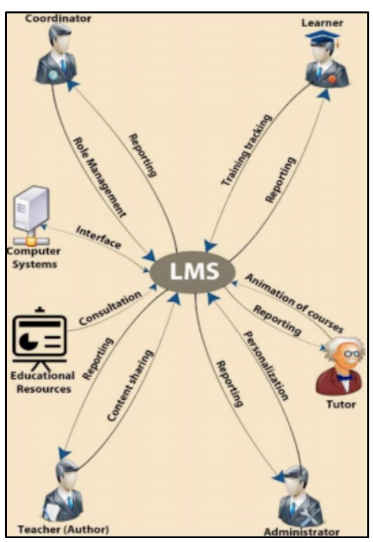
Module 8
Assessment and Tool proposal
James Brewer

Overview

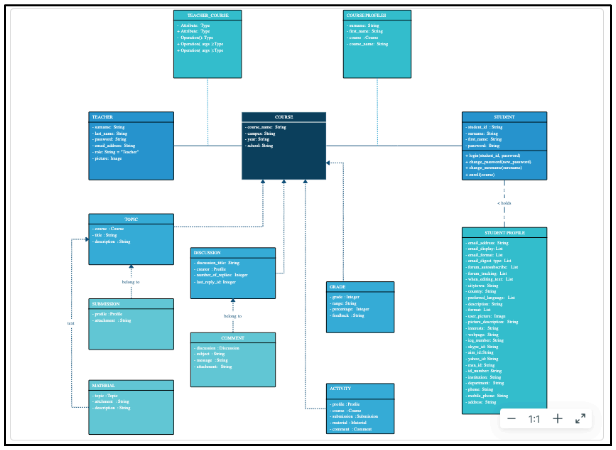
My approach to assessment will focus on the use of information and communication technologies (ICT) in the classroom to engage students in a collective solutioning exercise. This exercise is not limited to topics within the STEM domains, but could equally well be applied to other subjects, such as social science, language, literature, etc.

During this course, I have observed that many scholars concur that "current practices of schooling are somewhat outmoded in the new global working environment. Today, people work both individually and in groups to share complementary skills and accomplish shared goals—this practice contrasts to schools and assessments where students take tests individually" (Wilson & Scalise, 2016), given that even a "coherent set of science standards will not be sufficient to prepare citizens for the 21st century unless there is also coherence across all subject areas of the K-12 curriculum" (Pellegrino et al., 2014), and that "understanding of core ideas and crosscutting concepts be completely integrated with the practices of science—requires changes in the expectations for science assessment and in the nature of the assessments used. " (National Research Council, 2012, p. 15).

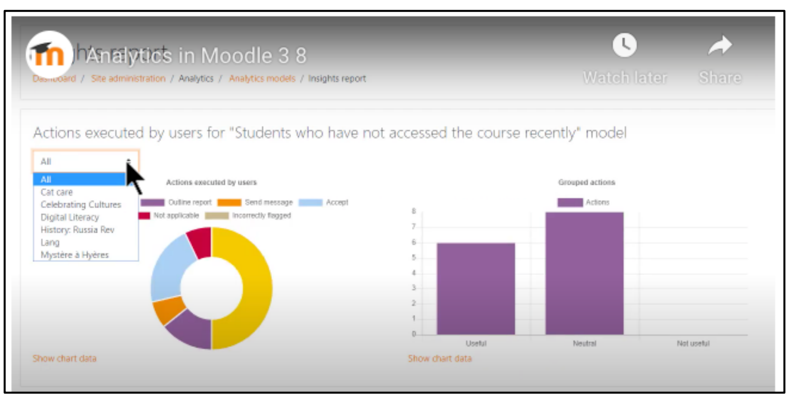
With this in mind, I propose a framework to organize collaborative and synthetic interactions between students, with teachers and external sources of expertise (scientists, engineers, and others) readily available through online connections. This framework is based upon the study from the ATC21S project group, and as summarized in Wilson and Scalise (2016). For its implementation, it is embodied in a Learning Management System (LMS). For the purposes of this assignment, I selected **Moodle**, an open-source LMS in wide use throughout the world. A conceptual diagram of Moodle, its functions, and the roles it supports, is provided below (Ouadoud, et al. 2016)



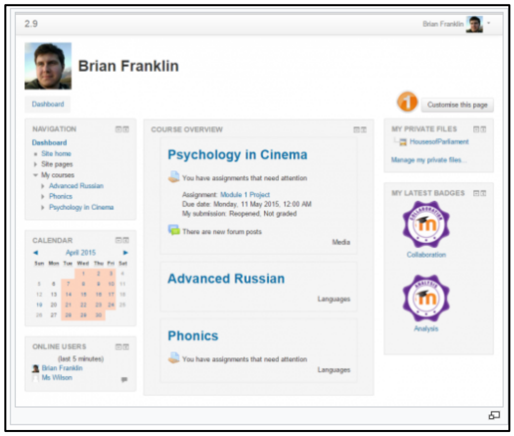
Furthermore, I propose a design for gathering information to measure the level of capability of the students throughout the LMS, and an analytics engine to help manage the raw data and render it into informative and relevant information sets to both assess and evaluate the strength of the curriculum and lessons, to assert the learning progression that occurs during the lesson, and to determine the impact of learning styles and other factors that bear on student performance. Moodle provides extensive support for data analysis and analytics. An illustration of a Moodle class diagram, which provides the foundation for the logical data model, is provided below (Jayarathne, 2023). Note that through Moodle's open architecture, this class structure is extensible to include new courses, curricula, assessment tools, etc.



Moodle supports a broad range of analytics and reporting and is fully capable of supporting ad hoc analysis to generate practically any report the administrator, teacher, or student may want (depending upon their privileges within the system). An example report is presented below (Moodle, 2023).



Finally, I propose that the LMS support an intuitive, flexible, and well-designed dashboard for students, teachers, and administrators to store course materials, submit assignments, review student competencies, and assess course design and overall value. An example dashboard for a student is presented below (Moodle, 2023).



What is an appropriate theoretical / philosophical framework to base your decisions about the characteristics of an effective STEM learning environment?

My framework draws from earlier research performed, and conclusions and recommendations posited, by the ATC21S project group, and as summarized in Wilson and Scalise (2016). This framework leverages open standards in its foundations, such as the Web3 standard, that includes technical capabilities such as "semantic constructions, crowdsourcing, peer evaluation, tagging, and the ability to judge the credibility and viability of sources" (Wilson and Scalise, 2016).

The framework itself is a based upon an extension to a construct-centered approach (Pellegrino, et al., 2012, p. 50). This approach is founded upon four principles and building blocks that will ultimately inform the assessment:

"Principle 1: Assessment should be based on a developmental perspective of student learning; the building block is a *construct map* of a progress variable that visualizes how students develop and how we think about their possible changes in response to items. Data analytics in this case are structured around the theoretical conception of the construct map.

Principle 2: There must be a match between what is taught and what is assessed; the building block is the *items design*, which describes the most important features of the format of the items—the central issue, though, is how the items design results in responses that can be analytically related back to the levels of the construct map.

Principle 3: Teachers must be the managers of the system, with the tools to use it efficiently and effectively; the building block is the outcome space, or the set of categories of student responses that teachers. These categories of student responses become the core of quantitative measures for conclusions through the data analytics.

Principle 4: There is evidence of quality in terms of reliability and validity studies and evidence of fairness, through the data analytics; the building block is an algorithm specified as a measurement model that provides for a visual representation of the students and the items on the same graph (called a "Wright Map"), and a number of other data analytic tools that are helpful for testing the quality of the measurement." (Wilson, 2005), (Wilson & Scalise, 2016).

Construct Maps

The learning progression framework cited above scales student capability development from rudimentary to the mastery of complex tools. In addition, the development of "soft skills" (i.e., judgement, leadership, innovation) is considered.

Following the template for construct maps put forth by the ATC21S project team, with the objective of assessing capability in "learning in networks" (Wilson & Scalise, 2016), "four strands of a learning progression" were defined:

1) Functioning as a consumer in network, involving the ability to gather, evaluate, and apply information from experts and virtual resources;
2) Functioning as a producer in networks, relating to the ability for "creating, organizing, and re-organizing" information and apply this to the group's information base;
3) Participating in the development of social capital through networks, indicating the ability to "moderating, leading and brokering the connections" between the team and external actors;
4) Participating in intellectual capital (i.e., collective intelligence) in networks, through cultivating and innovating the accrual of team intelligence and enhancement of personal perspective and metacognition (quotes from Wilson & Scalise, 2016).

Each of these "strands" will require the construction of construct maps (consumers, producers, social capital, and intellectual capital) that furnish an ordered set of capabilities that each student will be assessed against. These capabilities are to range from the most basic up to those expected from students who are skilled in ICT. A later version of this framework will include examples of such construct maps, aligned to a particular lesson, and calibrated to the system of implementation.

In the proposed LMS, the Construct Maps will assume a logical data structure that is expressed in database tables. The entities of the table will follow the structural example of a Construct Map is provided below (Duckor, et al., 2009)

	Respondents		*Responses to Items*
High	*Respondents* who can integrate normative and criterion referenced aspects of the construct map. They understand the construct map as a hypothesis about the empirical distribution of e.g. item difficulties and person proficiencies and try to align items design, outcome space, and measurement model with map.	Integrative 5	*Response to items* indicates understanding of where and when the particular construct map representation can be employed to strengthen/weaken inferential links between specific aspects of measurement system. Also demonstrates capacity to compare theoretical expectations against empirical findings.
	Respondents who can explain why some persons and items have more or less of the construct being measured. They may also be able to articulate the relationship between both.	Multi-structural 4	*Response to items* indicates understanding of how developing the orderliness of the Construct map aids in the development of items to populate scale, sketch out initial scoring strategy, provide validity check on content.
	Respondents who can describe the construct map in terms of a single concept or definition. They recognize the need for descriptions of ordered levels. They may also begin to develop sub- constructs to deal with complexity.	Definitional 3	*Response to items* indicates basic understanding of criteria for developing a Construct map. Shows that respondent can detect issues with construct definition, orderliness, dimensionality, etc.
	Respondents who can begin to describe all the goals, standards, factors, scales, etc. of interest but have not yet proposed to measure any single phenomena. They may be rigid and inflexible about the need to narrow and focus on a single construct map.	Discordant 2	*Response to items* indicates emerging notion of construct, but defined in multiple or vague ways. Shows that respondent may not be aware of inferential nature of measurement and the role of hypothesizing in advance.
Low	*Respondents* who ignore or are not attentive to any notion of cognitive or construct-based theory.	Pre-measurement 1	*Response to items* indicates a lack of concept or understanding of notion of construct or is off-topic.

Items Design

This building block requires scenarios to measure students' capabilities across each of the four construct maps. Given that this lesson will involve the development of ICT skills, each scenario should capture "process data from activities in the tasks" (i.e., mouse clicks, mouse positioning, events launched, etc.), "collaboration data" (i.e., instant messages, text messages, email, data retrieval logs, etc.), qualitative textual interactions between participants (students, teachers, external actors, etc.), as well as point-in-time responses to knowledge checks (i.e., quizzes, short essays, etc.).

The scenarios I have in mind will be provided in the next version of this document, with the understanding that they will involve interactive experiences founded upon industry standard collaboration tools (i.e., Google Drive, Canvas, etc.)

An essential characteristic of the tooling that underlies each scenario is the matter of diversity and inclusion. Distinctive student learning styles should be accommodated (i.e., visual, verbal, haptic) and rich contextual help should be available (i.e., dictionary, thesaurus, grammar) as well as the ability to connect information semantically through a knowledge engine.

Outcome Space

The outcome space will combine the structure of each of the four strands, with the criteria established within the items design. The difficulties of construction and population of the outcome space can vary widely, depending upon the nature of the items design. For example, if the responses are based upon a "fixed set (as in a multiple-choice item), the set can be planned ahead of time, but for open-ended tasks and activities, the resulting work product is more complex" (Wilson & Scalise, 2016a). Clearly, when evaluating student progression in ICT, there will be qualitative data that requires teacher review and analysis.

In the proposed LMS, the Outcome Spaces will assume a logical data structure that is expressed in database tables. The entities of the table will follow the structural example of an Outcome Space is provided below (Duckor, et al., 2009)

5	***Integrating the Normative and Criterion reference aspects of the Construct map (Persons and Items)*** •Hypothesizes person and item distributions on either side of Construct map •Notes that individual items and persons have relationship e.g. ability and difficulty •Describes expectation of the relationship between persons in terms of relative “ability” based on ranking •Describes expectation of the relationship between items in terms of relative “difficulty” based on criteria or characteristics at level •Recognizes implications for items design e.g. coverage and sampling strategy	
4	***Orderliness of the Construct map (Persons)*** •Suggests “respondents” as having more or less of the construct in some direction •Describes expectation of respondent “types” on right side of map e.g. “from very motivated to not motivated individuals”, “from experts to novices”	***Orderliness of the Construct map (Items)*** •Suggests “items” as having more or less of the construct in some direction •Describes expectation of responses as generalized groupings of item responses e.g. “levels”, “categories”, “buckets”
3	***Singular concept*** •States precise definition about the construct which suggests a continuum Identifies extremes on the Construct map e.g. high-low, a lot-a little, increasing-decreasing •Recognizes that construct can be split into sub-constructs e.g. “I broke it down further” or “looking at dimensions of construct” •Recognizes that construct described as orderable set of observations •Recognizes construct as latent, unobserved phenomena	
2	***Multiple/Vague concept*** •Describes goals, outcomes, standards, factors, variables, rubrics, scales, etc. •Presents many concepts without specification of a single dimension e.g. “I got bogged down with many definitions” or “ I used Wiggins criteria” or “Previous research suggests there are three pathways…” •Includes more than one dimension in description •May maintain misconception related to attempt to “measure” manifest phenomena, changes in pre-post states, etc.	
1	***Lack of concept*** •Offers no concept of Construct (“a test is a test is a test” or “a test measures what it is designed to measure”) •Presents empirical results of data analysis without reference to Construct •Presents outcome space or item type with no reference to Construct	
0	**No response (irrelevant or off-topic)**	

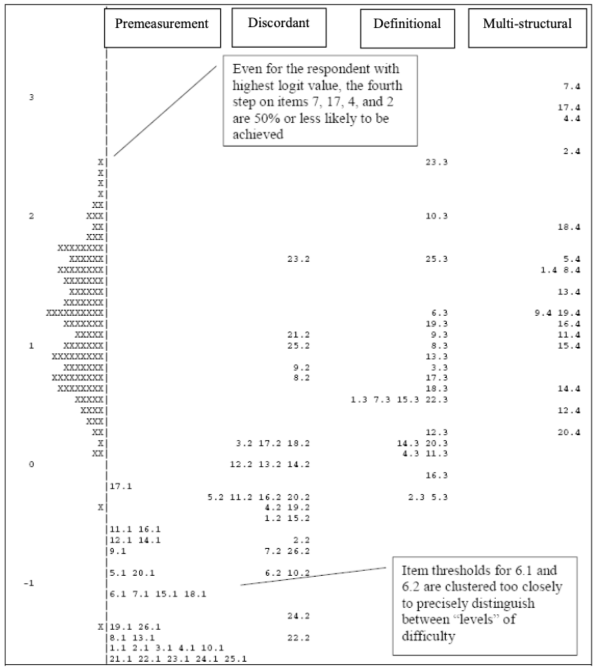
The Measurement Model

A technique that is popular among leading researchers involves the development of a measurement model. This model provides the mathematical and statistical foundation for the aggregation of results from the outcome space, as well as for predictive and inferential modeling regarding student performance and learning progression.

However, to become accessible and relevant, the measurement model must yield assessments that “clearly align with learning goals through these constructs, to produce valid and reliable evidence of what students know and can do for the development perspective, and to generate evidence useful to teachers and students. One potential mechanism to achieve these goals is to model assessment practice through a set of exemplary classroom materials”. (Wilson and Scalise, 2016b). This observation recalls the need to develop an Item Design that ensures fidelity between the constructs, the outcomes, and that fits into the analytics requirement for both assessment and evaluation of students, teachers, and the curriculum itself.

For my measurement model, I intend to explore the intersection between measurement and learning analytics, defined as the “collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Wilson & Scalise, 2016b).

For an example of a measurement technique, a Wright map to examine the empirical ordering of persons and items in order to compare those to our theoretical expectations based on the Construct Model. The figure below depicts the distribution of respondent and item locations for the Construct Model scale, following the above examples cited for the Construct Model and the Outcome Space (Duckor, et al., 2009).



What are the appropriate practical implications of the theoretical / philosophical framework that you have developed?

I think that the construct centered approach to assessment is the most conceptually sound of any that I have seen thus far, and offers several actionable advantages:

- It furnishes a logical and nearly seamless progression from curriculum planning, through lesson development, following the "design, construction, and mapping" aspects of the "three-arena model" (Remillard, 2009). This leverages the involvement the teacher had in the curriculum/lesson development and immerses them thoroughly into the assessment approach.

- The students learning ICT will understand both the lesson objectives and the rubric upon which they are scored directly and intuitively. Moreover, this approach provides the conceptual basis for technologies capable of providing feedback in near-real time, thus promoting enhanced learning and retention practices.

- A diverse population of students can participate on equitable basis, given the flexibility of the construct-item mapping, as well as in consideration of the ability to measure multidimensional data using a platform similar to the BEAR Assessment System (Wilson, 2005). This platform was created using design principles based upon the four principles and building blocks described above. This coherence gives teachers and researchers the flexibility to devise analytics that can account for both individual differences as well as learning styles.

- The selection of an open-source LMS was a conscious decision, aligning with Moodle's architects to adopt the four open-source freedoms:
    - Open source is the right choice for education, since educators share traits with open-source practitioners such as collaboration, creativity, and a love for lifelong learning.
    - Open source is long-lasting and given that a large community has grown around the project, this ensures its survival even if the company were to dissolve.
    - Open source is reliable and secure and enables education institutions and organizations to have complete control over their data, including how and where they run their Moodle sites and store their data. They can choose whether they will use their own resources for hosting and support.
    - Open source is flexible and customizable facilitating the extension or modification of solutions that conform to unique requirements, from the user interface to extensions to external software. (Moodle, 2023)

What are the appropriate assessment practices that result from the theoretical / philosophical framework you are employing?

As noted above, I intend to pattern my assessment approach after the BEAR Assessment System (Wilson, 2005), due to its conceptual compatibility with the construct centered approach.

What are the implications for the physical layout of classroom or school spaces in an effective STEM learning environment?

For this proposal, students will interact largely through a computer or through another similar user interface. This means that the classroom configuration can be completely flexible with regard to seating and table arrangement, teacher placement, and so on. In fact, students need not participate in the same room, building, or environment, given the virtual nature of the exercise.

Given the nature of virtual learning, it may be advantageous to meet in a common space, where teachers and mentors can congregate with students to address questions, help with technical issues, and to have facilitated breakout sessions to amplify on lesson content and implications.

What are the student actions you should observe?

Chiou, et al. (2008) studied the implications of utilizing a construct-centered approach that is similar to the principles stated earlier in this paper. They found that this approach supports several desirable outcomes for STEM education. First, since STEM "stresses the continuity of knowledge…construct mapping can help in making better inter-connections between courses".

Second, this approach engenders "creative and independent learning abilities…through construct mapping, which focuses on freely associated connections…students' logical thoughts and deductive and self-learning abilities can be enhanced". Third, the "meta-learning strategy of construct mapping and the experimental design in this study can be easily extrapolated to other…curriculum areas", a frequently cited recommendation among STEM researchers (Chiou, et al., 2008).

The specific behaviors and actions regarding performance will be cataloged into the Item Designs and captured with the Outcome Model. I think that content that each student provides, the degree of interaction each student generates with their peers, the amount of information they obtain should be of primary importance.

In a virtual setting, a student's ability to perform both cognitively and emotionally depends on the e-learning approach that is employed within a virtual curriculum. The table below illustrates three approaches; for the purposes of this exercise, I think the hybrid approach is desirable due to its emphasis on cooperation, shared resources, the ability to leverage different modalities, and the sharing of observations, hypotheses, and outcomes (Kallel, et al., 2019).

Teacher-centred Approach	**Learner-centred Approach**	**Hybrid Approach**
Individual and isolated work	Collaborative work	Cooperative work
Information transfer	information exchange	Information sharing
Single media	Multiple media	Multiple media
Passive learning	Exploration learning	Learning by sharing

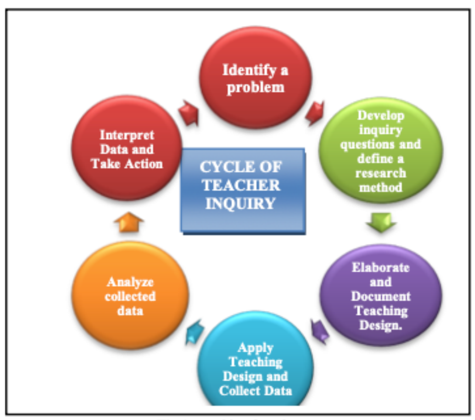
Given this choice, there are several methods for observing and assessing student actions against desirable behaviors. Since this exercise is designed to support fully virtual interaction (i.e., little to no physical interactions), the students will need to demonstrate their abilities not only for cognitive abilities (as demonstrated through evaluation of their Outcome Space ratings), but for their emotional awareness and intelligence as well. Kallel et al. (2019) developed a Moodle-based application that can identify degrees of emotional content, both positive and negative, from interactive messaging. This involves using both lexical and keyword statistical analysis to isolate, evaluate, and rate the emotional maturity of each student as they progress through each lesson. (Kallel, et al., 2009, pp. 4-7).

What are the teacher actions you should observe?

The teacher's role in this program should be one of facilitation and mentorship. That is, teachers should be the exemplar of the capabilities defined in the Item Designs at the highest levels. They should promote competence, virtual etiquette, teamwork, and collaboration. They should not expect to stand and deliver content to students except in broad terms, or if the exercise appears to stall.

Beyond these basic aspects of teacher performance, it is worthwhile to follow Schon (1983) and his concept of reflective practice. The key concept of reflective practice is that "purposeful and systematic reflection on experience is a key for continuous learning…an important instrument for practice-based professional development as well as organizational learning and improvement" (Haroun, et al. 2018, p. 44). For example, it could be observed whether the teacher is following "reflection-in-action", when the teacher reacts to a situation during the lesson (i.e., students not engaging in the material, thus intervening to refocus the students) or "reflection-on-action", which involves a retrospective analysis of the lesson and an evaluation of whether it was his/her performance, the lesson plan itself, the delivery mechanism, or some other factor (i.e., determining the root causes that inhibited student engagement and performance) (Schon, 1983).

Teachers could be evaluated on their ability to successfully apply the principles of reflective practice in both scenarios. Furthermore, in the case of "reflection-on-action", the capability of the teacher to navigate the "cycle of teacher inquiry", as depicted in the graphic below, could be part of the evaluation criteria ((Haroun, et al. 2018, p. 45).



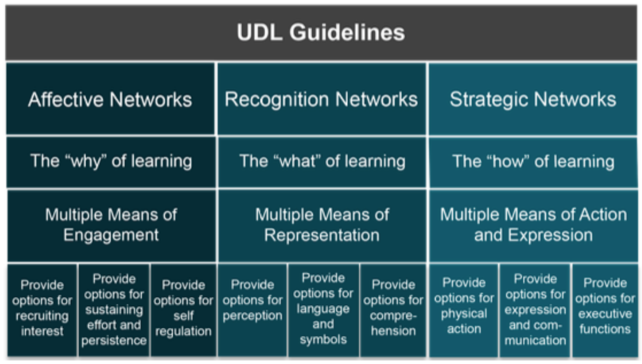
How can the learning environment be more inclusive and inviting for all students?

A crucial driver for ensuring an inclusive platform, particularly for virtual learning, is to adopt the principles of Universal Design for Learning (UDL). UDL is a "curriculum design, development, and delivery framework used to create equitable, inclusive, and accessible learning environments. UDL assumes all learning environments are diverse and that all learners have variable learning needs. UDL works to provide learning spaces (both physical and virtual) where all students can effectively learn and demonstrate their learning while creating expert learners who are purposeful, motivated, resourceful, knowledgeable, strategic, and goal-directed" (Kearney, 2014).

The Center for Applied Special Technology (CAST), which developed the UDL framework, established three principles to inform design, development, and delivery in practice to address different brain pathways used in learning:

- Affective networks – The “why” of learning
- Recognition networks – The “what” of learning
- Strategic networks – The “how” of learning

The graphic below illustrates how these pathways are actuated and supported by the CAST principles of employing multiple means of engagement, representation, and expression.



As noted above, students display a wide variety of learning styles, language skills, social intelligence, and sophistication in ICT generally. A key factor in creating inclusive environments is actually promoted by the virtual nature of this proposed exercise: the physical distance that the user interface inherently provides. Among other advantages, it creates a “safe zone” where students can utilize various accommodations without revealing their use to others.

The assignment of mentors to guide and facilitate lessons is also known as a factor for successful virtual sessions. A teacher can certainly fulfill this role, but it can be more effective for a student with advanced capabilities does so.

How should disciplines outside of STEM be included or integrated?

As noted above, the construct centered method can accommodate curricula outside of the ICT that is proposed here. Using this method in concert with the BAS has been documented as being performed successfully in the areas of poetry analysis, second language learning, and others (Wilson & Scalise, 2016a).

There is considerable debate within the STEM community as to what disciplines ought to be included. Li et al. cite an NSF publication that documents its “approved fields…under the umbrella of STEM” (NSF, 2014). This includes the traditional “core” disciplines, such as physics, chemistry, and materials research, but also includes disciplines in psychology and social sciences (Li et al., p. 2). This taxonomy is at variance with other entities, even some within the federal government (i.e., Department of Homeland Security and Immigration and Customs Enforcement) that exclude social sciences from the STEM disciplines.

I think that such exclusions reflect greater attention to domains and not nearly enough to the framework that STEM education should follow. As should be obvious, the framework presented above for pedagogy and assessment can readily accommodate the instruction in subjects outside the traditional “science, technology, engineering, and mathematics” domains. As an example, Chiou (2008) developed a construct map, and a rudimentary items design and outcomes space, for economics. This convinces me that economics can and should be part of a STEM curriculum.

I think that, based upon the evidence cited above, STEM educators could make a strong case for inclusion of social sciences in the STEM umbrella, not to mention performance disciplines such as music, painting and drawing, sculpture, drama, dance, and more.

From an implementation perspective, an LMS such as Moodle makes the inclusion of additional construct maps for additional disciplines a relatively simple matter of adapting the existing entity-relationship data model and appending new content to the tables.

What is the appropriate level of analysis for determining whether or not an effective STEM learning environment exists?

That question can be answered only in terms of outcomes. The trend toward greater reliance on virtual learning, it seems, is driven in part by economics. If more students can be reached, and less money spent on classroom environments, through virtual means, that is an attractive position for administrators and state regulators to take. Without strong assessment and evaluation mechanisms, as outline above, the efficacy of such an approach, and the factors that support or weaken it, are left to the better (or lesser) judgement of non-educators.

Clearly, the implementation of a well-configured and robust LMS can provide the views into student and teacher performance against established criteria for success. A recent study noted that “Monitoring students in Learning Management Systems (LMS) throughout the teaching–learning process has been shown to be a very effective technique for detecting students at risk. Likewise, the teaching style in the LMS conditions, the type of student behaviours on the platform and the learning outcomes” (Sáiz-Manzanares et al., 2021).

In another study, the authors evaluate the potential to “ecologize instruction, which means situating learning in its authentic contexts of use” in evaluating the efficacy of STEM programming, which they refer to as Caliper (Caliper, 2021, p. 1). The methodology of Caliper involves five phases including all stakeholders (e.g., students, teachers, administrators, external actors, etc.). The phases involve having been “(a) Informed about available opportunities for improved measurement information quality, (b) Consulted on the measurement content decisive to ecosystem success, and (c) Involved in the production of data testing hypotheses as to the measurability of ecosystem success, the resulting calibration of Caliper will support (d) new kinds of Collaboration leading to (e), stakeholder Empowerment. The delineation of a continuum of ecosystem success will enable the administration of individually customized assessments and management options tailored to local circumstances” (Caliper, 2021, p 3).

The Caliper team believes that this method can engender “meaningful and reproducible decision processes across the STEM learning ecosystems. Broadly speaking, the effect is to advance cultural progress by improving the quality of the cognitive supports scaffolded by the knowledge infrastructure" (Caliper, 2021, p. 3). This translates into a "step in a long-overdue new direction toward creating sustainable change by coordinating and aligning the behaviors and decisions of every key stakeholder group" (Caliper, 2021, p. 3).

References

Chiou, C. C., (2008). The effect of concept mapping on students' learning achievements and interests. Innovations in Education and Teaching International · November 2008 DOI: 10.1080/14703290802377240

Haroun, M. H. A., HACHEM, H., Ziti, C. (2018). Teaching and Learning Analytics to Support Teacher Inquiry. International Journal of Engineering & Technology, 7 (4.32) 44-47

Jayarathne, R. Moodle Conceptual Class Diagram. Downloaded on 1 August 2023 from https://creately.com/diagram/example/hmssopa41/moodle-conceptual-class-diagram

Kearney, D. B. (2014). Universal Design for Learning (UDL) for inclusion, diversity, equity, and accessibility (idea). eCampusOntario, Toronto

Li, Y.; Wang, K.; Xiao, Y.; & Froyd, J. (2020). Research and trends in STEM education: a systematic review of journal publications. International Journal of STEM Education. 7:11. https://doi.org/10.1186/s40594-020-00207-6

Moodle 2023b. https://moodle.com/news/moodle-lms-open-source/#:~:text=Open%20Source%20allows%20organisations%20to,sites%20and%20store%20their%20data.

Morrison, J., Fisher, W. P. Jr. (2021). Caliper: Measuring success in STEM learning ecosystems. https://doi.org/10.1016/j.measen.2021.100327

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Committee on Science Learning, Kindergarten Through Eighth Grade. R.A. Duschl, H.A. Schweingruber, and A.W. Shouse (Eds.). Board on Science Education, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Committee on Conceptual Framework for the New K-12 Science Education Standards. Board on Science Education. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards*. Washington, DC: The National Academies Press. https://doi.org/10.17226/18409.

National Science Foundation (2014). NSF Approved STEM Fields. Found at https://btaa.org/docs/default-source/diversity/nsf-approved-fields-of-studycac2.pdf?sfvrsn=642824bc_2 (28 July 2023).

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. Review of Educational Research, 75(2), 211–246.

Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Rodriguez-Diez. J. J., Rodríguez-Arribas, S., Díez-Pastor, J. F., Ji, Y. J. (2021). Improve teaching with modalities and collaborative groups in an LMS: an analysis of monitoring using visualisation techniques. Journal of Computing in Higher Education (2021) 33:747–778 https://doi.org/10.1007/s12528-021-09289-9

Schon, D. 1983. Reflective Practitioner: How Professionals Think in Action. New York: Basic Books.

Wilson, M. (2005). *Constructing measures: An item response modeling approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal for Research in Science Teaching, 46*(6), 716-730.

Wilson, (2009a, December). *Assessment for Learning* and *for Accountability*. Paper presented at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda, at ETS, Princeton, NJ.

Wilson, M., & Scalise, K. (2016a). Assessment of Learning in Digital Networks. In Griffin, P., & Care, E. (Eds.), *Assessment and Teaching of 21st Century Skills Volume 2 - Methods & Approaches*. Dordrecht: Springer.