Measure STEM Caliper Development Initiative

J Morrison[1] **and W P Fisher Jr**[2]

[1] Teaching Institute for Excellence in STEM, Cleveland, OH, USA
[2] LivingCapitalMetrics LLC, Sausalito CA, & BEAR Ctr, UC Berkeley, CA, USA

E-mail: jmorrison@tiesteach.com

Abstract. Ecologized learning combines educational resources across diverse in-school, out-of-school, and after-school programs. The distributed and multilevel complexity encountered when learning is situated in its authentic contexts poses daunting challenges to measuring, communicating, managing, and improving outcomes. Measure STEM is the umbrella under which multiple data aggregation, statistical summary, and measurement efforts in the 89 international STEM Learning Ecosystems have been brought together. Though STEM typically focuses on Science, Technology, Engineering, and Mathematics, a wide range of educational programs across areas encompassing the arts, coding, social skills, literacy, and other fields are also included under the STEM heading. Educators, policymakers, funders, and other key stakeholders convinced of the value of cross-sector collaborations defined five pillars of ecosystem success: partnerships, measures, administrative systems, teaching and learning practices, and workforce development connections. The ecosystem success construct was mapped, existing items from an earlier ecosystem assessment were revised and augmented, four parallel forms with common items were devised, and an online administration system was created in early 2020. A data set of over 300 responses became available in early February, and will be reported in April. The measurement system design, calibrations, stakeholder comparisons, interpretation guidelines, and reports are reported in a partner presentation by the same authors.

1. Introduction

A growing movement in education aims to ecologize instruction, which means situating learning in its authentic contexts of use [1]. In the context of STEM (science, technology, engineering, and mathematics, which often includes arts, communications, and other areas, as well), students engage in out of school and after school activities in zoos, museums, aquariums, insectariums, parks, workplaces, internships, etc., where they can experience the practical value of knowledge in use. Community partners working across sectors must not only align and coordinate their efforts, they must work at new, deeper levels to leverage in-, after-, and out-of-school opportunities to provide more students with quality learning. For more information, see https://stemecosystems.org/.

Begun in 2015 by the STEM Funders Network at the meeting of the Clinton Global Initiative America, the STEM Learning Ecosystems have rapidly expanded in the U.S. and abroad [2]. As of February, 2020, there are 89 STEM learning ecosystems in five countries (Canada, Israel, Kenya, Mexico, and the U.S.) involving almost 1,900 school districts, 33 million pre-k to 12th grade children,

---

[1] To whom any correspondence should be addressed.

over 1,200 out-of-school partners, 850,000 educators, and over 4,000 business, philanthropic, and industry partners.

For three years (2016-2018), an Ecosystem Indicator Tool (EIT) was administered annually to obtain information on ecosystem performance in five areas deemed the pillars of ecosystem success: administrative systems, measurement, teaching and learning; partnerships, and workforce development (see Figure 1). EIT items were, unfortunately, not written with the intention of achieving meaningful measurement. Efforts aimed at constructing measures from the three years of available data failed [3].

The EIT was then redesigned as Caliper, a measure of ecosystem success defined in terms of stakeholder empowerment. The construct was mapped [4-5] on the basis of the intended object of measurement built into the EIT but not clearly articulated there. A total of 161 new survey items derived from the EIT were written to be rated on a six-point scale ranging from Very Strongly Disagree to Very Strongly Agree, and a Caliper calibration study was launched in early 2020.

Over 2,500 stakeholders were invited to respond to one or more of four forms. Forms were linked by 31 common items, with another 32 or 33 items specific to each. Items were distributed across the forms proportionately from each of the five pillars of success, and from each of five hypothesized levels in the ecosystem success construct (Table 1).

The calibration study will evaluate the dimensionality of the measured construct within and across ecosystems and stakeholder groups. Do items pertaining to the different pillars and the success levels work together to measure the same thing? Do the items measure the same thing across different stakeholder groups and ecosystem locations? If not, what constructs are empirically supportable, and can they be theoretically justified?

2. The EIT redesign

The EIT items were written with the intention of incorporating into each item every different ecosystem's specific variations in administration and focus. The items were then multivalent: pointing in several different directions at the same time. It was impossible to reproduce from any given response which specific combination of possible subjects, verbs, and objects had been addressed. Separating out each subject-verb-object combination in any given item may result in as many as eight or more different statements. There was no way to tell from the data if respondents had consistently averaged across all possible permutations, or if they had instead based their answers on selected subsets of the multiple questions asked.

Even rating scale choices on the EIT were multivalent, as with one item's rating of 4 labelled "Infrastructure is robust, placing many educators, and growing to meet demand, and is iterating in response to in-field observations." When the rating options also present multiple possible single rating criteria, and even more combinations of rating option pairs and triples, the number of possible items represented in any single rating can be 20 or more.

EIT items were furthermore not written from a construct map with an intention of defining a range of variation from less to more. The number of items was determined by considerations of content coverage and respondent burden, not by reverse engineering from the precision needed to support a decision process. Considered from the alternative perspective of measurement theory, a definitive bank of relevant items could be developed, calibrated, and adaptively administered according to the needs of individual ecosystems and stakeholders, to support the precision demands of the relevant management decision processes.

Finally, though Not Applicable and No Opinion response options were not used, all items were associated with five categories instead of an even number that forces respondents to make a decision about disagreeing or agreeing with each assertion about ecosystem circumstances.

Given (a) the predominance of items asking multiple questions, (b) rating categories specifying multiple responses, (c) no deliberate intention of posing questions varying in their expression of a construct from least to most, and (d) no concern for measurement uncertainty relative to variation, it was not likely that data from the EIT would fit a model requiring separable parameters. Multiple analyses supported that conclusion [3].

3. Construct model formulation and expected scaling results

Previous research across a number of various domains involving surveys of customers, employees, patients, students, faculty, parents of students with disabilities, etc. led to the formulation of a theory of relationship development [6-8]. That theory was augmented by a review of the literature that revealed an independent body of research [9-13] focused on what may be the same construct, though it was not formulated in relation to measurement theory and data.

Elbaum and Fisher [7-8] categorized phases in the development of improved partnerships between schools and parents of children with disabilities in terms of Access, Involvement, Dialogue, Choices, and Impact. The literature on government efforts to cultivate local citizen population took up much the same theme, but in terms of Informing, Consulting, Involving, Collaborating, and Empowering. These latter labels, referred to as ICICE in the aggregate, are employed in the Caliper Development Initiative. Table 1 is an initial effort at defining an ICICE construct map.

Caliper will be the medium through which the ICICE construct is implemented. Reports will be provided at three levels representing

1. the overall accountability of the ecosystem at the construct map [4] level,
2. the managerial focus on quality improvement measurement comparisons across stakeholder groups and ecosystems facilitated in terms of Wright maps [14-17], and
3. support for both individual stakeholder and stakeholder group learning processes reported in modified kidmaps [15-22].

Concerning the latter, means and standard deviations of Caliper measures by ecosystem will be augmented by group-level kidmaps showing rating response patterns for stakeholder groups. Special strengths and weaknesses evident in these patterns will be communicated along with recommendations for action. Formative assessment principles will be applied to guide advancement along the ICICE continuum. Variations in construct definition across ICICE item groups and/or across stakeholder groups may necessitate some combination of rating scale, multidimensional, item bundle, or mixture measurement models [4, 24-26].

The overall trajectory for formative applications will proceed as follows:

1. Having (a) INFORMED the Measure STEM stakeholders of basic measurement principles and how each original EIT item could be parsed into varying numbers of specifically focused statements, and
2. having (b) CONSULTED the Measure STEM stakeholders as to their sense of the ecosystem development narrative communicated by the revised EIT (now Caliper) items,
3. the stage is set for (c) INVOLVING the Measure STEM stakeholders in the pilot calibration of those items.

Upon completion of this pilot and the successful calibration of the Caliper STEM Ecosystem measurement infrastructure, the Caliper Development Initiative will transition into a new phase of (d) COLLABORATION, in which Caliper will be applied as a shared formative and summative resource informing Lead STEM Capstones, process and outcome communications at convenings, and quality improvement efforts in general. As these collaborations mature with the blossoming of individual micro-level, stakeholder and local meso-level, and ecosystem-wide macro-level applications efficiently and transparently integrating processes and outcomes, (e) EMPOWERMENT will be manifest in a variety of forms as an abundance of fulfilled potential. .

4. Conclusion

Calibrating Caliper at a high level of measurement separation reliability in relation to a predictive construct theory sets the stage for supporting meaningful and reproducible decision processes across the STEM learning ecosystems. To paraphrase the famous observation of Fujio Cho [27], past Chairman of Toyota Motor Corporation, the Caliper measurement infrastructure is intended to facilitate brilliant results from everyday people managing a well-designed process, instead of producing only typical results from brilliant people managing a broken process. Broadly speaking, the effect is to advance cultural progress by improving the quality of the cognitive supports scaffolded by the knowledge infrastructure [28-29]. These decision processes will informed, then, by a shared knowledge infrastructure distributing a multilevel scaffold of cognitive supports, following previous suggestions [30-33]. Pilot results are reported in an associated paper [34].

Table 1. Stakeholder Participation and Empowerment Construct Map

LEVEL	STAKEHOLDER PARTICIPATION GOAL	PROMISE TO THE STAKEHOLDERS	EXAMPLE TOOLS
5. EMPOWER	To place final decision-making in the hands of stakeholders	What you decide is implemented	Authority Responsibility Accountability Returns on investments Information infrastructure
4. COLLABORATE	To partner with the stakeholders in each aspect of the decision including the development of alternatives and the identification of the preferred solution	We will look to you for direct advice and innovation in formulating solutions and incorporate your advice and recommendations into the decisions to the maximum extent possible	Advisory Committees Consensus-building Participatory decision-making Infrastructure R&D
3. INVOLVE	To work directly with stakeholders throughout the process to ensure that issues and concerns are consistently understood and considered	We will work with you to ensure that your concerns and issues are directly reflected in the alternatives developed and provide feedback on how your input influenced the decision	Workshops Deliberate polling
2. CONSULT	To obtain feedback on analysis, alternatives and/or decisions	We will keep you informed, listen to and acknowledge concerns and provide feedback on how your input influenced the decision	Public comment Focus groups Surveys Public Meetings
1. INFORM	To provide balanced and objective information to assist in understanding problems, alternatives and/or solutions	We will keep you informed	Factsheets Websites Open Houses

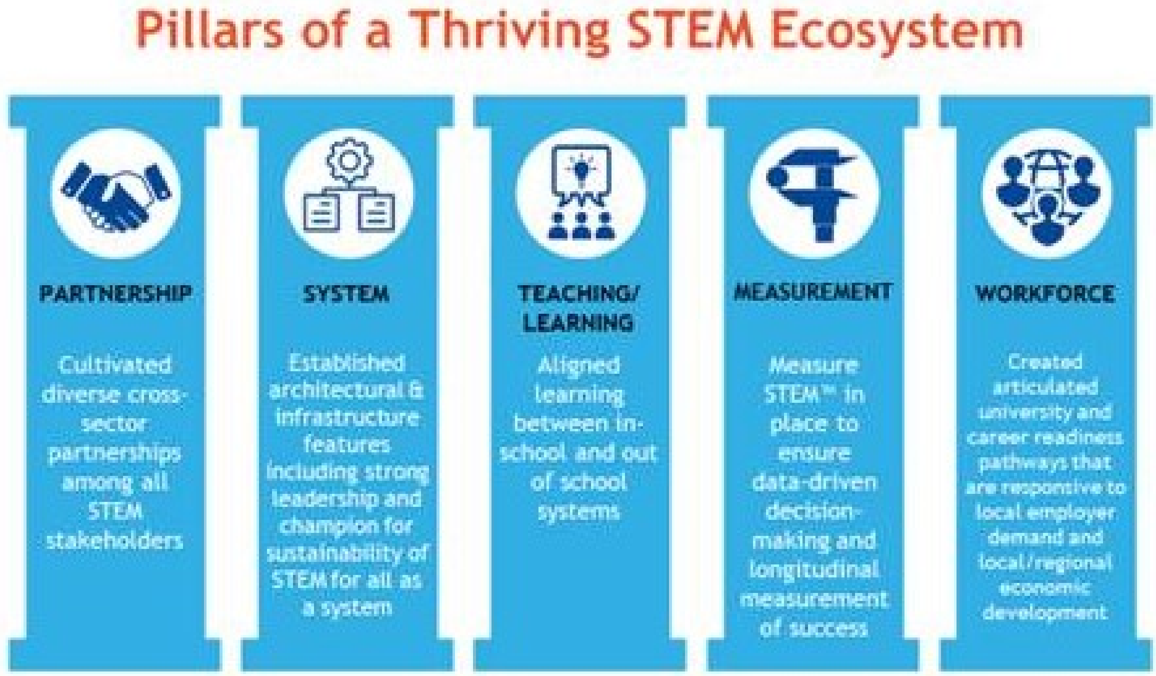

Figure 1. Pillars of Thriving STEM Ecosystems

5. References

[1.] Barab S A and Roth W M 2006 *Educational Researcher* **35** 3-13
[2.] Morrison J and Fisher W P Jr 2018 *J Physics: Conf Series* **1065**(022009)
[3.] Morrison J and Fisher W P Jr 2019 *J Physics: Conf Series* **1379**(012042)
[4.] Fisher W P Jr 2006 *Rasch Meas Trans* **20** 1072-1074
[5.] Wilson M 2005 Constructing measures (Mahwah, New Jersey: Lawrence Erlbaum)
[6.] Elbaum B 2014 *J Disability Policy Studies* **24** 206-217
[7.] Elbaum B, Fisher W P Jr and Coulter W A 2011 *J Appl Meas* **12** 261-278
[8.] Fisher W P Jr. Elbaum B and Coulter W A 2012 *J Appl Meas* **13** 57-76
[9.] Arnstein S R 1969 *J American Institute of Planners* **35** 216-224
[10.] Chappell B 2008 *Community engagement handbook: a model framework for leading practice in local government in South Australia*. Government of South Australia. https://www.dpti.sa.gov.au/__data/assets/pdf_file/0003/152535/Community_Engagement_Handbook.pdf
[11.] Jami A A and Walsh P R 2017 *Energy Res & Sol Sci* **27** 14-24
[12.] Jankowski P 2009 *J Environmental Management* **90** 1966-1971
[13.] Ortiz D and Huber-Heim K 2017 *Internat J Management Educ* **15** 318-331
[14.] Wilson M R 2011 *Rasch Meas Trans* **25** 1331
[15.] Black P, Wilson M and Yao S 2011 *Measurement: Interdiscip Res & Persp, 9*, 1-52
[16.] Boone W J 2016 *CBE--Life Sciences Educ* **15** 1-7
[17.] Holster T A and Lake J W 2015 *Shiken* **19** 32-41.
[18.] Mead R J 2009 *J Appl Meas* **10** 208-224.
[19.] Wright B D, Mead R J and Ludlow L H 1980 KIDMAP: person-by-item interaction mapping, MESA Memorandum #29 (Chicago: MESA Press)
[20.] Chien T-W, Linacre J M and Wang W-C 2011 *Communications in Computer and Information Science: Vol. 201* H Tan and M Zhou eds, pp. 578-585 (Berlin: Springer Verlag)
[21.] Chien T-W, Shao Y and Kuo S-C 2017 *BMC Med Res Methodol* **17**

[22.] Draney K and Wilson M 2011 *J Appl Meas* **12** 298-309
[23.] Andrich D 1978 *Psychometrika* **43** 561-573
[24.] Briggs D and Wilson M 2003 *J Appl Meas* **4** 87-100
[25.] De Boeck P and Wilson M 2004 *Explanatory item response models: A generalized linear and nonlinear approach* (New York: Springer-Verlag)
[26.] von Davier M 2006 *Multivariate and mixture distribution Rasch models* (New York: Springer)
[27.] Shore D A 2014 *Launching and leading change initiatives in health care organizations: Managing successful projects* (San Francisco, CA: Jossey-Bass)
[28.] Commons M L and Goodheart E A 2008 *World Futures* **64** 406-415
[29.] Fisher W P Jr 2019 *TMQ--Techniques, Methodologies, and Quality* **10** 21-40
[30.] Bateson G 1972 *Steps to an ecology of mind* (Chicago: University of Chicago Press)
[31.] Sijtsma K and Emons W H M 2013 *Theory & Psychol* **23** 786-796
[32.] Fisher W P Jr and Stenner A J 2018 *J Physics: Conf Series* **1044**(012025)
[33.] Sutton J, Harris C B, Keil P G and Barnier A J 2010 *Phenomenol and the Cog Sci* **9** 521-560
[34.] Fisher W P Jr and Morrison J 2020 *J Physics: Conf Series* in review