoo or aquarium), networks (e.g., state afterschool system-builder), or councils (e.g., technology-focused advisory board). The sample also represented responses from CEOs/vice presidents/presidents at STEM-expert institutions such as science centers (10.5%), program managers/coordinators of extended education programs (13.16%), an educational consultant (2.63%), a senior research scientist (2.63%), an educational policy advisor (2.63%), and a superintendent of a K-12 school district (2.63%).

SLEs represented a great variety of regional economies and education systems with 19 states, plus Washington, D.C., from all four U.S. regions as defined by the U.S. Census Bureau: Northeast (29.7%), including Maine, Massachusetts, New York, Pennsylvania, and Rhode Island; Midwest (18.9%), including Illinois, Michigan, Missouri, Ohio, and Indiana; South (18.9%), including Florida, Maryland, Oklahoma, Washington, D.C., Texas; and West (24.3%), including Arizona, California, Colorado, New Mexico, and Oregon. Additionally, Cohort 1 SLEs represented cities of all sizes ranging from small (e.g., Augusta, ME and Camarillo, CA), to mid-size (e.g., Providence, RI and Salem, OR), to large (e.g.,

Chicago, IL and New York City, NY) as defined by National Center for Education Statistics (NCES) classifications and criteria.

Measures

A 34-item survey was designed to understand where SLEs are in their evaluation planning, implementing, and capacity-building processes. Questions included a mixture of qualitative open-ended questions (“What do you hope to learn by using data collection tools in your ecosystem?”), discrete categorical questions (e.g., a question about an evaluation-related action or resource followed by the response options “No,” “No, but considering,” or “Yes”), and quantifiable Likert-type items (e.g., rating the usefulness of various data collection tools on a scale of 1-4, ranging from “Not at All Useful” to “Very Useful”).

To understand the progress SLEs have made with their evaluation plans, SLE leaders were asked about actions that could advance their evaluation and assessment-related objectives (“Has your ecosystem hired an evaluator to help you measure STEM program quality or student outcomes?”).

To understand resources that SLEs have at their disposal, SLE leaders were asked about data collection tools and systems potentially in use in their communities (“Please tell us the names of the data collection tool(s) you are CURRENTLY using [within the last 2 years] to evaluate STEM learning in your community.”).

To understand drivers and hinderers of evaluation and assessment, SLE leaders were asked questions related to their motivations (“What are your ecosystem’s goals for evaluation and assessment?”), expectations (“What do you hope to learn by using data collection tools in your ecosystem?”), and obstacles (“Are there any challenges to evaluation or assessment in your community?”).

To understand the disposition of SLEs to build a common evidence base for the SLECoP, leaders were asked about their communities’ willingness to adopt a shared vision of evaluation and assessment (“How willing do you think your partners would be to use data collection tools that are common across ecosystem [to look at ecosystem development, program effects, and youth impacts]?”).

Procedures

Survey items were drafted and revised in consultation with the Teaching Institute for Excellence in STEM (TIES), the educational consulting organization that provides leadership and technical assistance for the SLECoP. The leaders of each SLE were contacted by email and asked to voluntarily complete a survey that asks questions about evaluation and assessment in their SLE. Survey responses were collected over an eight-week period, and all SLE leaders contacted answered the survey. There were two communities where two responses were received from co-leaders of the SLE. The survey was not designed to probe attitudes or content knowledge that would require psychometric properties or normative comparisons. Instead, we designed a broad survey of the plans, practices, and procedures of SLEs to advance their evaluation and assessment agendas.

Data Analysis

Quantitative data from Likert questions were analyzed using SPSS (version 24) to determine descriptive statistics among variables. Qualitative data from open-ended responses were analyzed thematically using a recursive six-phase process (Braun, Clarke, & Rance, 2015): (1) becoming familiar with the data; (2) assigning preliminary codes; (3) searching for patterns or themes; (4) reviewing themes; (5) defining and naming themes; (6) writing up the theme.

All study procedures were reviewed and approved by the institutional review board at our home institution.

Results

Our key findings are organized around the following questions: 1) How intentional and systematic are SLE evaluation practices?; 2) What are the motivations and challenges around SLE evaluation and assessment?; and 3) Are SLEs willing to adopt a shared vision around evaluation and assessment?

Question 1: How intentional and systematic are SLE evaluation practices?

We found high levels of interest and intention among SLE leaders to pursue evaluation and assessment within and across SLEs. One indication was the 100% response rate (n=37 SLEs) to this survey about evaluation and assessment, but the context shared by SLE leadership explicitly linked evaluation and assessment with the achievement of their communities' goals. Leaders frequently reported the goal of understanding the effectiveness of different educational models and strategies within SLEs. As a northeastern school superintendent said: "As we continue to develop innovative STEM learning models that emphasize hands-on, mind-on 'project-based' learning to foster curiosity, questioning, creativity and innovation, meaningful assessments of and for learning are vital." Many SLEs were also interested in data collection to support diversity, inclusion, and equity in STEM, especially in terms of youth awareness of college and career pathways. A midwestern executive director of an OST program reported the need to use data to understand "...how to best build out STEM learning pathways in a city with diverse community assets, needs and supports." Others also expressed the importance of adopting evidence-based practices as part of their strategy, as a northeastern executive director of an OST program shared: "The degree to which we can collectively embrace evidence-based practice is dependent upon the strength of our evaluation system."

The evidence showed that there were SLEs that were collecting data in a purposeful, systematic manner, however it was clear that some SLEs were more advanced in terms of implementation than others based on specific actions reported (e.g., hiring an evaluator, partnering with other sectors to conduct evaluations, using data collection tools, adopting a data management system).

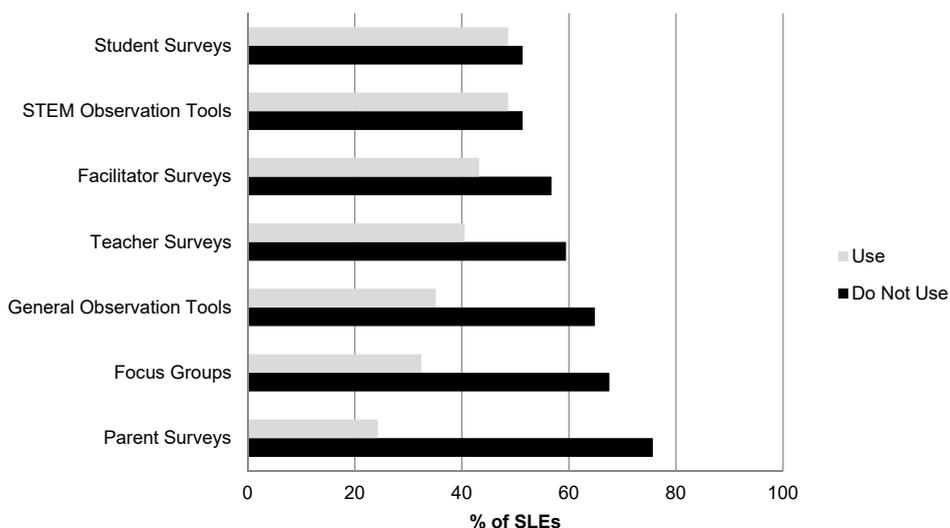
Table 2. List of Most to Least Commonly Used Assessments by Type as Reported by STEM Learning Ecosystem Leadership

General Observation	STEM Observation	Student Survey	Teacher Survey	Facilitator Survey	Parent Survey	Other
Weikart Center Youth Program Quality Assessment (YPQA)	PEAR Dimensions of Success (DoS)	PEAR Common Instrument Suite- Student Survey	Teaching Rubrics (e.g., Project-Based Learning, PBL/Projected Based Teaching Rubric, State Rubrics)	PEAR Common Instrument Suite- Educator Survey	Interviews of families (e.g., to find out about learning from OST STEM resources)	Student achievement data/testing (e.g., state mandated testing)
Thoughtful Classroom Teacher Effectiveness Framework	Youth Program Quality Assessment (YPQA) for STEM	Holistic Student Assessment	ExpandED Schools Educator Survey— Collaboration among Educators	ExpandED Schools- Educator Survey— Collaboration among Educators	Homegrown/local	Network analysis
Assessment of Program Practices Tool (APT)	Thoughtful Classroom Teacher Effectiveness Framework	National Assessment of Educational Progress (NAEP)	Friday Institute Teacher Efficacy and Attitudes toward STEM Survey (T-STEM)	Easy CBM (Benchmarking and Progress Monitoring System) - Fluency and Comprehension		Group and individual interview (e.g., focus groups)
University of Cincinnati Evaluation Services Center Observation Tool	University of Cincinnati Evaluation Services Center Observation Tool	Student Attitudes toward STEM Surveys (S-STEM): MISO/ North Carolina State/Friday Institute	Survey of Academic and Youth Outcomes- Teacher Version (SAYO-T)	Homegrown/local		Landscape Survey— Needs and Wants of STEM Community
Onsite visits by a Program Evaluator	Homegrown/local	Survey of Academic and Youth Outcomes Youth Survey (SAYO Y) State-wide knowledge-based assessments (e.g., M-STEP in Michigan)	Science Teaching Efficacy Belief Instrument (STEBI) - STELAR - EDC			District preparedness surveys (e.g., Carnegie Science Center STEM Pathways Assessment tool)
Policy Studies Association, Inc. (PSA) OST Observation Instrument		STEM Readiness Self-Assessment	Federal Annual Performance Report - Teacher Survey			Pre-screen check-lists
The Danielson Group Observation Template (modified)		University of Cincinnati Evaluation Services Center - Student STEM Survey	Pedagogical Content Knowledge (PCK) Assessment			
Homegrown/local		Homegrown/local	Homegrown/Local			

We found that about 43% of SLE leaders reported they had hired a formal evaluator to coordinate and collect data on STEM program quality or student outcomes, with another 35% of SLE leaders considering evaluator options (see Table 2). Additionally, about 37% of SLE leaders reported a partnership with their local school districts to conduct evaluations of STEM learning, with about 40% actively considering this possibility. Cross-sector partnerships are another action that more advanced SLEs take to advance evaluation and assessment. A midwestern STEM coordinator noted: “We’re still in the early stages of figuring out how we will work with several school districts around the state. Certainly, evaluation will be an element of those relationships, but it’s too early to define what exactly those practices will be.” As an example of an SLE in advanced stages of planning, a midwestern program manager offered: “All [OST] programs include evaluation elements and school partners are involved in gathering and reviewing STEM learning data, to continually strengthen and increase the impact of STEM programs.”

We next mapped the assessment landscape to find out how many SLEs were collecting data from K-12 schools, OST programs, or other sectors, and if they had used observation tools, self-report survey tools, and other data collection tools within the last two years. Approximately 50% (n = 18 SLEs) reported using at least one kind of data collection tool, but the percentage of SLEs using each type of tool varied from about 24% to 48% (see Figure 1). STEM-focused program quality observation tools (48.6% of SLEs) and student self-report surveys (48.6% of SLEs) were two of the most commonly used tools across the SLEs, while parent surveys, general non-STEM observation tools, and focus groups/interviews were the least common (see Figure 1).

Figure 1. Percent of STEM Learning Ecosystems using Data Collection Tools¹



1 The percentage of different types of data collection tools that are used and not used by ecosystems. Data are organized from most frequently used tool to least frequently used tool.

SLE leaders named more than 40 unique measures that were being used to assess either STEM program quality, student attitudes, or content knowledge. Additionally, about 40% of SLEs reported creating their own assessment tools (Table 2). The quality of the measures varied greatly; some demonstrated strong psychometric properties and were published in peer-reviewed journals, while others were developed for local evaluation purposes but were not psychometrically tested. Examples of higher-quality (vetted) observation tools include the Assessment of Afterschool Program Practices Tool (APT) (Tracy, Surr, & Richer, 2012) and the Dimensions of Success (Shah, Wylie, Gitomer, & Noam, 2018), which focus on general and STEM-specific aspects of OST quality, respectively. Examples of higher-quality (vetted) student self-report surveys that assess STEM-related attitudes include the Common Instrument Suite for Students (CIS-S) survey (Allen et al., 2019; Sneider & Noam, 2019) and the Student Attitudes toward STEM (S-STEM) survey (Unfried, Faber, Stanhope, & Wiebe, 2015). In addition to STEM attitudes, there was interest among SLEs to capture social and emotional attitudes and skills using surveys such as the Holistic Student Assessment (HSA) survey (Malti, Beelmann, Noam, & Sommer, 2018; Noam, Malti, & Guhn, 2012) and the Survey of Academic and Youth Outcomes-Youth Version (SAYO-Y) survey (Stavsky, 2015).

Table 1. Actions Taken or Considered by STEM Learning Ecosystem Leadership to Reach Evaluation and Assessment Goals

Action	No, and not considered	No, but considering	Yes
Hired an evaluator	8 (21.62%)	13 (35.14%)	16 (43.24%)
Partnered with school district(s)	8 (21.62%)	15 (40.54%)	14 (37.84%)
Developed local assessment tools	10 (27.02%)	11 (29.73%)	16 (43.24%)
Adopted a data management system	13 (35.14%)	20 (54.10%)	4 (10.81%)

When asked how useful each category of data collection tool was for informing decisions, leaders were unanimous in their beliefs that all data collection tools were useful (when combining ratings for “Somewhat Useful” and “Very Useful”). STEM-focused program quality observation tools were rated the most useful and one of the most commonly used data collection tools across SLEs (see Figure 2). When asked to elaborate on usefulness, four practices emerged: 1) ensuring STEM learning opportunities and resources were equitably distributed to reach and support for all youth; 2) measuring progress made to achieve the SLEs’ specific goals around improving STEM learning outcomes for youth; 3) identifying underperforming programs so that SLE collaborators can share support, resources, and strategies; and 4) estimating the long-term impact of SLE partnerships on students’ college and career readiness.

Lastly, we asked about the adoption of data management systems and found that only about one in ten SLEs (10.8%) reported a system dedicated to collecting SLE data. SLEs that reported having an existing data management system used City Span, The Connector, or CAYEN Afterschool. SLEs that did not have a data management system but were ex-

ploring options reported interest in Pivotal Tracker, Qualtrics, and Salesforce, in addition to those previously mentioned. Although a majority of SLEs are just beginning to think about how to collect, store, and track data, several ecosystems noted that this was “not an if but when” scenario. A western policy advisor noted: “We are working on statewide tracking and monitoring of STEM outcomes and may seek additional investments. Our Legislature has funded a statewide longitudinal database project with visualizations and we will likely utilize that when complete.” Among the few that reported having an established system, there was the sentiment that the system was “...not used to its full potential” and that there was not a single format for all settings and sectors within the SLE.

Question 2: What are the motivations and challenges around SLE evaluation and assessment?

When SLE leaders were asked about their motivations to develop an evaluation and assessment strategy, two common themes emerged: 1) Demonstrate value to stakeholders and prospective partners to secure increased funding, capacity, cross-sector partnerships, and improve educational strategies; 2) Inform continuous improvement and provide program- and system-level decision-makers with realistic ideas for effective interventions or gradual modifications to programs, curricula, standards to improve youth STEM-related outcomes.

The remaining motivations were: 3) Demonstrate impact and assess the effectiveness of interventions or modifications to programs, curricula, standards as well as the effectiveness of partnerships cultivated through the ecosystem. Leaders suggested that the demonstration of impact was important on both a program/sector-level and a SLE/regional-level. A southern K-12 STEM education director specified: “We would like to measure quality of programs, shifts in beliefs about STEM careers and STEM learning, and quantitative measures such as numbers of students impacted and increases in assessment scores.” At the regional-level, an executive director of a STEM-focused council in a southern SLE indicated that they wanted to “...better understand the larger impacts created by the partnerships cultivated through our ecosystem...[and also] to learn how our training impacts [community service organizations’] efforts in their school and in the community.” 4) Ensure high-quality student learning experiences (e.g., strong minds-on activities, exposure to STEM career options, etc.). An executive director of an OST program leading a New England SLE noted: “We have been assessing the quality of the programs using both the PQA and the DoS... We also use the SAYO T and Y to assess the practices that undergird next generation science standards.” 5) Improve educator effectiveness by using data from educators and students to target professional development and resources for teachers and OST facilitators on local as and regional levels. For instance, a western executive director of an OST program is motivated “...to determine if STEM Ecosystems move the dial on improving teacher practice and students’ achievement.” 6) Improve access to STEM learning opportunities by increasing the number of cross-sector partnerships and STEM learning opportunities. A scientist leading a northeastern SLE performed evaluations “To characterize the ways in which the ecosystem uses partners, resources, and STEM Guides (brokers) to connect youth to out-of-school STEM opportunities.”

When asked about barriers to the implementation of SLE evaluation and assessment strategies, several themes emerged. One commonly reported challenge was a lack of infrastructure, with many SLE leaders citing a need for a common, centralized data management

system to collect, store, analyze, and report data findings. Most SLE leaders were unsure or had not yet thought about what kind of infrastructure was needed because they were not yet fully established. According to a midwestern STEM coordinator: "...we're very early in figuring out what our evaluation processes will be and how the data will be collected and filed."

Another related challenge included limited resources (i.e., time and money). Many SLE leaders felt additional resources were necessary to achieve their evaluation and assessment goals. A western executive director of a museum noted that they needed "dedicated staff and resources to ensure success" whereas a northeastern director of an OST intermediary reported that "...it takes a significant amount of time and effort to achieve goals; this can obstruct the formation of partnerships and development of relationships between programs and schools." Several leaders also cited a need for affordable assessment tools that can be used from year to year that are shared across the SLEs and better access to student data (both in school and outside of school). A midwestern senior program director noted that "securing funding is difficult in and of itself..." and another had a related thought that: "Systems such as databases, portals, and dashboards are expensive to create, and even more expensive to maintain; many funders are more interested in supporting programs made directly available to students as opposed to organizational capacities." When asked what would better enable their community to advance evaluation and assessment of STEM learning, a western director of education for a science museum itemized the following solutions: "1. A pre-existing infrastructure that works for our community for data collection and sharing...2. Dedicated staff and resources to ensure success."

Another common barrier was obstacles within formal educational systems. For example, in school settings, performance-based measures can overshadow other types of assessments. An executive program director of a northeastern SLE lamented: "Within the K-12 education system, high stakes testing has soured many discussions related to assessment and evaluation...conversations can very quickly disintegrate into a debate over state assessment linked to school performance." There are also limitations with schools and programs sharing student-level data due to existing privacy and confidentiality policies. A western STEM director acknowledged that: "Data privacy issues are pervasive and recent policy makes it harder to collect student level data from districts." Challenges with data sharing/privacy, lack of a clear/shared vision for evaluation and assessment, as well as limited resources and competition for funding reduce the strength of partnerships within ecosystems. A northeastern executive director cautioned that SLEs "...need to be mindful of the time and administrative burden that we ask of teachers and program staff to invest in surveying youth...That can be an obstruction to developing partnerships/relationships between programs and schools."

Lastly, leaders indicated that there are many different tools currently in use for some organizations, which makes it difficult to compare across organizations. There is a general feeling that there is a need for shared metrics and assessment tools to perform evaluations well. For instance, a northeastern program director indicated that there was "no common data system, no commonly defined goals or metrics for STEM learning." A western director of STEM initiatives noted: "Useful assessment and evaluation always require...carefully selected instruments upon which the various constituencies agree and approve, and the development of a common language/purpose of assessment."

Question 3. Are SLEs willing to adopt a shared vision around evaluation and assessment?

The data indicate that SLEs favor a shared vision around measurement. In response to the question “How willing do you think your partners would be to use data collection tools that are common across ecosystems (to look at ecosystem development, program effects, and youth impact)?” the majority of SLE leaders (about 89%) reported that their partners would be somewhat willing (n = 21 SLEs) or very willing (n = 12 SLEs) to share data collection tools. Table 3 provides brief summaries of the common reasons for why SLE leaders believe that their partners would be willing or unwilling to use shared measures across the SLECoP. For those willing to use shared measures, the most common reasons are wanting standardized measures that can communicate between SLEs and to funders and a shared metric that will cultivate cross-sector and regional partnerships to support STEM pathways and opportunities for youth.

The data revealed a few concerns related to a shared vision among a minority of SLEs, with about 11% of SLEs reporting that their partners would be somewhat unwilling (n = 3 SLEs) or very unwilling (n = 1 SLE) to use the same tools that are used by others across the SLECoP. The most common explanations for a hesitation to adopt shared measures relate to a lack of resources, difficulty obtaining alignment among schools/programs, commitment to current tools already in use, and uncertainty about the reliability or validity of available tools (Table 3).

Table 3. Partners’ Willingness to Use Data Collection Tools that are Common Across STEM Learning Ecosystems

Rating	Reasons for Willingness or Unwillingness
Very Willing	<ul style="list-style-type: none"> Partners have a desire to standardize evaluation tools to better collect data and share metrics with stakeholders. Some funders have pre-existing research requirements that may interfere with introducing a new tool.
	<ul style="list-style-type: none"> Partners have a strong commitment to implementing common data collection tools. Many partners are eager and excited to be able to track and measure progress. Ecosystems have a vision and desire to cultivate cross-sector and regional partnerships to support youth pathways and generate opportunities for growth
Somewhat Unwilling	<ul style="list-style-type: none"> There are concerns about the time and energy it takes to attain alignment across all schools. The quality and type of tools would have to be better than the established suite of tools already being used by some ecosystems. Ecosystems feel it is challenging to convince people who are resistant to change to get on board.
	<ul style="list-style-type: none"> Limited resources and a lack of information about return on investment impedes implementation

Although most leaders saw value in creating a shared vision around evaluation and assessment, they also favored collecting data that are specific to local conditions to ensure that the needs of the local community are met. For example, a southern director of a STEM alliance noted: “The Network wants to ensure that we have easily accessible and actionable data for the [SLE] that will empower the Network and its partners to meet the needs of the commu-

nity.” A southern executive director of a STEM council pointed to the agenda of local foundations: “Right now most of our funders are not pushing this sort of [common] evaluation as much as what they get out of things related to their individual organization. This is however something we are internally interested in doing.” The topic of a shared vision around evaluation and assessment is actively under consideration in SLEs, as one midwestern executive director of a STEM alliance shared: “Currently, the primary STEM organizations utilized a variety of evaluation tools. There have been conversations around common language for STEM learning...”

Discussion

While the formation of STEM learning ecologies began in the U.S., the SLECoP has now grown into an international movement with the recognition that the challenges to STEM are a global concern that communities must work together to solve. The approach goes beyond the boundaries of formal school settings to include afterschool and summer programs, libraries, museums, businesses, and homes, among other sectors. The present study showed that evaluation and assessment is being built into SLE strategies to support their growth, effectiveness, and sustainability, but as expected, SLEs are in different phases of their planning and have different capacities. Several SLEs with sophisticated implementation strategies have significant potential to inform the field about national trends in STEM teaching and learning. Other SLEs early in their evaluation planning show promise but need additional support from the research and evaluation communities to build infrastructure and capacity to support data collection. Importantly, a positive culture around evaluation and assessment is developing across the SLECoP; SLEs are highly motivated to collect data and most endorse a shared vision around evaluation and assessment.

There were high levels of interest and motivation among ecosystem leaders to tackle the challenge of evaluation and assessment, as evidenced by the 100% response rate and in-depth answers to open-ended questions. Some primary motivators include: demonstrating value of programming to stakeholders; using data to guide the process of implementing system- and program-level changes, assessing program impact on student outcomes; ensuring quality of student learning experiences; using data to improve teaching effectiveness; and increasing quality STEM learning opportunities. The SLE leaders' motivations translated into action, with about half of SLEs having made a commitment to data collection and actively collecting data from K-12 or OST STEM learning environments. The group of SLEs that are not yet ready to collect data still indicated that evaluation and assessment are important and useful.

Additionally, our mapping of data collection tools shows that the OST field has gone beyond the collection of simple demographic data. About half of SLEs were collecting data on STEM program quality and outcomes—with most self-report surveys focused on understanding youth STEM interest, motivation, and attitudes. There was less interest in collecting data from other informants, including educators and parents, although SLEs reported parent surveys as most useful, suggesting that interest in family engagement is rising. There was also little interest in performance measures that assess content knowledge and STEM skills. Sev-

eral SLE leaders noted negative perceptions of standardized assessments in their communities because there is a belief that overassessment is a problem in formal school settings.

Although data management platforms are becoming less expensive and more user-friendly, we found that only about 10% of SLEs are using a data management system. One way to create a common data management system across the SLECoP is to choose some measures that make it possible to have shared outcomes while allowing some customization to address questions specific to local contexts. While common measures and systems may help advance the evaluation and assessment goals the SLECoP, there is a risk for strong opposition to using common measures in favor of using measures in common. A “measures in common” approach would rely on more complicated and less precise secondary analysis strategies (e.g., meta-analysis) that require more time and statistical expertise to analyze and interpret. If the goal is to have a management tool where stakeholders can see how the whole initiative performing, then there is a need for some common instrumentation or a hybrid model. This means SLEs can only ask a limited set of questions of youth, as there is only so much time and space on a survey; therefore, a common database management system would need to include short versions of different measures so that those interested in different assessments can choose more than one.

This leads to the pivotal question of whether SLE leaders believe their partners would be willing to create a shared vision around data collection. Consistent with our hypothesis, most respondents—about 90%—reported that they believed their partners would be willing to use common data collection tools. Those reporting their SLE was in favor of a shared vision cited the need for standardized measures that can track progress and facilitate communication and learning among educational stakeholders. The 10% of SLE leaders that were hesitant about their partners’ willingness to adopt shared tools and evaluation plans cited a lack of resources, resistance to changing measures already in use, difficulty obtaining alignment among schools and programs, and uncertainty about the reliability or validity of available tools. It will be important to address existing concerns so that all rally around a shared vision.

When considering SLEs’ implementation approaches we found that some communities had more capacity and experience than others. The data point to three distinct clusters: 1) an advanced level—exemplified by a northeastern SLE that has an evaluator, a clear evaluation plan, established data collection tools and management system, and has performed evaluations across sectors (in school and outside of school) for several years prior to joining the SLECoP; 2) an intermediate level—exemplified by a midwestern SLE that has defined goals and evaluation plans, a collaboration with a research institute, has piloted data collection tools, and has started small scale evaluations within the last year; and 3) a beginner group—exemplified by a southern SLE that has started the early stages of planning/goal setting around evaluation and assessment, but has not used data collection tools, an evaluator, or a data system.

SLEs reported that the most common barriers to progress in this area is a lack of infrastructure and funding to support the expenses associated with evaluation and assessment. Building a common data management system that is readily available to SLEs could improve capacity to collect data from many young people and reduce expense. There would still be a need for interpretation by independent evaluation or research experts, who tend to

charge premium rates, but the costs would be substantially lower. However, SLEs need to search for opportunities to finance this critical infrastructure.

Conclusions and Recommendations

The present study of the U.S. SLECoP demonstrates the significant interest and investment of SLEs in evaluation and assessment. The extended education/OST sector will be important to consider as the growth of STEM learning opportunities outside of schools has become a phenomenon across the world. The present results suggest that the OST sector can become a champion of data collection within many of the SLEs, especially as many SLEs are led by organizations representing the OST sector and much of the data being collected is from OST programs.

There are also international implications for a shared vision around evaluation and assessment. The SLECoP now includes four international ecosystems—in Canada, Israel, Kenya, and Mexico—and the initiative is rapidly scaling. While we expect to find challenges very similar to those we described here, one important difference is that the American education system is localized. Every state has different educational standards, different data privacy/confidentiality requirements, different assessments, and different data systems, which may make evaluation more challenging in the U.S. relative to other countries that have national data collection systems.

The willingness of most SLEs in the U.S. to use common measures represents a major opportunity for the SLECoP initiative to build rich international datasets, to examine the strength of interventions and effects, and to track individual and collective progress (i.e., for children and ecosystems). If connected to the research community, this could advance the STEM education fields' understanding of effective models and approaches. The first step toward improving readiness and capacity for data collection will be to identify funding opportunities to provide needed infrastructure, including a centralized online data management system desired by SLE leadership. Time is of the essence to avoid fragmentation in approach as SLEs begin to seriously consider selecting an existing system or developing their own. A shared vision needs to maintain flexibility to balance local goals and metrics with national/international goals and metrics. One solution may be to support SLEs in conducting in-depth case studies at the local level to preserve local context and innovation, in addition to implementing shared measures for a few quantifiable outcomes that are prioritized by SLEs. Having different grain sizes of analysis will simultaneously help build the evidence base for the SLECoP and generate new hypotheses to improve and innovate STEM teaching and learning.

Acknowledgments

The authors thank the STEM Next Opportunity Fund for funding this research. We also want to express our sincere thanks to the Teaching Institute for Excellence in STEM (TIES) for its leadership in STEM education and for the opportunity to conduct research on the STEM Learning Ecosystem Community of Practice (SLECoP).

References

- Afterschool Alliance. (2015). *Full STEM ahead: Afterschool programs step up as key partners in STEM education*. Afterschool Alliance. <http://www.afterschoolalliance.org/AA3PM/STEM.pdf>
- Allen, P. J., Lewis-Warner, K., & Noam, G. G. (2020). Partnerships to transform STEM learning: A case study of a STEM learning ecosystem. *Afterschool Matters*, 31, 30-41.
- Allen, P. J., Chang, R., Gorrall, B. K., Waggenpack, L., Fukuda, E., Little, T. D., & Noam, G. G. (2019). From quality to outcomes: A national study of afterschool STEM programming. *International Journal of STEM Education*, 6 (1). <https://doi.org/10.1186/s40594-019-0191-2>
- Braun, V., Clarke, V., & Rance, N. (2015). How to use thematic analysis with interview data (process research). In A. Vossler & N. Moller (Eds.), *The counselling and psychotherapy research handbook*. SAGE.
- Chittum, J. R., Jones, B. D., Akalin, S., & Schram, Á. B. (2017). The effects of an afterschool STEM program on students' motivation and engagement. *International Journal of STEM Education*, 4 (1). <https://doi.org/10.1186/s40594-017-0065-4>
- Dabney, K. P., Tai, R. H., Almarode, J. T., Miller-Friedmann, J. L., Sonnert, G., Sadler, P. M., & Hazari, Z. (2012). Out-of-school time science activities and their association with career interest in STEM. *International Journal of Science Education, Part B*, 2 (1), 63-79. <https://doi.org/10.1080/21548455.2011.629455>
- Grack Nelson, A., Goeke, M., Auster, R., Peterman, K., & Lussenhop, A. (2019). Shared measures for evaluating common outcomes of Informal STEM education experiences: Shared measures for evaluating common outcomes. *New Directions for Evaluation*, 2019 (161), 59-86. <https://doi.org/10.1002/ev.20353>
- Joyce, A., & Dzoga, M. (2012). *Intel White Paper: Science, technology, engineering and mathematics education—Overcoming challenges in Europe*. Intel.
- Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education*, 95 (5), 877-907. <https://doi.org/10.1002/sce.20441>
- Malti, T., Beelmann, A., Noam, G. G., & Sommer, S. (2018). Innovation and integrity in intervention research: Conceptual issues, methodology, and knowledge translation. *Prevention Science*, 19 (3), 271-273. <https://doi.org/10.1007/s11211-018-0868-7>
- National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. The National Academies Press. <https://doi.org/10.17226/11463>
- National Science & Technology Council. (2018). *Charting a course for success: America's strategy for STEM education*. 48.
- National Science Foundation. (2010). *Integrated postsecondary education data system completions survey [data file]*. WebCASPAR: <https://caspar.nsf.gov/>
- Sneider, C., & Noam, G. G. (2019). The Common Instrument Suite: A means for assessing student attitudes in STEM classrooms and out-of-school environments. *Connected Science Learning*, 11. [csl.nsta.org/2019/07/the-common-instrument-suite](https://doi.org/10.1002/csl.nsta.org/2019/07/the-common-instrument-suite)
- Noam, G. G., Malti, T., & Guhn, M. (2012). From clinical-developmental theory to assessment: The Holistic Student Assessment tool. *International Journal of Conflict and Violence*, 6 (2), 201-213. <https://doi.org/10.4119/UNIBI/ijcv.276>
- Noam, G. G., & Triggs, B. B. (2019). Expanded learning: A thought piece about terminology, typology, and transformation. *International Journal for Research on Extended Education*, 6 (2-2018), 165-174. <https://doi.org/10.3224/ijree.v6i2.07>
- OECD. (2010). *PISA 2009 results: What students know and can do – Student performance in reading, mathematics and science* (Vol. 1). OECD Publishing.

- http://www.oecd-ilibrary.org/education/pisa-2009-results-what-students-know-and-can-do_9789264091450-en
- OECD. (2019). *PISA 2018 Results (Volume I): What Students Know and Can Do, PISA*. OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- Parkhurst, M., & Preskill, H. (2014). Learning in Action: Evaluating Collective Impact. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/evaluating_collective_impact
- Potvin, P., & Hasni, A. (2014a). Analysis of the decline in interest towards school science and technology from grades 5 through 11. *Journal of Science Education and Technology*, 23 (6), 784-802. <https://doi.org/10.1007/s10956-014-9512-x>
- Potvin, P., & Hasni, A. (2014b). Interest, motivation and attitude towards science and technology at K-12 levels: A systematic review of 12 years of educational research. *Studies in Science Education*, 50(1), 85-129. <https://doi.org/10.1080/03057267.2014.881626>
- Sahin, A., Ayar, M. C., & Adiguzel, T. (2013). STEM related after-school program activities and associated outcomes on student learning. *Educational Sciences: Theory & Practice*, 14 (1). <https://doi.org/10.12738/estp.2014.1.1876>
- Shah, A. M., Wylie, C., Gitomer, D., & Noam, G. G. (2018). Improving STEM program quality in out-of-school-time: Tool development and validation. *Science Education*, 102 (2). <https://doi.org/10.1002/sce.21327>
- Stavsky, S. (2015). Measuring social and emotional learning with the Survey of Academic and Youth Outcomes (SAYO). Wellesley, MA: National Institute on Out-of-School Time.
- STEM Learning Ecosystems. (2020). STEM Learning Ecosystems. *Cultivate. Learn. Innovate*. <http://stemecosystems.org/>
- Tracy, A., Surr, W., & Richer, A. (2012). *The assessment of Afterschool Program Practices Tool (APT): Findings from the APT validation study*. National Institute on Out-of-School Time.
- Traill, S., & Traphagen, K. (2015). Assessing the impacts of STEM learning ecosystems: Logic model template & recommendations for next steps. *Noyce Foundation*. <http://stemecosystems.org/>
- Traphagen, K., & Traill, S. (2014). How cross-sector collaborations are advancing STEM learning. *Noyce Foundation*. <http://stemecosystems.org/resource-category/key-resources/>
- Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment*, 33 (7), 622-639. <https://doi.org/10.1177/0734282915571160>
- Wulf, R., Mayhew, L. M., Finkelstein, N. D., Singh, C., Sabella, M., & Rebello, S. (2010). Impact of Informal Science Education on Children's Attitudes About Science. 337-340. <https://doi.org/10.1063/1.3515238>
- Young, J. R., Ortiz, N., & Young, J. L. (2017). STEMulating Interest: A meta-analysis of the effects of out-of-school time on student STEM interest. *International Journal of Education in Mathematics, Science and Technology*, 5 (1), 62. <https://doi.org/10.18404/ijemst.61149>