



# Developing an Integrated STEM Classroom Observation Protocol Using the Productive Disciplinary Engagement Framework

Yann Shiou Ong<sup>1</sup> · Jaime Koh<sup>1</sup> · Aik-Ling Tan<sup>1</sup> · Yong Sim Ng<sup>1</sup>

Accepted: 11 March 2023 / Published online: 3 April 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

STEM education and research has gained popularity internationally over the last decade. However, there is a lack in specifications in existing K-12 STEM classroom observation protocols of how features of an integrated STEM experience/lesson would lead to desired outcomes and how those outcomes should be measured. To bridge this gap, we propose the development of a new integrated STEM classroom observation protocol (iSTEM protocol). This article describes the ongoing development work of the iSTEM protocol, which features two creative attempts. Firstly, the productive disciplinary engagement framework is adapted to design a classroom observation protocol that provides a coherent frame of design principles to be met to achieve desired 3-dimensional pedagogical outcomes. Secondly, *interdisciplinarity* of student engagement was interpreted in terms of the extent to which students take a systematic and disciplinary-based approach to make and justify decisions during STEM problem-solving. The iSTEM protocol comprises 15 items (4-point scale) rated holistically for the extents to which evidence was found in the observed lesson for (1) the 3-dimensional pedagogical outcomes of productive interdisciplinary engagement (five items) and (2) problematising, resources, authority, and accountability design principles (10 items). The accompanying iSTEM profile visually represents and communicates the strengths and inadequacies in design principles, thus providing explanations for extents of students' productive interdisciplinary engagement. The iSTEM protocol will contribute as a research tool for STEM education researchers and as a pedagogical guide for STEM classroom teachers to improve their design of STEM learning experiences.

**Keywords** Interdisciplinary · Observation protocol · Productive disciplinary engagement · STEM education

---

✉ Yann Shiou Ong  
yannshiou.ong@nie.edu.sg

Extended author information available on the last page of the article

## Introduction

STEM education and research has gained popularity internationally over the last decade (Li et al., 2020). While initially US-based, STEM education has gained worldwide interest, including among Asian economies (Ong, 2021) and Australasia (Murphy et al., 2019; Papua New Guinea [PNG] Education News, 2022). A review of US-based STEM K-12 education programmes noted inadequate specifications of how features of an integrated STEM experience/lesson would lead to desired outcomes and how those outcomes should be measured (National Research Council [NRC], 2014), a problem also encountered in efforts to promote STEM education in Singapore. This problem is partly contributed by the lack of consensus on what counts as STEM education, what STEM integration means (English, 2016; Kelley & Knowles, 2016; NRC, 2014), and what STEM education outcomes are worthy of pursuit.

Observational tools such as a STEM classroom observation protocol present a viable way to provide guidance on what makes a “good” integrated STEM experience for STEM education stakeholders, including researchers, teachers, teacher educators, curriculum developers, and policymakers (Dare et al., 2021). However, our literature search for existing K-12 integrated STEM (rather than monodisciplinary STEM) classroom observation protocols (Dare et al., 2021; Peterman et al., 2017; Wheeler et al., 2019) revealed that none articulated a set of design principles/features with corresponding pedagogical outcomes. To bridge this gap, we propose a new integrated STEM classroom observation protocol (iSTEM protocol) informed by the productive disciplinary engagement (PDE) framework (Engle, 2012; Engle & Conant, 2002). Two novel features of the iSTEM protocol are highlighted. Firstly, items were constructed based on the PDE framework comprising four design principles—problematising, resources, authority, and accountability—theorised to foster the three-dimensional student outcomes of engagement, interdisciplinarity (modified from disciplinarity), and productivity. Secondly, interdisciplinarity of students’ engagement in STEM problem-solving is interpreted as the extent to which students take a systematic and disciplinary-based approach towards decision-making. In this paper, we report on the design considerations and challenges in ensuring the validity, reliability, and usability of the iSTEM protocol.

## Literature Review

### Defining STEM Integration

Drawing upon definitions of multi-, inter-, and transdisciplinary proposed by STEM educators (Vasquez et al., 2013) and STEM professionals (Choi & Pak, 2006), the iSTEM protocol focuses on STEM integration at the *interdisciplinary* level. That is, students apply concepts, skills, and practices from two or more STEM disciplines to solve a real-world problem by interacting between disciplines and blurring disciplinary boundaries such that the problem cannot be easily categorised as a science, technology, engineering, or mathematics problem nor can the solution be sought through a single, disciplinary approach. While Vasquez et al. (2013) considered students working on real-world problems/projects by applying knowledge and skills from two or more disciplines as sufficient conditions for transdisciplinary integration, we interpret transdisciplinarity as involving “a common

perspective that ‘transcends’ those that are standard in the two disciplines” (Choi & Pak, 2006, p.355). Our experiences as K-12 educators and researchers suggest it is more likely for schools to achieve interdisciplinary integration, while transdisciplinary integration is more likely at the post-secondary levels as students have greater disciplinary expertise. Hence, our proposed protocol—intended for use in elementary and secondary STEM classrooms—focuses on interdisciplinary integration.

## Existing Protocols

Existing STEM classroom observation protocols mostly target post-secondary/college levels and monodisciplines in STEM (for examples, see Anwar & Menekse, 2021). Nevertheless, we identified three integrated STEM observation protocols intended for K-12 classrooms. These protocols—Classroom Observation Protocol for Engineering Design (COPED) by Wheeler et al. (2019), Engineering-Infused Lesson (EIL) Rubric by Peterman et al. (2017), and K-12 STEM Observation Protocol (STEM-OP) by Dare et al. (2021)—are compared as follows and their key features are summarised in Table 1.

## Frameworks and STEM Integration

All three protocols involved integrated STEM to different extents. COPED and EIL Rubric focused on science and engineering integration; STEM-OP considered integration of all four STEM disciplines generally (see Table 1 for integration-related items). Emphasising engineering design integration in secondary science classrooms, COPED’s framework was based on (1) engineering design process components (EDPC: problem, brainstorming, researching, planning, building, testing, evaluating, redesigning, and sharing) and the extents to which processes are teacher-driven versus student-driven, as well as (2) engineering habits of mind (EHOM: creativity, divergent thinking, systems thinking, optimism, collaboration, communication, and attention to ethical considerations). EDPC could be considered an integrated STEM lesson’s design features while the EHOM would be the pedagogical outcomes.

EIL Rubric was intended to “help teachers infuse engineering content and activities into their [science] lessons” (Peterman et al., 2017, p.1916). EIL Rubric comprises three sections: curriculum materials (i.e., lesson design features reflected in materials, such as alignment with NGSS standards, the nature of the design challenge presented to students, presence of science content, science-engineering connection, and assessment), design-centred pedagogical practices (for implementing design challenges), and engagement with engineering concepts (use of engineering terminologies and making connections to real-world engineering applications). Aligned to its intention, the EIL Rubric framework emphasises design features of an engineering design challenge lesson but does not consider pedagogical outcomes.

STEM-OP was designed based on the characteristics of integrated STEM education as informed by existing literature. These include (according to item names) relating content to students’ lives, contextualising student learning, developing multiple solutions, cognitive engagement in STEM, integrating STEM content, student agency, student collaboration, evidence-based reasoning, technology practices in STEM, and STEM career awareness. The ensemble of ten characteristics seemed eclectic rather than based on a coherent framework. For example, there is a mixture of STEM lesson design features (e.g., integrating STEM content; technology practices in STEM), possible pedagogical outcomes (e.g., cognitive

**Table 1** Key features of three K-12 integrated STEM observation protocols

Protocol	EIL Rubric (Peterman et al., 2017)	COPEd (Wheeler et al., 2019)	STEM-OP (Dare et al., 2021)
Framework	<p>Items in three sections:</p> <ul style="list-style-type: none"> <li>• Curriculum materials (lesson design features)</li> <li>• Design-centred pedagogical practices</li> <li>• Engaging with engineering concepts</li> </ul>	Engineering design process components (EDPC) and extents to which the processes are teacher-driven versus student-driven, as well as engineering habits of mind (EHOM)	Characteristics of integrated STEM education represented by 10 items. Each item represents a characteristic
STEM integration	Infusion of engineering content and activities into science lessons	Science and engineering integration (based on teacher's intention)	Items: 5. Integrating STEM Content (students draw upon content from more than one STEM discipline); 8. Evidence-based Reasoning (for claims/design choices); and 9. Technology Practices in STEM (students' use of digital technology to construct a scientific model or design solution)
Type of items	19 items (checklist)	Codes for presence of EDCM (9), EHOM (7), and grouping based on any student's observed action, as well as write lesson description for every 2-min interval	10 items; 4-point scale All but one item (Item 9) focuses on teacher action
Nature of analysis	Holistic: analyse lesson documents Summary scores per section (based on checked items) + total score	Determine extents of science and engineering integration and extents to which EDPC are teacher-driven versus student-driven, post-observation	Holistic No mention of adding up item scores for overall score
Reliability	Krippendorff's alpha = 0.74 ( $n = 16$ , 3 raters)	Segmented (2-min interval) for EDPC and EHOM Holistic for extent to which design component is teacher-driven vs student-driven	Krippendorff's alpha = 0.580 to 0.870 ( $n = 104$ , 7 raters)
Validity	No mention	Cohen's kappa = 0.81 (final) ( $n = 760$ coded entries for two 40-min videos, 2 raters) Content validity: two cycles of expert panel (engineers, engineering educators, science educators, and instrument development experts) review and revision	Credibility and trustworthiness: common characteristics of integrated STEM were based on literature Face and content validity: items reviewed by 3 STEM education experts

engagement in STEM; evidence-based reasoning), and emphasis on students' STEM identify development through "STEM career awareness".

### **Types of Items and Nature of Analysis**

The three protocols vary in their types of items and how lessons are analysed. For COPED, the observer codes for the presence of EDPC, EHOM, and grouping (whole class, small group or individual) according to what is observed of any student's action in 2-min intervals. For example, if one student in a group demonstrated creativity during brainstorming within a 2-min interval, codes for creativity and brainstorming, and small group will be circled for that interval. Additionally, the observer is requested to write a lesson description for each interval. Post-observation, the observer indicates (based on observation codes) the extent to which each observed EDPC was teacher- versus student-driven on a 4-point ordinal scale. Thus, COPED comprises segmented analysis of EDPC and EHOM and holistic analysis of extents of teacher-versus student-driven EDPC. Conversely, EIL Rubric is used for analysing lesson documents, such as lesson plans, worksheets, and assessments. The observer determines whether a set of lesson documents meets the requirements on a 19-item checklist grouped into three sections as previously described. A section score is generated based on checked items. Section scores are then added to produce a total, holistic score for the set of lesson documents. Finally, STEM-OP comprises 10 items each on a 4-point Likert scale with specific statements for each rating. All but one item focuses on teacher action (exception is Item 9: Technology practices in STEM, which emphasises student action). Thus, the observer scores an observed lesson holistically using STEM-OP.

### **Reliability and Validity**

All three protocols measured reliability by considering inter-rater reliability and utilised either Cohen's kappa (COPED) or Krippendorff's alpha (EIL rubric and STEM-OP). For validity, COPED and STEM-OP ensured content validity by seeking STEM-related experts' opinion while EIL Rubric did not discuss validity considerations. Furthermore, STEM-OP authors argued for the protocol's credibility and trustworthiness as its items reflected common characteristics of integrated STEM from the literature.

In summary, among the identified K-12 integrated STEM classroom observation protocols, both EIL Rubric and STEM-OP did not include well-defined design features and pedagogical outcomes. While COPED's framework included both, its emphasis on engineering makes it unsuitable for a broader use with integrated STEM lessons. While COPED (and STEM-OP) established reliability and validity to reasonable extents, in terms of usability, COPED's segmented analysis and writing of lesson descriptions at 2-min intervals seems daunting even for trained researchers. As none of the reviewed protocols met the requirements of (1) including a set of design features and corresponding pedagogical outcomes and (2) can be easily used to analyse integrated STEM lessons, we were motivated to develop an integrated STEM observation protocol that addressed these gaps.

## Methods

We followed the approaches of existing STEM observation protocols (i.e., Dare et al., 2021; Milford & Tippett, 2015; Wainwright et al., 2003; Wheeler et al., 2019) in developing the integrated STEM classroom observation (iSTEM) protocol. Specifically, protocol development involves three phases: (1) preliminary protocol development; (2) iterative pilot testing, protocol review, and revision; and (3) reliability establishment. We unfortunately made limited progress in phase 3 due to COVID-19-related situations. Nevertheless, we provide an overview of the iSTEM protocol and a detailed description of phases 1 and 2.

### Preliminary iSTEM Protocol Development

The iSTEM protocol is intended for observation of individual integrated STEM lessons in primary/elementary and secondary classrooms and should meet three criteria: validity, reliability, and ease of use as a research instrument for analysing live integrated STEM lessons. Additionally, protocol items should be framed by a coherent set of design principles and pedagogical outcomes, which we achieved by applying the productive disciplinary engagement (PDE) framework (Engle, 2012; Engle & Conant, 2002). The PDE framework states four design or guiding principles should be fulfilled to foster students' engagement in specific disciplinary practices (e.g., scientific argumentation) in productive ways that demonstrate intellectual progress (i.e., the three-dimensional pedagogical outcomes). Students should be engaged in problems that are meaningful to the disciplinary community (problematising) and be provided resources to help them solve the problems. Students should also be given the authority to propose/construct their own ideas/solutions yet have their ideas/solutions held accountable to themselves (clarity of expressions), peers (especially differing ideas), the teacher, and disciplinary norms.

### Productive Interdisciplinary Engagement

We illustrate how we adapted the PDE framework's three-dimensional outcomes and design principles for the iSTEM protocol by using hypothetical examples of the case of students solving the STEM problem of making a pill coating for an oral medication that meets the criteria of durability (dissolving in the stomach after 2 min), cost, and ease of swallowing the pill (henceforth the pill-coating activity). Detailed analysis of the lesson using the iSTEM protocol is discussed subsequently.

For the three-dimensional outcomes, *engagement* refers to the extent to which students are cognitively engaged during a STEM lesson (one item in iSTEM protocol), working as a group (one item) towards solving a STEM problem. Drawing on the ICAP framework (Chi & Wylie, 2014) and Walton's dialogue types (Walton, 1998), students' cognitive engagement is highest when they engage in critical discussion (students build on and challenge ideas with justifications, or discuss/juxtapose multiple ideas), followed by idea-building (students build on ideas without challenging ideas), information-seeking (mainly one student shares ideas while others ask clarifying questions), and the least when they engage in exposition (mainly the teacher or one student shares/elaborates idea). Moreover, students should be working in their groups to demonstrate such cognitive engagement as group work motivates the need for making individuals' thinking visible to peers. Using the case of the pill-coating activity, a group where students mostly work as a group and

multiple students put forward and challenge ideas demonstrates high engagement in group work and critical discussion. Conversely, a group where students work individually on different parts of the problem with one student mainly telling others what to do demonstrates low engagement.

Defining interdisciplinary practice(s) proved challenging as there is no single, consensus definition. Definitions using the twenty-first-century skills such as critical thinking, communication, collaboration, and creativity (Partnership for 21st Century Skills, 2015) lose the disciplinary aspect of integrated STEM as such skills could apply to all disciplines, even beyond STEM. Conversely, opting for discipline-specific practices loses the integrative aspect. Hence, we interpreted *interdisciplinarity* of engagement as the extent to which students take a systematic, disciplinary-based approach to make and justify decisions relevant to solving the STEM problem (one item). High interdisciplinarity means students' decisions are based on weighing benefits and trade-offs, evidence (researched/given information or gathered data), and disciplinary reasonings (e.g., scientific/mathematical reasonings) tied to success criteria/solution requirements. Referencing the pill-coating activity, groups that evaluate all three solution criteria to reason towards an optimised solution based on evidence from scientific inquiry (e.g., testing each ingredient systematically in incremental amounts to determine minimum amounts to meet durability criterion) and mathematical reasoning (e.g., estimate cost of pill coating based on amount and unit cost of each ingredient used) exemplify high interdisciplinarity in decision-making. Groups that randomly select the amounts of each ingredient or do not justify their decision demonstrate low interdisciplinarity. We contend that our interpretation of interdisciplinarity considers both the disciplinary aspect (i.e., disciplinary reasonings) and integrative aspect (i.e., decision-making).

Finally, *productivity* refers to the extent to which students make intellectual progress or improvement in their decision/solution within the observed STEM lesson (one item) and the extent to which students' final solution meets the success criteria/solution requirements (one item). In the pill-coating problem, a group demonstrates high productivity if the students reach a new/improved solution that addresses an issue or meets more success criteria/solution constraints compared to their previous solution by the end of the observed lesson. However, a group demonstrates low productivity if the students do not reach a required decision/solution and no issues with their proposed decision/solution are identified by the end of the lesson. Students' final solution indicates high productivity in product quality if it meets all success criteria and satisfies all constraints. Conversely, a final solution indicates low productivity if it does not satisfy any criteria/constraints.

Our interpretation of *productive interdisciplinary engagement* (PIE) reflects our belief that fostering students' abilities to critically engage in systematic, disciplinary-based decision-making to achieve improved decision/solution to a STEM problem is a worthy goal of integrated STEM education (Sutaphan & Yuenyong, 2019). This echoes Bybee's (2013) view that STEM literacy, which includes using STEM knowledge, skills, attitudes, and values to solve real-life problems as responsible and reflective citizens, should be the goal of STEM education.

Turning to the four design principles, *problematising* considers the extent to which the nature of the problem taken up by students is a meaningful problem for the students and STEM communities (one item). A meaningful STEM problem is complex (requires concepts/skills from more than one STEM discipline to solve), authentic (relevant to students' lives/real-life), open-ended (has more than one possible solution), extended (requires prolonged working on the problem and cannot be solved by simple search for solution, e.g., via the Internet), and persistent (occurs in multiple contexts; not a once-off problem or problem that

only exists in textbooks) (Tan et al., 2019). The pill-coating activity meets the problematising design principle to a high extent. It is complex as students require scientific inquiry and mathematical reasoning, as described under high interdisciplinarity, to optimise their solution through an iterative engineering design process. It is authentic as the production of pill coating for oral medication is real-life problem. The problem is open-ended as more than one optimal pill-coating design could exist, and it is extended as students cannot simply search for an existing answer but must work over an extended duration to solve the problem.

Problematising should be balanced by *resources*: the extents to which material resources (one item), support from teacher and/or scaffolding (one item), and instructional time (one item) are provided to facilitate students in solving the STEM problem. The pill-coating activity meets the resources design principle to a large extent if (1) students are given all the necessary materials and equipment to produce and test their pill coating; (2) students are provided a handout with adequate scaffolds, such as tables for recording data and prompts for students' justification of their final solution, and their teacher engages in group discussions to support group decision-making; and (3) sufficient instructional time is planned for students to produce and share their final solution. In contrast, the activity design meets resources to a low extent if students have to source for their own materials and equipment, are not provided any assistance via a handout or the teacher, and are given little time to work on the problem.

*Authority* depicts the extent to which students are given epistemic authority to construct their solution to the STEM problem (one item) with minimal teacher modification (one item) and to determine success criteria/solution requirements (one item). The pill-coating activity satisfies the authority design principle to a high extent if (1) students get to propose their own ideas that are acknowledged by peers for discussion, (2) the teacher critiques or highlights good points about their idea/solution without suggesting what to change, and (3) students have the opportunity to propose additional success criteria with justifications. The activity meets authority to a low extent if students do not propose their own ideas but merely follow instructions to complete the activity, and students accept the criteria given by the teacher.

Authority should be balanced by *accountability*: the extent to which students' ideas/actions are held accountable to STEM disciplinary concepts and practices through critiques (one item) by peers/teacher (one item) and through success criteria/solution requirements (one item). Examples of disciplinary concepts include mathematically/scientifically accurate concepts/facts; examples of disciplinary practices include fair test, appropriate analysis, and 2D/3D sketch/drawing norms. The pill-coating activity meets the accountability design principle to a large extent if (1) critiques of a group's ideas/solution are made by the teacher and peers beyond the group, (2) multiple critique instances involve holding ideas/solution accountable to the soundness of STEM disciplinary concepts or practices (including satisfying criteria/constraints), and (3) the criteria/constraints involve sound concepts/practices from two or more STEM disciplines. Conversely, the activity meets the accountability principle to a low extent if there are no opportunities for critique beyond respective groups, there are few critique instances during group discussion, and success criteria are not present or not made explicit to students.

### Characteristics of STEM Lessons Captured in iSTEM Protocol

An engineering or technological design context, such as a design challenge, has been recognised as a productive and common context for STEM integration (English, 2016; NRC, 2014). Thus, iSTEM protocol can be used with a lesson in an integrated STEM unit of work or activity that involves a design challenge where students draw upon their conceptual



knowledge, procedural skills/methods, and/or practices from at least two STEM disciplines to solve a real-life problem and design its solution through design processes. The solution could take the form of a design sketch/drawing, a prototype or scaled/working model, or the actual object as the solution. We identified eight relevant design processes from the engineering practices and design thinking literature (Crismond & Adams, 2012; Cunningham et al., 2020; d.school at Stanford University, 2018; Wheeler et al., 2019), which we labelled as STEM tasks and categorised into three STEM phases. The *problem definition* phase involves the tasks of context introduction and problem definition (including identification of success criteria/solution requirements). The *research* phase focuses on gathering information or data to inform the solution design and includes search (for existing information) and investigate (e.g., carry out investigations to gather data). The *development* phase includes tasks relevant to the solution development: generate (proposing ideas/representations of intended solution), concretise/make (i.e., create/revise a form of the solution), test (of the solution), and feedback (i.e., present the solution to peers/teacher for feedback and/or reflect on feedback).

### Iterative Pilot Testing, Review, and Revision

The initial iSTEM protocol, comprising 13 items for PIE outcomes and design principles rated on a 4-point scale, was conceptualised by the first author. It was then reviewed by the research team including science education and design and technology education experts (other authors). The expert review included analysing a video-recorded 2-h STEM lesson using the protocol. A key issue deliberated by the team was whether the analysis should be holistic or segmented. We initially trialled a version of the protocol with segmented analysis. A STEM lesson was analysed for the extents to which the PIE design principles and pedagogical outcomes were observed in every 15-min interval. However, we encountered the challenge that the imposed time interval could not accommodate the natural segmentation of STEM tasks as a STEM lesson progresses, and the nature of the STEM task could impact extents of PIE design principles and pedagogical outcomes. While a shorter time interval might circumvent this issue (e.g., Wheeler et al., 2019), we contend this would compromise the protocol's ease of use as judgements of the PIE outcomes and design principle items within such short time intervals are cognitively demanding. We eventually decided on a holistic analytical approach whereby ratings are indicated for the whole observed lesson. The protocol requires the observer to write descriptive notes on the lesson, which serve as evidence to support item ratings. This enables the observer to focus on observing and documenting features during the lesson and then deliberate on the relevant evidence that support the ratings post-observation. The latter is important if multiple observers are involved and consensus ratings need to be reached, and when communicating the ratings to the teachers as feedback.

The holistic iSTEM protocol was tested by two of the authors with STEM lesson recordings at primary and secondary levels, and revised following discussions with the research team to strengthen and clarify the construct definitions. Based on contrasting features of the observed lessons, one design principle item—resources: time—was added to reflect the need for provision of adequate time for students to work on STEM tasks and one outcome item, engagement: group work, was added to consider the extent to which group work was observed during a STEM lesson. To improve clarity and thus reliability of items, examples of complex student behaviours were added to illustrate their meanings (e.g., engagement and interdisciplinarity items, Part D, Appendix A). The revised iSTEM protocol was

then reviewed by education experts in mathematics, engineering, and digital technology and two expert curriculum planning officers at the Ministry of Education, Singapore, specialising in primary and secondary science/STEM education, and shared at a meeting with other STEM education curriculum planning officers and at an international conference presentation. The collective feedback led to improved clarity in the protocol descriptions and organisation.

## iSTEM Protocol Overview and Structure

The current version of the iSTEM protocol (Appendix A) comprises descriptions of the STEM phases, STEM tasks, and the PIE dimensions along with its design principles, as well as four parts to be completed. Part A captures basic information of the observed integrated STEM lesson, including the sequence of STEM phases and tasks. In Part B, the observer (1) records the duration of an observed STEM phase, (2) indicates the enacted STEM tasks and the nature of instructional activities within the STEM phase (i.e., teacher instruction, class discussion/presentation, group discussion/hands-on, or individual seat work/hands-on) via checkboxes, and (3) writes descriptive observation notes of the lesson. Besides instructions to record the participants (i.e., teacher and students) and contents of their speech and action, the abovementioned descriptions of PIE and its design principles help guide the observer on what to notice. To gather evidence for extents of PIE enacted, the observer should observe one group's interactions during group activities, if any. Points (1) to (3) are provided in individualised tables labelled for each STEM phase; tables can be repeated as necessary as the lesson proceeds. This protocol design affords the capture of STEM phases in any sequence, including repeats. Tables for recapitulation of previous lesson and reflection on lesson are also provided as these were commonly observed features of STEM lessons that do not belong to any of the STEM phases. Parts C and D of the iSTEM protocol are completed post-observation. Using observation notes in Part A as evidence, the observer assigns a rating (between 0 to 3) based on the extent to which evidence was present for each of 10 items for the four design principles (Part C) and five items for the three-dimensional pedagogical outcomes (Part D). Rubrics with four levels of evidence for each item are provided, similar to the Dimensions of Success (DoS) observation tool (Shah et al., 2018).

## Validity and Reliability Considerations

The iSTEM protocol design prioritises content validity as the items were designed using a coherent framework informed by the PDE framework. The framework serves to provide an explanation for why students demonstrate high productive interdisciplinary engagement, or not, by accounting for the extents to which the four design principles are enacted and balanced in an observed STEM lesson. Review of the protocol by experts as abovementioned provided further support for its content validity.

Due to unforeseen delays associated with the COVID-19 situation, only initial efforts have been made towards establishing the protocol reliability. Two of the authors individually coded eight STEM lesson videos (four each at the primary and secondary levels, approximately 10 h of recordings in total). Due to the small sample size ( $n=8$ ) and well-trained raters with low likelihood of guessing the ratings, we chose to use percent agreement as the indicator of inter-rater reliability instead of Cohen's kappa, which requires a recommended sample size of at least 30 (McHugh, 2012). The percent agreement for the

protocol items ranged from 100% (three items), to 88% to 63% (nine items), to 50% to 38% (three items). We acknowledge that the latter three items—interdisciplinarity (50%), authority: teacher involvement (38%), and accountability: criteria nature (38%)—did not achieve satisfactory inter-rater reliability based on the existing sample. Nevertheless, we chose to retain the items as they are highly relevant to the PIE framework. Moreover, the two coders could reach consensus ratings on all the items through discussion, suggesting that the reliability is likely to improve as more video samples are coded. We have also indicated in the iSTEM protocol (Appendix A) that the abovementioned three items have low percent agreement among raters and recommend the protocol users to review these item descriptions carefully when deliberating their ratings.

## Analysis with iSTEM Protocol

The item ratings for one of the coded lessons are presented to illustrate how the iSTEM protocol could characterise and communicate the profile of an integrated STEM lesson in terms of the extents to which students' PIE and design principles were enacted. As the ratings are not intended to be interpreted as an interval scale, no single score is assigned to any design principle or the lesson (Wainwright et al., 2003).

### STEM Lesson Context

The video footage was recorded as part of a study that observed enacted integrated STEM lessons in Singapore K-12 classrooms. Video-recordings focused on whole class instruction and one group during group activities, modelling what an observer would likely attend to during classroom observation. The reported 1-h STEM lesson was the second of a two-part STEM lesson (henceforth, Pill Coating lesson 2) (Koh & Tan, 2019). Working in groups of fours/fives, secondary school students (eighth grade equivalent) worked on the previously described pill-coating problem. Students were given a set of materials to make the coating for a chocolate candy that simulated the pill and a cup of vinegar to simulate the stomach. In lesson 1, the teacher instructed to the class on how an ingested pill makes its way through the human digestive system. Students also had some time to explore making their own pill coating in their groups.

### iSTEM Lesson Profile

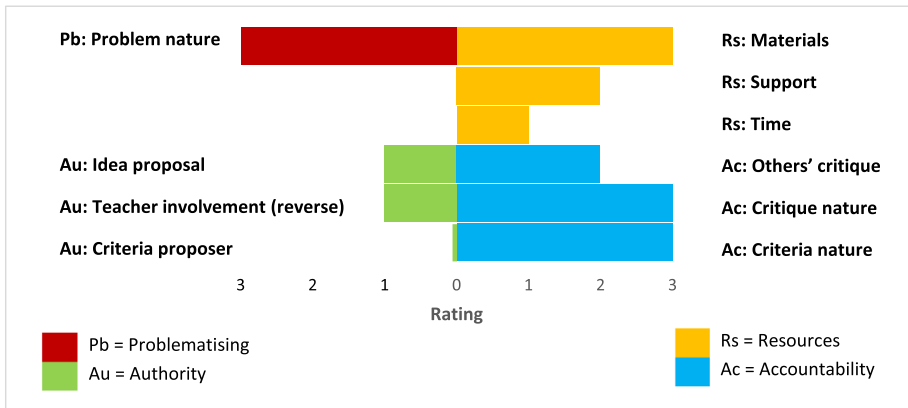
Pill Coating lesson 2 progressed through the STEM phases, tasks, and activities sequence shown in Fig. 1. The teacher, who is trained in science teaching, recapped the problem and called on a group with the longest lasting pill coating from lesson 1 to share their “recipe”. Students then spent most of the lesson researching and developing the pill coating in their groups as they first made pill coatings that satisfied one criterion at a time before making the optimal coating that incorporated all three criteria. The observed group engaged in episodes of members-only and teacher-involved discussions, as well as brief discussions with peers from other groups, during the hands-on activity. The teacher concluded the lesson by summarising and reflecting on how the STEM activity resembled real-life drug development by pharmaceutical companies.

The iSTEM profile visually displays a lesson's design principles (Fig. 2) and PIE extents (Fig. 3). For brevity, we focus on how to interpret the iSTEM profile. Readers should refer

STEM Phase	Recap	Research and Development	Reflection
STEM Task Activity	<i>Context, Problem, Development</i> Teacher instruction	<i>Investigate, Concretise/Make, Test</i> Group members-only + teacher-involved discussion Group hands-on; Inter-group discussion (limited)	Teacher instruction

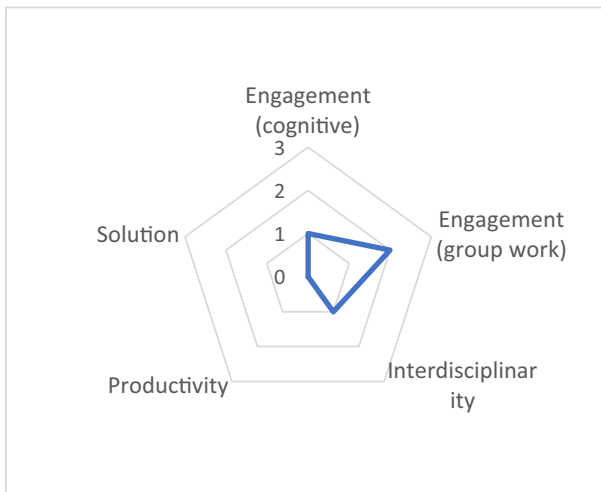
Fig. 1 Pill Coating lesson 2 iSTEM profile—design principles

*Pill Coating Lesson 2 iSTEM Profile—Design Principles*



**Fig. 2** Pill Coating lesson 2 iSTEM profile—PIE

*Pill Coating Lesson 2 iSTEM Profile—PIE*



**Fig. 3** Pill Coating lesson 2 iSTEM profile—PIE

to the iSTEM protocol items (Appendix A) to interpret specific ratings. The pill-coating problem was considered a meaningful problem for students and STEM communities (rating=3 for the problematising item in Fig. 2). Students were provided the necessary materials (rating=3) to produce the coating and adequate teacher support (rating=2), such as whole class instruction on how to proceed (produce three coatings that separately met a criterion before producing the final, optimised coating; how to divide the work among group members) and suggestion for the observed group on how to proceed with a particular investigation. Although the teacher modified the task by asking group members to

divide the work by having a student pair produce a pill coating that met two criteria while the other pair produced a pill coating that met the remaining criterion as well as the optimised coating, the observed group only had time to produce pill coatings for one criterion (rating = 1). Thus, problematising was not well-balanced by the resources to solve the problem, specifically with a lack in time as depicted in Fig. 2. Notably, all three authority items were rated low. Students only proposed ideas with the teacher's support (rating = 1) as the teacher mostly told students what to do (rating = 1; this item has a reverse rating: the greater the teacher involvement, the lower the rating), including telling the class to include oil as an ingredient (based on the "recipe" shared by the group with longest lasting pill coating) and telling the observed group how much vinegar to use in their tests. Furthermore, all the criteria were given by the teacher and accepted by students without justification (rating = 0). Conversely, accountability items were rated high as the group's ideas were mostly critiqued by the teacher (rating = 2), with some critiques by their own members. Multiple critique instances referenced empirical data (checking amounts of flour/oil previously used to determine subsequent amount to use) and held the solution accountable to criteria (does coating need to last four minutes or could some ingredients be reduced to meet cost criteria) (rating = 3 for critique nature). Furthermore, criteria considered sound disciplinary concepts/practices in two STEM disciplines (rating = 3): scientific concept (that the pill coating should last a certain duration is based on the scientific concept of protecting the drug from the stomach acid) and mathematical reasoning (reducing amount of ingredients reduces cost of pill coating proportionally, based on ingredient costs).

The observed group demonstrated low PIE (Fig. 3). Group members mostly engaged in information-seeking discourse (engagement (cognitive) rating = 1) and completed tasks through division of labour, working in subgroups within the group (engagement (group work) rating = 2). Students' decisions were often random and unjustified (e.g., students stated random amount of flour and oil to use, without justification), although the group did demonstrate disciplinary-based decision-making through the teacher's support (interdisciplinary rating = 1). In terms of productivity, the group only produced the pill coating that met the criterion for durability, and it was not obvious how this was an improvement over their initial solution since lesson 1 (productivity rating = 0). Thus, the group did not produce a final solution optimised to satisfy all criteria (solution rating = 0).

Overall, the imbalance of authority and accountability, as well as a lack of time as a resource, might explain the group's low PIE. From the evidence for authority and accountability items, students in the observed group were not forthcoming in putting forward and deliberating their ideas. A possible reason might be that the teacher's insistence that groups produced separate solutions for the criteria before producing the optimal solution might not have made sense to the students (low student authority), especially if students did not understand the concept of optimisation. This is evident as students did not realise that they should reduce coating duration before the teacher prompted them to think so. Consequentially, students did not have sufficient time to produce all the necessary solutions (c.f. low rating for time as resource), resulting in low productivity. The teacher's insistence and thus assertion of authority at the expense of students' authority might also reflect the conventional classroom norms that the students and teacher were used to. Conventional science classroom, which was familiar to the teacher, typically positions the teacher as the authority over the final form ideas in science which students are to acquire from the teacher (McNeill & Berland, 2017). Thus, students in the observed lesson might not be used to asserting their own authority and their teacher might not be comfortable with sharing authority with the students, for fear that the students might not succeed in their learning or task. The latter point is corroborated by the teacher's whole class instruction during Pill Coating lesson 1

where she insisted the students follow specific procedures to test their pill coatings, so that they would succeed in obtaining the results. As such, students might be unfamiliar with discourses involving critical discussions of their own ideas in systematic and disciplinary-based decision-making, i.e., interdisciplinary engagement. Indeed, students and teachers might require extended practice before they adapt to the discourse norms of shared authority (Preston et al., 2020) that foreground epistemic aspects of a learning activity, such as why/how a phenomenon occurs in science inquiry (Preston et al., 2022) or how reasoned decisions are reached in integrated STEM activities. Another possible reason for the low PIE might be that the students did not find the problem personally meaningful, which resulted in the students not being engaged with seeking the best solution to the problem.

## Conclusions

The proposed iSTEM protocol addresses a main shortcoming of previous K-12 integrated STEM classroom observation protocols (e.g., Dare et al., 2021; Peterman et al., 2017; Wheeler et al., 2019) by articulating a coherent set of design principles and corresponding pedagogical outcomes. Informed by the PDE framework (Engle, 2012; Engle & Conant, 2002), the iSTEM protocol serves as a viable tool to provide guidance on how to design integrated STEM experiences to meet desired pedagogical outcomes. The iSTEM protocol posits students' PIE in deliberating decisions towards a solution as a worthy integrated STEM education outcome, which was not articulated in the previous protocols. Additionally, the iSTEM protocol items corresponding to problematising, resources, authority, and accountability suggest ways through which these four design principles could be achieved and balanced to foster the 3-dimensional PIE outcomes. By articulating details of our protocol development, including our design considerations, rationale, and challenges, we hope to engage other researchers in the conversation of how to characterise and design “good” integrated STEM learning experiences.

## Limitations

We highlight three main limitations with the current iSTEM protocol. Firstly, the inter-rater reliability should be further improved with a larger sample of video-recorded integrated STEM lessons. Since the iSTEM protocol is intended for use with live lessons, the reliability of live coding of lessons should be determined and compared with coding of video-recordings. A second limitation is that the protocol needs to be tested with a variety of K-12 integrated STEM lessons to ensure its usability across lesson designs and nature of integrated STEM problems. Thirdly, we recognise that it is possible for the four design principles to be met in ways not captured by the protocol items and, likewise, for the indicators of productive interdisciplinary engagement as described in the items. Nevertheless, the items provide a relevant and coherent explanatory framework that highlights some important design principles to consider in integrated STEM learning experiences.

## Potential Use of iSTEM Protocol

The fully developed iSTEM protocol will contribute towards STEM education in two ways. Firstly, the protocol serves as a research tool for STEM education researchers to analyse K-12 integrated STEM lessons. We have illustrated how the iSTEM profile for a lesson could be used to identify strengths and inadequacies in design principles, imbalance in pairs of design principles, and the corresponding 3-dimensional pedagogical outcomes. Using the iSTEM protocol with a variety of integrated STEM lesson designs, we can characterise and identify trends across integrated STEM learning experiences, as well as theorise explanations for productive interdisciplinary engagement using problematising, resources, authority, and accountability design principles as an initial model.

Secondly, we intend for the protocol to be used as a pedagogical guide for STEM classroom teachers to improve their design of STEM learning experiences. The iSTEM profile, along with evidence from the written observation notes, could be used to communicate strengths and areas of improvements and initiate conversations with teachers on how to improve their instructional design. As teachers might have concerns about using a protocol with a rating scale that implies evaluation of their teaching, a suggestion would be to convert the ratings into nominal categories for communication with the teachers (Wainwright et al., 2003).

Finally, when interpreting and communicating the iSTEM protocol ratings, we should be mindful that the iSTEM profile presents a snapshot of a single, integrated STEM lesson enacted likely as a part of a series of lessons. For greater reliability, observations of more than one lesson within a STEM activity or unit of work should be made to identify potential profile variations due to the nature of the STEM tasks.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11165-023-10110-z>.

**Acknowledgements** We would like to acknowledge our expert members Tan Ying Chin and Wee Beng Tay from the Curriculum Planning and Development Division, Ministry of Education, Singapore.

**Author Contribution** All authors contributed to the study conception and design. Data collection and analysis were performed mainly by the first and second author, with contributions from the third and fourth authors. The first draft of the manuscript was written by the first author and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (OER 05/21 OYS) and administered by National Institute of Education (NIE), Nanyang Technological University, Singapore.

**Data Availability** The datasets generated during and/or analysed during the current study are not publicly available due to the need to uphold the anonymity of participants. An editable version of the iSTEM protocol is available from the corresponding author upon reasonable request.

## Declarations

**Ethics Approval and Consent to Participate** IRB approval had been sought and given from the ethics review board of the authors' affiliated institution. Informed consent (and assent from minors) had been sought and given from teacher and student participants.

**Competing Interests** The authors declare no competing interests.

**Disclaimer** Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE and NIE NTU, Singapore



## References




- Anwar, S., & Menekse, M. (2021). A systematic review of observation protocols used in postsecondary STEM classrooms. *Review of Education*, 9(1), 81–120. <https://doi.org/10.1002/rev3.3235>
- Bybee, R. W. (2013). The case for STEM education: Challenges and opportunities. *The Case for STEM Education: Challenges and Opportunities*.
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Choi, B. C. K., & Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clinical and Investigative Medicine*, 29(6), 351–364.
- Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, 101(4), 738–797. <https://doi.org/10.1002/j.2168-9830.2012.tb01127>
- Cunningham, C. M., Lachapelle, C. P., Brennan, R. T., Kelly, G. J., Tunis, C. S. A., & Gentry, C. A. (2020). The impact of engineering curriculum design principles on elementary students' engineering and science learning. *Journal of Research in Science Teaching*, 57(3), 423–453. <https://doi.org/10.1002/tea.21601>
- Dare, E. A., Hiwatig, B. M. R., Keratithamkul, K., Ellis, J. A., Roehrig, G., Ring-Whalen, E. A., Rouleau, M., Faruqi, F., Rice, C., Titu, P., Li, F., Wieselmann, J. R., & Crotty, E. A. (2021). Improving integrated STEM education : The design and development of a K-12 STEM observation protocol ( STEM-OP ) ( RTP ). *ASEE Virtual Annual Conference Content Access, Virtual Conference*. <https://peer.asee.org/37307>. Accessed 1 Dec 2022
- dschool at Stanford University (2018). *Design Thinking Bootleg*. [https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/5b19b2f2aa4a99e99b26b6bb/1528410876119/dschool\\_bootleg\\_deck\\_2018\\_final\\_sm+%282%29.pdf](https://static1.squarespace.com/static/57c6b79629687fde090a0fdd/5b19b2f2aa4a99e99b26b6bb/1528410876119/dschool_bootleg_deck_2018_final_sm+%282%29.pdf). Accessed 1 Dec 2022
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, 20(4), 399–483. [https://doi.org/10.1207/S1532690XCI2004\\_1](https://doi.org/10.1207/S1532690XCI2004_1)
- Engle, R. A. (2012). The productive disciplinary engagement framework: Origins, key concepts, and developments. In D. Y. Dai (Ed.), *Design research on learning and thinking in educational settings* (pp. 161–200). Taylor & Francis.
- English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education*, 3(1), 1–8. <https://doi.org/10.1186/s40594-016-0036-1>
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, 3(1), 1–11.
- Koh, J. Q. D., & Tan, A.-L. (2019). Students as pharmaceutical engineers: A biology-centric STEM task. *Teaching Science*, 1(65), 26–32.
- Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education*, 7(1), 1–16.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica (zagreb)*, 22(3), 276–282. <https://doi.org/10.11613/BM.2012.031>
- McNeill, K. L., & Berland, L. (2017). What is (or should be) scientific evidence use in k-12 classrooms? *Journal of Research in Science Teaching*, 54(5), 672–768. <https://doi.org/10.1002/tea.21381>
- Milford, T., & Tippett, C. (2015). The design and validation of an early childhood STEM classroom observational protocol. *International Research in Early Childhood Education*, 6(1), 24–37.
- Murphy, S., MacDonald, A., Danaia, L., & Wang, C. (2019). An analysis of Australian STEM education strategies. *Policy Futures in Education*, 17(2), 122–139. <https://doi.org/10.1177/1478210318774190>
- National Research Council. (2014). *STEM Integration in K-12 Education*. The National Academies Press.
- Ong, Y. S. (Ed.). (2021). Executive Report on Envisioning STEM Education in Asia Seminar. Multi-centric Education Research and Industry STEM.
- Partnership for 21st Century Skills. (2015). *Framework for 21st century learning*. [https://static.battelleforkids.org/documents/p21/P21\\_Framework\\_Definitions\\_New\\_Logo\\_2015\\_9pgs.pdf](https://static.battelleforkids.org/documents/p21/P21_Framework_Definitions_New_Logo_2015_9pgs.pdf). Accessed 1 Dec 2022
- Peterman, K., Daugherty, J. L., Custer, R. L., & Ross, J. M. (2017). Analysing the integration of engineering in science lessons with the Engineering-Infused Lesson Rubric. *International Journal of Science Education*, 39(14), 1913–1931. <https://doi.org/10.1080/09500693.2017.1359431>

- PNG Education News (2022). *Home-grown 'STEM' curriculum for PNG school of excellence*.<https://edu.pngfacts.com/education-news/home-grown-stem-curriculum-for-png-school-of-excellence>.&nbsp;Accessed 1 Dec 2022
- Preston, C. M., Hubber, P. J., & Xu, L. (2022). Teaching about electricity in primary school multimodality and variation theory as analytical lenses. *Research in Science Education*, 52(3), 949–973.&nbsp;<https://doi.org/10.1007/s11165-022-10047-9>
- Preston, C., Hubber, P., Bondurant-Scott, M., & Gunsekere, I. (2020). A representation construction approach to learning about electrical energy in year 6. *Teaching Science*, 66(2), 5–19.
- Shah, A. M., Wylie, C., Gitomer, D., & Noam, G. (2018). Improving STEM program quality in out-of-school-time: Tool development and validation. *Science Education*, 102(2), 238–259.&nbsp;<https://doi.org/10.1002/sce.21327>
- Sutaphan, S., & Yuenyong, C. (2019). STEM education teaching approach: Inquiry from the context based. *Journal of Physics: Conference Series*, 1340(1), 012003.
- Tan, A.-L., Teo, T. W., Choy, B. H., & Ong, Y. S. (2019). The S-T-E-M Quartet. *Innovation and Education*, 1(1), 1–14. <https://doi.org/10.1186/s42862-019-0005-x>
- Vasquez, J., Sneider, C., & Comer, M. (2013). *STEM lesson essentials, grades 3–8: Integrating science, technology, engineering, and mathematics*. Heinemann.
- Wainwright, C. L., Flick, L., & Morrell, P. D. (2003). Development of instruments for assessment of instructional practices in standards-based teaching. *The Journal of Mathematics and Science: Collaborative Explorations*, 6, 21–46.
- Walton, D. N. (1998). *The new dialectic: Conversational contexts of argument*. University of Toronto Press.&nbsp;<https://doi.org/10.3138/978144268185>
- Wheeler, L. B., Navy, S. L., Maeng, J. L., & Whitworth, B. A. (2019). Development and validation of the Classroom Observation Protocol for Engineering Design (COPEd). *Journal of Research in Science Teaching*, 56(9), 1285–1305.&nbsp;<https://doi.org/10.1002/tea.21557>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## Authors and Affiliations

Yann Shiou Ong<sup>1</sup>  · Jaime Koh<sup>1</sup> · Aik-Ling Tan<sup>1</sup>  · Yong Sim Ng<sup>1</sup> 

Jaime Koh  
jaime.koh@nie.edu.sg

Aik-Ling Tan  
aikling.tan@nie.edu.sg

Yong Sim Ng  
yongsim.ng@nie.edu.sg

<sup>1</sup> Natural Sciences and Science Education, meriSTEM@NIE, National Institute of Education, Singapore, Nanyang Technological University, Singapore, Singapore