Journal of Pre-College Engineering Education Research (J-PEER)

Volume 13 | Issue 1 Article 2

2023

Development and Validation of an Integrated STEM Teacher Classroom Observation Protocol

Shuqi Zhou
Donghua University, shuqizhou11@gmail.com

Hillary E. Merzdorf
Texas A&M University, hmerzdorf@tamu.edu

Kerrie A. Douglas
Purdue University, douglask@purdue.edu

See next page for additional authors

Recommended Citation

Zhou, S., Merzdorf, H. E., Douglas, K. A., & Moore, T. J. (2023). Development and Validation of an Integrated STEM Teacher Classroom Observation Protocol. *Journal of Pre-College Engineering Education Research (J-PEER), 13*(1), Article 2.
https://doi.org/10.7771/2157-9288.1357

Development and Validation of an Integrated STEM Teacher Classroom Observation Protocol

Abstract

This study aimed to develop a K-12 classroom observation protocol to assess K-12 teachers' implementation of science, technology, engineering, and mathematics (STEM) integration. The intended purpose of the observation protocol is for researchers to examine how K-12 teachers implement the STEM integrated curriculum. Based on research on STEM integration, the protocol covers eight categories: *Define, Learn, Plan, Try, Test, Decide, Improve,* and *Instructional Strategies*. The protocol was reviewed by a panel of integrated STEM education experts, engineering design experts, and educational assessment experts ($n = 15$) and revised accordingly. To test the reliability of using the protocol, two raters applied the protocol to videos of 12 lower-level elementary teachers located at seven schools who implemented an integrated STEM curriculum over a total of 196 class periods. The inter-rater reliability Gwet AC1 coefficient was 0.73. The inter-rater reliability of each category's coding was also calculated and reported. We discuss the implications and limitations of the protocol.

Keywords

classroom assessment, observation protocol, STEM integration, teacher implementation

Document Type

Research Article

Cover Page Footnote

We are grateful to the experts who participated in the review of the protocol and shared their insights with us about the protocol. Disclosure of potential conflicts of interest: The authors declare that they have no conflict of interest. Ethical Statement: The work described is original and has not been previously published. The manuscript was approved by all authors. Consent Statement: An appropriate consent form was used for the participants.

Authors

Shuqi Zhou, Hillary E. Merzdorf, Kerrie A. Douglas, and Tamara J. Moore

Available online at http://docs.lib.purdue.edu/jpeer

Journal of Pre-College Engineering Education Research 13:1 (2023) 20–36

Development and Validation of an Integrated STEM Teacher Classroom Observation Protocol

Shuqi Zhou[1], Hillary E. Merzdorf[2], Kerrie A. Douglas[3], and Tamara J. Moore[3]

[1]*Donghua University*
[2]*Texas A&M University*
[3]*Purdue University*

Abstract

This study aimed to develop a K-12 classroom observation protocol to assess K-12 teachers' implementation of science, technology, engineering, and mathematics (STEM) integration. The intended purpose of the observation protocol is for researchers to examine how K-12 teachers implement the STEM integrated curriculum. Based on research on STEM integration, the protocol covers eight categories: *Define*, *Learn*, *Plan*, *Try*, *Test*, *Decide*, *Improve*, and *Instructional Strategies*. The protocol was reviewed by a panel of integrated STEM education experts, engineering design experts, and educational assessment experts ($n = 15$) and revised accordingly. To test the reliability of using the protocol, two raters applied the protocol to videos of 12 lower-level elementary teachers located at seven schools who implemented an integrated STEM curriculum over a total of 196 class periods. The inter-rater reliability Gwet AC1 coefficient was 0.73. The inter-rater reliability of each category's coding was also calculated and reported. We discuss the implications and limitations of the protocol.

Keywords: classroom assessment, observation protocol, STEM integration, teacher implementation

---

Development of an Integrated STEM Teacher Classroom Observation Protocol

STEM (i.e., Science, Technology, Engineering, and Mathematics) integration aims to engage students more meaningfully through personal and real-world experience by connecting different disciplines, students' prior knowledge, and new experiences (e.g., Brooks & Brooks, 1993; Capraro & Slough, 2008). STEM integration provides students with a richer understanding of engineering practices and how mathematics and science are combined with engineering, which encourages students' interest in STEM (Mustafa et al., 2016; National Research Council [NRC], 2013; Riskowski et al., 2009), motivation toward STEM learning (Wang et al., 2011), and achievement (Hurley, 2001). Through students' exposure to different disciplines, STEM integration can prepare students for college and future career choices (NRC, 2013). Integrated STEM education effort is called worldwide, and teaching of integrated STEM is more prevalent (e.g., Estapa & Tank, 2017; NRC, 2014). However, not all integrated STEM promotes students' critical thinking or engages students with the content in a meaningful way (Moore et al., 2020). K-12 teachers are responsible for implementing integrated STEM education and play an important role in students' learning. Yet, little is known about the quality of teachers' implementation of integrated STEM education and how to assess it. In order to ensure that students have learning opportunities to engage meaningfully in integrated STEM (Stohlmann et al., 2011), there is a need to develop tools to observe and assess the pedagogical practices of teachers in integrated STEM learning environments so that they can be provided with feedback regarding their practices and make improvements.

http://dx.doi.org/10.7771/2157-9288.1357

The Next Generation Science Standards (NGSS) highlight the core ideas of engineering design "by raising it to the same level as scientific inquiry in science classroom instruction" (NRC, 2013, p. 1). Scholars (e.g., Bryan et al., 2016; Guzey et al., 2020; Moore et al., 2020; Roehrig et al., 2021) proposed that the STEM integration model should explicitly apply engineering design processes to solve real-world problems. High-quality integrated STEM curricula foster a deeper understanding of science and mathematics concepts through solving an engineering design challenge. However, researchers have also found that solving an engineering design challenge does not necessarily foster deeper content knowledge or skill, as successful integration depends on teachers' knowledge and reflection on the design process, and on their ability to translate design cognition research into classroom practice (Crismond & Adams, 2012). Equally important is that students are expected to learn and apply specific content to make design decisions. Students' ability to successfully apply science and mathematics towards engineering design depends in part on the classroom instruction and support of teachers. Thus, this study aims to develop a classroom observation protocol for teachers' implementation of STEM integration in the classroom. Integrated approaches to STEM education are more common in K-12 classrooms for students of diverse backgrounds, examining many factors such as problem-solving abilities, academic achievement, and attitudes and career interest towards STEM (e.g., Guzey et al., 2017; Kurt & Benzer, 2020; Lie et al., 2019). The main purpose of the observation protocol is to act as a tool for researchers to examine how K-12 teachers implement the STEM integrated curriculum developed by teachers and researchers. Our long-term goal of this research is to more widely inform and improve teachers' classroom practices by developing a process for observing teachers' best practices in STEM integration.

STEM Integrated Approaches

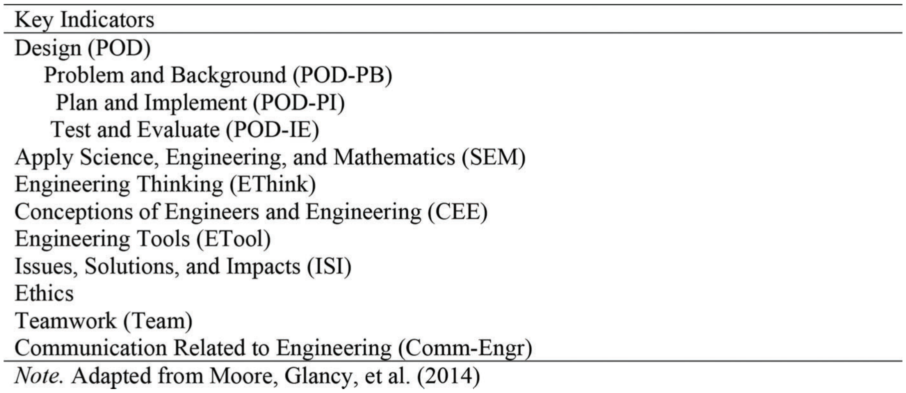
Content and Context Integration

Teachers have flexibility on how to integrate STEM in their classrooms (Moore, 2008). Content and context are two different ways for STEM integration to integrate content and thinking. Content integration emphasizes the merging of the content fields into a single curricular activity or unit and highlighting "big ideas" from multiple content areas (Roehrig et al., 2012). Content integration involves multiple STEM disciplinary learning objectives (Moore et al., 2014c) for the activity or unit, which integrates not only mathematics and science content but also technology and engineering learning. Through content integration, students learn content from different STEM disciplines and solve problems using these different disciplines together. Context integration focuses on the content of one discipline and uses contexts from others to make the content more relevant (Roehrig et al., 2012). The content may lead to different knowledge points, but the context is responsible for iterations and designing of solutions (Roehrig et al., 2012). Both approaches represent ways that teachers may integrate STEM in the classroom based on curriculum and learning outcomes. For context integration, the process or project provides a means of scaffolding STEM learning, which allows students to explore different aspects of knowledge (Nadelson & Seifert, 2017). The context gives students a shared learning experience for studying various STEM topics, which facilitates information transfer by enabling students to build on shared understanding and experience (Nadelson & Seifert, 2017). The use of context in STEM integration also increases the meaningfulness and relevancy of STEM content and is likely to increase students' learning motivation and engagement (Nadelson & Seifert, 2017). In addition, context integration has been shown to have a larger effect on students' STEM achievement during STEM integration than content integration (Zhou et al., 2023). It should be noted that the complexity of the integrated STEM context should be aligned with students' knowledge level to ensure the success of integrated STEM learning (Nadelson & Seifert, 2017).

Conceptual Frameworks

Engineering Design-Based Science/Mathematics Integrated Instruction

In this study, our integrated classroom observation protocol is based on STEM integration defined by Moore et al. (2014a). Design is one of the effective practices to help students learn science and engineering (National Academies of Sciences, Engineering, and Medicine, 2019). Moore and Smith (2014) specifically defined STEM integration as students participating in engineering design as a means to develop relevant technologies that require meaningful learning through integration and application of mathematics and science. Based on previous research (Brophy et al., 2008; Frykholm & Glasson, 2005; Moore et al., 2014b; Morrison, 2006), Moore et al. (2014c) put forward that high-quality integrated STEM learning experiences include, but are not limited to, a motivating and engaging context that engages students in engineering design challenges, allowing for students to learn from failure and participate in redesign and to use relevant contexts for the engineering challenges to which students can personally relate. Integrated STEM learning experiences also require the teaching and learning of relevant science and mathematics content, the ability to engage students in content using student-centered pedagogies, and the opportunity to promote communication skills and teamwork. Moore et al. (2014a) proposed a framework for quality K-12 engineering education and highlighted the key indicators for quality STEM education



Key Indicators
Design (POD)
Problem and Background (POD-PB)
Plan and Implement (POD-PI)
Test and Evaluate (POD-IE)
Apply Science, Engineering, and Mathematics (SEM)
Engineering Thinking (EThink)
Conceptions of Engineers and Engineering (CEE)
Engineering Tools (ETool)
Issues, Solutions, and Impacts (ISI)
Ethics
Teamwork (Team)
Communication Related to Engineering (Comm-Engr)

Note. Adapted from Moore, Glancy, et al. (2014)

Figure 1. The framework for quality K-12 engineering education.

(see Figure 1 for detailed indicators). In this framework, engineering design problems are the fundamental context to integrate science and mathematics.

Engineering integrates other subjects (Moore et al., 2014a) in STEM education, as engineering provides a real-world context for learning mathematics and science knowledge. Engineering design is crucial to connecting STEM learning activities such as mathematics and science thinking by providing an authentic context for mathematics and science content knowledge (Kelley & Knowles, 2016). It encourages learners to apply content knowledge to make informed decisions during engineering design processes. Previous studies have shown that students have positive science content knowledge gains throughout design-based STEM integration. For example, Apedoe et al. (2008) found that students better grasped fundamental chemistry concepts after completing an eight-week high school engineering design-based unit on heating and cooling. Fortus et al. (2004) saw improvements in students' understanding and application of electricity, thermal energy, and environmental safety after implementing high school engineering design-based physical science units producing 3D prototypes, diagrams, and models. Engineering design is a highly iterative process and includes steps such as: identify the need of a problem, research the need or problem, develop possible solutions, select the best possible solutions, construct a prototype, test and evaluate the solutions, communicate the solutions, and redesign (Massachusetts DOE, 2006). By participating in design, students translate principles to practice and see how knowledge from multiple science and mathematics disciplines can inform a solution to a complex engineering problem.

There are also challenges to using engineering as a context to teach science and mathematics. For instance, engineering, mathematics, and science all come from distinct epistemological traditions. While engineering does rely on the application of math and science ideas, it is distinct from learning about math and science ideas within the framing of those disciplines. Integrating engineering also needs teachers who are familiar with their discipline's content and engineering practices (Brand, 2020). Teachers' lack of practical knowledge of integrating engineering, insufficient class time, and a lack of materials, training, space, resources, and administrative support are all barriers to integrating engineering into the classroom (Anderson-Rowland et al., 2003; Cejka & Rogers, 2005; cited in Brand, 2020). Integrated STEM in the classroom needs teachers who are thoroughly knowledgeable about the scientific and mathematics principles that students are asked to apply, and who are equally confident in guiding students through the design process to solve engineering problems.

Engineering design challenges are different from scientific inquiry, as scientific inquiry is motivated by an interest in generalizable findings and does not necessarily create a tangible solution to a real-world problem (Householder & Hailey, 2012). Engineering design challenges provide the context for developing problem-solving skills and promoting communication skills and teamwork (Roehrig et al., 2012). Engineering design pedagogy is also one of the required methods for science teachers according to the *Framework for K-12 Science Education Practices, Crosscutting Concepts, and Core Ideas* (NRC, 2012). The NGSS differentiate scientific inquiry from engineering design by stating that scientific inquiry "involves the formulation of a question that can be answered through investigation," while engineering involves "the formulation of a problem that can be solved through design" (NRC, 2012). According to the NGSS, both disciplines ask questions and define problems, develop and use models, plan and carry out investigations, analyze and interpret data, use mathematics and computational thinking, construct explanations and design solutions, engage in argument from evidence, and obtain, evaluate, and communicate information (NRC, 2012). Scientific inquiry explains phenomena with empirically testable questions, employs models for making hypotheses, and conducts investigations to collect data and draw conclusions. In contrast, engineering practice translates needs in societies and systems into problems with measurable criteria and constraints, creates models for analyzing systems and predicting failures, and investigates the performance of a solution in meeting design criteria. Both practices rely on mathematics to conduct tests and explain behavior, and engineering solutions meet the concrete needs of society with a range of possible solutions. While scientific argumentation involves reasoning from data and established theory, engineering reasoning evaluates the strengths and weaknesses of a solution on its ability to meet the needs of stakeholders in the problem, based on test data and multiple revisions as evidence.

Evidence-Based Reasoning

Evidence-based reasoning derives from Toulmin's (2003) argument pattern (i.e., claim, data/grounds, warrant, backing, a qualifier, rebuttal) and links to scientific and mathematics principles. Engineers use evidence to justify design ideas and decisions to each other and their relevant stakeholders from a variety of sources such as scientific theory, experimental data, mathematical reasoning, societal needs, or technical criteria (ABET, 2016; Dyba et al., 2005; Van Epps, 2013). Evidence-based reasoning is a means of connecting this information into a coherent rationale for the design solution. This narrative will serve to guide the design process and communicate needs, decisions, and results to stakeholders. It is also flexible throughout engineering design as new evidence is acquired through research, discussion, and prototype testing. Thus, evidence-based reasoning is adopted in our definition of engineering design. Evidence-based reasoning is mainly reflected in the Plan and Decide categories in our protocol as students develop a rationale for their design features and decisions and interpret test results to draw conclusions about their design performance. The observation protocol is developed from the research mentioned above regarding what makes for high-quality integrated STEM learning (Moore et al., 2014a) and includes pedagogies that are supportive of the engineering design process (i.e., Define, Learn, Plan, Try, Test, Decide, and Improve). Instructional strategies are also listed as an important part of the protocol. Other important aspects of STEM integration such as evidence-based reasoning and integration of mathematics and science content are also incorporated under different categories as necessary.

Preexisting Teacher Classroom Observation Instrument in STEM Field

Several existing classroom observation instruments have been developed to capture different perspectives of classroom teachers in the STEM field and are widely used. Most instruments are for prospective teachers, e.g., the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002), UTeach Observation Protocol (UTOP; Walkington et al., 2012), and Classroom Observation Protocol (COP; Horizons Research Inc., 2000). Additionally, there are other researchers exploring observation protocols for college STEM classrooms, e.g., the Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al., 2013), Mathematics Classroom Observation Protocol for Practices (MCOP^2; Gleason et al., 2017), and Practical Observation Rubric to Assess Active Learning (PORTAAL; Eddy et al., 2015). Shekhar et al. (2015) developed an observation protocol to study instructor and student engagement in college STEM classrooms. Recently more protocols have been developed to capture engineering design instruction: Classroom Observation Protocol for Engineering Design (COPED; Wheeler et al., 2019) and Pre-Kindergarten Engineering Observation Protocol (PREEOP; Bagiati & Evangelou, 2018). COPED is focused on secondary science classrooms while PREEOP is pre-kindergarten. Chen and Terada (2021) developed Interactive-Constructive-Active-Passive (ICAP) to Measure by Observation NGSS Science Practice Implementation in the Classroom from eight practices. This protocol facilitates an understanding of how students can effectively engage in scientific practices in the classroom aligning NGSS. None of these protocols are designed to capture teachers' implementation of STEM integration.

Few existing STEM observation protocols have been developed to explore teachers' implementation of STEM integration in K-12 education settings. Judson (2013) developed Mathematics Integrated into Science: Classroom Observation Protocol (MISCOP) to examine the quality of science lessons when mathematics is integrated. Capobianco and Rupp (2014) adapted the Inquiring into Science Instruction Observation Protocol (ISIOP) and designed an Engineering Design-based Classroom Observational Rubric for Grade 5 and Grade 6 students, aiming to observe and characterize design-informed pedagogical methods employed by STEM teachers. However, neither of these studies provides clear validity evidence of the observational rubric. This study addresses the need for a well-developed observation protocol in K-12 education examining teacher implementation of STEM integration from an engineering design-based perspective, including content and support of student reasoning as well as pedagogical practices.

Phase 1: Development of the Draft Protocol and Panel Review

Development of the Draft Protocol

The development of the observation protocol was an evolutionary process extending for more than a year. Our research group consisted of experts in engineering design, educational assessment, and STEM education. We created an initial draft of the protocol based on the well-established theoretical framework in STEM education by Moore et al. (2014a) and iteratively revised our protocol through eight different versions through weekly discussions with the experts in our team,

http://dx.doi.org/10.7771/2157-9288.1357

following experts' suggestions and insights. The format, procedure, and rating strategy changed, and the overall content of categories and subcategories continued to evolve.

When trying to develop an assessment tool that can be broadly used, there are trade-offs between creating a tool sensitive enough to capture nuances and reducing subjectivity to increase consistency between raters or users. Our protocol is intended to capture differences between classroom experiences while also reducing the amount of rater subjectivity. We used dichotomous scoring instead of scoring with more than two scale points since we found that a checklist-type protocol was easier for raters to indicate whether a behavior was observed. Rather than asking raters to indicate their observations by degrees, specific sub-objectives are less ambiguous for raters to rate through dichotomous rating. A binary yes/no scoring approach is more consistent with the purpose of an observation protocol, as we have detailed descriptions of expected teacher actions and can therefore indicate with a high level of specificity whether or not those actions were performed. We described different levels of achievement corresponding to each category, and Table 1 gives one example of the indicator. However, we found that separate rubrics were easier and more feasible to observe and rate. This led us to change the content of level achievement into several separate rubrics, and to change the rating of the protocol into a binary checklist (i.e., "yes" and "no"). The final format of the protocol has columns with categories, pedagogical content goals, rubric, and notes (see the Appendix for the final format).

There are seven categories, namely Define, Learn, Plan, Try, Test, Decide, and Instructional Strategies, in our draft protocol. These categories were based on the STEM integration framework by Moore et al. (2014a) who outlined the core processes of engineering design and the learning activities which take place within them. We derived our categories from the framework's key indicators (see figure 1), primarily the Process of Design but also Apply Science, Engineering, and Mathematics, Teamwork, and Communication. Moore et al.'s (2014a) framework is student-centered, but our observation protocol is teacher-focused. Therefore, our challenge was to adapt these general definitions into specific, observable behaviors used by teachers during STEM integration. For each component of the framework, we translated Moore et al.'s (2014a) key indicators into actions performed by teachers to support the learning objectives and goals of a STEM integration classroom unit. In addition to the Process of Design key indicators, panel experts strongly suggested a separate Improve section for the protocol (see details in the section Detailed Panel Experts' Feedback). Thus, there are eight categories in our final protocol. Pedagogical content goals are the overarching description for each category. We then divided pedagogical content goals into separate objectives and objectives into sub-objectives. A total of 28 objectives and 104 sub-objectives are listed in our final version of the protocol. All objectives in the protocol are STEM integration best practices, but we do not anticipate every class will address every key objective in the protocol. The format of the current protocol is presented in the Appendix. The overall description of each category is presented as follows.

In the Define category, the teacher presents the engineering challenge and introduces/reviews the design process. The teacher facilitates students to understand the scope of the problem by asking questions and interacting with the client. The teacher encourages students to consider what is the fundamental need of stakeholders in the problem. In the Learn category, the teacher gives students an opportunity to build their knowledge for the design through research for the purpose of developing design solutions. The research informs their design and allows them to further scope the problem. This research can include mathematics and science concepts, existing design solutions, and other strategies or background knowledge that students already have that can be of use. In the Plan category, the teacher instructs students to generate multiple design ideas and select potential solutions based on evidence learned about the problem during idea selection. The teacher instructs students to communicate the details of their design ideas, the decisions that students made, and how they justify their decisions with evidence. In the Try category, the teacher facilitates students to put the plan into action by building a testable representation of a solution (e.g., building a prototype, process). The teacher instructs students to make modifications to their prototype in order to better meet criteria/constraints if needed. In the Test category, the teacher instructs students to consider testable questions (e.g., hypotheses or predicted outcomes) about the performance of their design prototype, model, or product. The teacher has students collect specific results from tests and instructs students to analyze their collected results for evaluating solution performance. In the Decide category, the teacher asks students to decide the overall performance of

Table 1
Initial format of the protocol with examples.

Category	Indicator	Level of achievement			
		0	1	2	3
Define	Teacher requires students to use engineering design processes (e.g., problem scope, background, plan, implement, test, evaluate).	Elements of the design process are not communicated to students.	Teacher mentions design elements.	Teacher tells students what is expected in each design element.	Teacher tells students what is expected in each design element and structures lessons according to design process elements.

their design. The teacher instructs students to communicate and justify whether their design decision is appropriate based on evidence. The Instructional Strategies category is meant to observe whether the teacher promotes collaborative and cooperative learning in teamwork, whether the teacher implements a student-centered approach, and how teachers provide feedback. The teacher encourages students to look at problems in multiple ways and evaluates their conceptual understanding.

The observation protocol developed in this study can be used for the purpose of research. For instance, researchers can evaluate teachers' fidelity to implementing a curriculum that is designed from a STEM integration perspective. In addition, the tool can also be used to identify teachers' implementation of STEM integration across different class lessons according to the objectives and sub-objectives.

Panel Review

After we created the draft of the protocol, a panel of experts from STEM education was formed to give feedback on the draft protocols. There were 15 panel experts consisting of highly regarded professors from STEM education (H-index ranging from 9 to 33), graduate students, and undergraduates who majored in STEM education at a Midwest research university. Professors were experts on educational theory, curriculum and pedagogy, STEM education, assessment and evaluation, and design theory. Graduate students also had extensive teaching experience in K-12 and college STEM classrooms. We collected experts' feedback through questionnaires about whether the seven categories captured the essential pedagogies in engineering design-based STEM integration for observation from a big picture, whether each category in the observation protocol included all the essential pedagogical goals of lessons, and whether sub-objectives were aligned to their corresponding learning objectives. We also asked what other important pedagogical goals should be added, whether the sub-objectives were clear, and whether there were any needed sub-objectives. In addition, we asked experts to give suggestions beyond the above questions. Besides the questionnaire, we also interviewed two professors about the protocol. We received detailed and useful feedback from the panel review and revised our protocol according to the feedback. We added the "Improve" category since multiple experts strongly recommended adding this category. There are eight categories in our final version of the protocol: Define, Learn, Plan, Try, Test, Decide, Improve, and Instructional Strategies. The following section discusses the detailed panel experts' feedback.

Detailed Panel Experts' Feedback

The protocol went through 10 rounds of revisions: eight rounds of draft development process and two rounds of feedback. Panel experts gave detailed feedback, and we revised the protocol according to their feedback. The specific feedback is the following. For the Define section, subject matter experts observed that asking students to explain the problem, without requiring a rationale, is not realistic as a separate item. Having students define why the problem is worth solving, relating the problem to students' own lives, and asking questions about the problem were recommended for this section. One expert also noted that not all engineering problems will have a client. Experts also gave feedback on alignment and objectives which did not fit. In the Learn section, sub-objectives were proposed for learning about stakeholders, and collecting documentation to inform design. This stage was suggested to be about teacher–student discussion to prepare for the design, as well as students identifying what they need to learn. The experts proposed that teachers do more to reinforce essential mathematics and science concepts, which translated to connecting them to the design problem.

For the Plan section, experts suggested that students' ability to use science and mathematics knowledge as justification should not necessarily be differentiated. One expert also noted that communication of design ideas should be separate from design decisions and that multiple design solutions should be developed. Experts also emphasized the importance of having teachers remind students of criteria, constraints, and simplifying assumptions during planning, and suggested that students weigh their relative importance. Having supporting evidence for justifications was noted. One expert recommended that the use of evidence to make decisions be a separate section in the Plan. Feedback was also given about when to observe students' generation and implementation of multiple design ideas.

Multiple experts strongly recommended a separate Improve section for the protocol. Iterative design improvement based on test results was differentiated from modifications in the Try stage, where it may occur but is not limited to design construction. Try was noted as potentially yielding designs that are not physical products, with significant changes unlikely to occur at this stage. One expert suggested ways for teachers to ask students about the problems they may have encountered during Try and how students intend to update the design, as well as match between the initial design and current prototype.

For the Instructional Strategies, expert feedback was given on teamwork, clarifications that students address team conflicts, and suggestions for how each student contributes to the team. Alignment between sub-objectives and pedagogical content goals was also reinforced by experts. Overall, experts made recommendations for clarity, precision, and ease of use by teachers which streamlined the protocol. Experts asked questions to ensure that our meaning and language were specific for each item. We adopted experts' suggestions and revised the protocol accordingly.

Phase 2: Inter-rater Reliability

Inter-rater reliability (IRR) refers to the degree of the consistency of independent coders' or reviewers' decisions and is an important component of instruments' validity. In this study we used Gwet AC1 (Gwet, 2008) as the IRR coefficient.

Background and Participants

The videos we used in our research were from Grade 1 and Grade 2 classes implementing STEM integration curricula. Student ages were from six to eight years old. The videos were part of a larger NSF-funded curriculum development and teacher professional development project for P-12 teachers implementing integrated STEM units in their classrooms. Our access to the videos was provided by coauthors who are in this NSF-funded project. We chose to observe Grade 1 and Grade 2 classrooms due to the engineering design process being more simplified than middle school grades and the design challenge being less complex, making it easier to observe the activities of interest. Three different units of the curriculum were observed, and teachers were recorded when they implemented the units. The development of the curriculum units was guided by a STEM integration research paradigm to (a) deepen student understanding of STEM disciplines by contextualizing concepts, (b) broaden student understanding of STEM disciplines through exposure to socially and culturally relevant STEM contexts, and (c) increase student interest in STEM disciplines to expand their pathways for entering STEM fields (Roehrig et al., 2012). The units were also built from the Framework for Quality STEM Integration Curriculum (Moore et al., 2014a), with each unit intentionally including a motivating and engaging context, meaningful mathematics and science content, student-centered pedagogies, an engineering design task, and teamwork and communication skills. The observed three units use three separate engineering design challenges but have equivalent structures. A client with a problem was introduced by the teacher, and students were asked to develop a design solution following the same engineering design process steps. Teachers spent around two weeks on the entire curricular unit. Most often, classrooms had a single teacher, but occasionally included a teaching aide. Twelve teachers from seven elementary schools and 196 class periods were observed in this research. In total, we observed approximately 4,200 hours of classroom instruction.

Methods

Training

Before the training, trainers pre-coded with experts' ratings as a reference. First, observers became familiar with the protocol. Trainers introduced the purpose and content of the rubrics, and observers were advocated to ask questions. Then, two observers were asked to code the video lectures. After coding, we asked observers to compare their codes and discuss them with each other. We also strongly advocated observers put forward questions at this point. After no uncertainty remained for either coding, the raters were asked to code other classroom videos again. We calculated the IRR at this point, and the training procedure was not completed until the IRR of the two coders reached at least 0.75. Classroom videos from two teachers (23 videos in total) were coded as a smaller sample to reach consistency. Then, we asked the observers to apply the protocol and code the remaining classroom videos from additional teachers. During training, observers made notes about observing protocol items that were applied to revisions.

Analysis Method

After observers coded all the videos, we calculated the IRR. The Gwet AC1 coefficient is the most accurate of all agreement coefficients when there are only two raters (Gwet, 2008). This statistic accounts for an unequal distribution of observations across categories (Hallgren, 2012) as well as the possibility of two raters agreeing by chance (Tong et al., 2019). Since there were only two raters in our study, we used the Gwet AC1 coefficient as the IRR statistic. We also used the Gwet AC1 coefficient as the IRR statistic in the training phase. First, we calculated the overall IRR of all 196 videos' coding. Then, we calculated IRR for each category. We conducted all the analysis through R software Rel packages. We used Landis and Koch (1977) benchmarks (see Table 2) for evaluating the magnitude of IRR. Their benchmark is the most classic and popularly adopted (Tong et al., 2019). Since our protocol will not be used as a high-stakes assessment,

Table 2
Landis and Koch kappa benchmark (Landis & Koch, 1977).

Kappa statistics	Strength of agreement
<0.00	Poor
0.00–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–1.00	Almost perfect

we selected the moderate level (0.41–0.6) from Table 2 following the study of Tong et al. (2019). For each category level, IRR should meet the lower bound of the moderate range of 0.41, and the overall IRR should be above the upper bound of 0.6.

Reliability Results

The overall IRR (i.e., Gwet AC1 coefficient in this research) for all categories and sub-categories of 196 videos' coding is 0.73, which is substantial (Landis & Koch, 1977). The range of coefficients for the categories was 0.54 (Test) to 0.89 (Improve), all of which are in an acceptable range. Table 3 displays the overall Gwet AC1 coefficient and each category's coefficient, along with the strength of agreement interpretation as described by Landis and Koch (1977).

Discussion

While integrated STEM has been increasingly popular in pre-college settings, resources for evaluation and research on how teachers implement STEM integration in their classrooms are scarce. This research developed a classroom observation protocol focusing on teachers' implementation of STEM integration in K-12 classroom settings. Our integrated STEM classroom observation protocol is the first classroom observation instrument to collect information about teachers' STEM integration practices. Because teachers have varied approaches to classroom instruction in general and STEM integration specifically (Douglas et al., 2018), the quality of student learning depends on teacher implementation. This makes the information collected through our protocol valuable for helping to inform and improve teachers' instruction.

Validity of the Instrument

The integrated STEM teacher classroom observation protocol is intended to assess how teachers implement STEM integration in the classroom from an engineering design-based perspective, while it is not designed to be related to any other external criteria. During the early stages of instrument development, we support validity by showing alignment between observation items and the theory they are intended to represent. Validity can be defined as building a case with evidence to support claims about an instrument and its scores. There are several claims we make about our protocol and its outcomes. We propose that effective teacher implementation of STEM integration is represented by the protocol items. The protocol is developed based on the existing well-established theoretical framework in STEM education (Moore et al., 2014a). Its categories are grounded in theory and evidence of learning indicators for STEM integration, as well as effective STEM integration pedagogical practice sources of evidence. It was also reviewed by experts with extensive experience with STEM education (i.e., faculty and graduate students in related fields). Preliminary validity is established through the revisions according to the feedback from these experts in STEM-related education, so that the protocol more fully captures teaching and learning behaviors for STEM integration. In addition, we have preliminary evidence that the protocol was sensitive enough to detect differences in teaching behaviors between classrooms. Future research will continue investigating the performance of protocol criteria and categories through group comparison.

The IRR results demonstrated agreement by raters on construct definitions, measurement, and indicators at each level. However, it should be noted here that the IRR of some sections is relatively low (e.g., Test, Decide) though it is still moderate. One possible reason is that some of the sub-objectives listed in the section are not obvious for the observer to catch. For instance, it is difficult for the observer to judge whether the teacher "instructs students to consider the criteria and constraints of the problem during testing." Since "criteria and constraints" are sometimes also mentioned in previous sections, the observer tends to check yes even if they do not appear in the test section. The observers need to be well trained and consistent in checking these sub-objectives.

Table 3
Final IRR result by category.

Category	Gwet AC1 coefficient	Strength of agreement
Define	0.74	Substantial
Plan	0.58	Moderate
Learn	0.61	Substantial
Try	0.61	Substantial
Test	0.54	Moderate
Decide	0.54	Moderate
Improve	0.89	Almost perfect
Instructional Strategies	0.58	Moderate
Overall protocol	0.73	Substantial

Use for Research and Evaluation

The integrated STEM teacher classroom observation protocol is based on engineering design-based science and mathematics integrated instruction and includes pedagogies supportive of the engineering design process (i.e., Define, Learn, Plan, Try, Test, Decide, Improve). Instructional strategies are also listed as a separate category in our protocol to better understand teachers' practices. The protocol can be used to characterize teachers' implementation of STEM integration across different class lessons using evidence collected from the protocol, i.e., the rating of sub-objectives. Since different class lessons only cover separate categories of the protocol, the rating of sub-objectives of these categories provides evidence about teachers' overall implementation of STEM integration based on engineering design perspective. The observers can use the observation tool and rubric in tandem, such as for debrief and reflection on teaching practice (Dillon et al., 2020; Gabriel, 2017). Observers can easily tell teachers who are closely following best practices in STEM integration and those who are not through the protocol via dichotomous checking. For instance, we have found that teachers spent most time guiding students to learn about the problem. However, teachers normally did not connect scientific or mathematics concepts to the design problem. In this case, observers can refer to the Learn part, "Discusses relevant science/mathematics concepts" would be checked as "yes," while "Connects scientific or mathematics concepts to design problems" would be checked as "no." And any other comments or notes can be noted down in the "Note" column.

Our protocol also captures nuanced aspects of engineering design between students and teachers, such as decisions, conversations, and reflections, which are difficult to capture but have strong implications for the success of learning and STEM integration. The protocol also reflects teachers' fidelity to implementing a curriculum that is designed from a STEM integration perspective. Through the sub-objective ratings, for instance, if teachers did not ask students to justify based on a correct understanding of mathematics/science concepts or instruct students to apply results (e.g., data) gathered from testing to evaluate their design, these show that the teacher did not fully implement the curriculum as intended. In addition, the information collected from the protocol could also provide feedback about teachers' instruction practices. It would be a beneficial resource for teachers' professional development by increasing their awareness of best practices in STEM integration and allowing them opportunities to reflect on their own teaching practices. This observation protocol has potential research value for gaining more detailed insight into how teachers perform STEM integration in K-12 classrooms, and how these actions support student learning. Research is needed to continue beyond quality of implementation to explore teachers' motives and decisions during STEM integration. The categories and actions articulated by our protocol provide a strong foundation for identifying strengths and challenges, and the resulting values can give researchers insight into aspects of STEM integration where teachers may need support. It may also be used as one measure in larger investigations into STEM integration effectiveness using many sources of evidence.

Although the IRR values meet the benchmarks we selected, some categories' reliability values are less than ideal. Engineering design is not a linear process with discrete steps, making it difficult to divide STEM integration teaching behavior into exclusive categories and determine which activities happened when. This protocol used dichotomous yes/no scoring for each sub-objective which improved the overall reliability of each category rating. This design is much different from traditional presentation of multiple scale scoring systems, which is more ambiguous and requires more training of raters to achieve reliability. Raters have greater subjectivity when using the traditional multiple scale scoring systems. However, for the dichotomous scoring system we use in this study, raters only need to check yes or no, which is more objective.

Limitations

There are several limitations to this study. First, due to the restriction of the data, we are only able to develop the protocol based on teacher instruction of Grade 1 and Grade 2 classes, and therefore we cannot use factor analysis or generalizability theory to develop our instrument with our small sample size of observed teachers. Future research may develop the protocol for use with teachers across different grade levels and use more advanced psychometrics and measurement approaches to evaluate its performance. This could provide insight into whether all categories of STEM integration instruction continue to be relevant in the higher grades, and whether more are needed to capture the expectations of more sophisticated reasoning and justification. Second, some sub-objectives in the protocol may not apply to other grade bands. For instance, in the Plan category, for the sub-objective "Instructs students to consider simplifying assumptions in order to help generate ideas," students in lower grades may not be asked to simplify assumptions in our observed classrooms. Some steps, such as Learn, may be guided by teachers of younger grades, while older students are more capable of self-directing inquiry, organizing information, and developing hypotheses. As older students practice skills used by engineering designers, the process may become longer and more iterative as they handle complex real-world challenges.

Conclusion

This study develops a K-12 classroom observation protocol to assess teachers' implementation of STEM integration. The protocol covers eight categories: *Define*, *Learn*, *Plan*, *Try*, *Test*, *Decide*, *Improve*, and *Instructional Strategies.* A draft protocol was reviewed by STEM education experts and revised accordingly. The protocol was also developed through rating class videos by two raters. The overall IRR in our study (Gwet AC1 coefficient) was 0.73, which was substantial. Each category's IRR was also calculated, and the strengths of the agreement are all above moderate. Future work can revise lower categories based on rater disagreement to improve clarity.

This study shows that it is feasible to develop an integrated STEM teacher classroom observation protocol that is aligned with theory and best practices for teaching science and mathematics through engineering design contexts. The observation protocol developed in this study can be reliably used by multiple raters for identifying teachers' implementation of STEM integration across different class lessons and evaluating teachers' fidelity to implementing a curriculum that is designed from a STEM integration perspective. This study developed the protocol from an engineering design-based perspective where students employ evidence-based reasoning, problem-solving, communication skills, and teamwork as they apply scientific and mathematics principles towards developing real-world solutions.

Future studies may also develop the observation tool from other perspectives of scientific inquiry, curriculum integration, and engineering design pedagogy. Contributions from other disciplines in engineering, science, and mathematics can add to our checklist of actions contributing to effective integrated STEM teaching. Input from teachers on decisions and during STEM integration would also add perspective to our definitions. Integrated STEM education is a challenging instructional approach requiring a great deal of teacher knowledge and preparation, making high-quality observation instruments an important resource for measuring teacher implementation.

Acknowledgments

We are grateful to the experts who participated in the review of the protocol and shared their insights with us about the protocol. Disclosure of potential conflicts of interest: The authors declare that they have no conflict of interest. Ethical Statement: The work described is original and has not been previously published. The manuscript was approved by all authors. Consent Statement: An appropriate consent form was used for the participants.

Author Bios

Shuqi Zhou currently is an assistant professor at Donghua University, China. She received her Ph.D. degree in educational psychology and research methodology at Purdue University. Her research area is applied educational measurement and assessment. Email: shuqizhou11@gmail.com

Hillary Merzdorf is a postdoctoral researcher at Texas A&M University with the Institute for Engineering Education and Innovation. She received her Ph.D. degree in engineering education from Purdue University. Her research interests are in assessment methods for student-technology interactions in engineering education. Email: hmerzdorf@tamu.edu

Tamara Moore is a professor in the School of Engineering Education at Purdue University. Her research interests include: how engineering and engineering thinking promote learning through the paradigm of STEM integration;

support STEM integration pedagogies through curriculum development and teacher professional development. Email: tamara@purdue.edu

Kerrie Douglas is an associate professor in the School of Engineering Education at Purdue University. She conducts studies about the evidence and rationale that inform how evaluation and assessments are used to make decisions and support learning in engineering education. Email: douglask@purdue.edu

References

ABET. (2016). *Criteria for accrediting engineering programs, 2016-2017*. http://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2016-2017/#outcomes

Anderson-Rowland, M. R., Baker, D. R., Banks, D. L., Crouch, P. E., Evans, D. L., Garcia, A., & Yasar, S. (2003). Design, engineering, and technological expansion for K-12 teachers. *In 2003 ASEE Annual Conference and Exposition: Staying in Tune with Engineering Education* (pp. 7211–7227).

Apedoe, X. S., Reynolds, B., Ellefson, M. R., & Schunn, C. D. (2008). Bringing engineering design into high school science classrooms: The heating/cooling unit. *Journal of Science Education and Technology*, *17*(5), 454–465. https://doi.org/10.1007/s10956-008-9114-6

Bagiati, A., & Evangelou, D. (2018). Identifying engineering in a PreK classroom: An observation protocol to support guided project-based instruction. In L. English & T. J. Moore (Eds.), *Early engineering learning* (pp. 83–111). Springer.

Brand, B. R. (2020). Integrating science and engineering practices: Outcomes from collaborative professional development. *International Journal of STEM Education*, *7*(1), 1–13. https://doi.org/10.1186/s40594-020-00210-x

Brooks, G. J., & Brooks, G. M. (1993). *In search of understanding: The case for constructivism classroom*. Association for Supervision and Curriculum Development.

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *Journal of Engineering Education*, *97*(3), 369–387. https://doi.org/10.1002/j.2168-9830.2008.tb00985.x

Bryan, L. A., Moore, T. J., Johnson, C. C., & Roehrig, G. H. (2016). Integrated STEM education. In C. C. JohnsonE. E. Peters-Burton, & T. J. Moore (Eds.), *STEM road map: A framework for integrated STEM education* (pp. 23–37). Routledge.

Capobianco, B. M., & Rupp, M. (2014). STEM teachers' planned and enacted attempts at implementing engineering design-based instruction. *School Science and Mathematics*, *114*(6), 258–270. https://doi.org/10.1111/ssm.12078

Capraro, R. M., & Slough, S. W. (2008). Why PBL? Why STEM? Why now? An introduction to project-based learning: An integrated science, technology, engineering, and mathematics (STEM) approach. In R. M. Capraro & S. W. Slough (Eds.), *Project based learning: An integrated science technology engineering and mathematics (STEM) approach* (pp. 1–6). Sense.

Cejka, E., & Rogers, C. (2005). Inservice teachers and the engineering design process. *Proceedings of the 2005 American Society for Engineering Education Annual Conference & Exposition.* https://ceeo.tufts.edu/ documents/conferences/2005eccr.pdf

Chen, Y. C., & Terada, T. (2021). Development and validation of an observation-based protocol to measure the eight scientific practices of the Next Generation Science Standards in K-12 science classrooms. *Journal of Research in Science Teaching*, *58*(10), 1489–1526. https://doi.org/10.1002/tea.21716

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, *101*(4), 738–797. http://doi.org/10.1002/j.2168-9830.2012.tb01127.x

Dillon, H., James, C., Prestholdt, T., Peterson, V., Salomone, S., & Anctil, E. (2020). Development of a formative peer observation protocol for STEM faculty reflection. *Assessment & Evaluation in Higher Education*, *45*(3), 387–400. https://doi.org/10.1080/02602938.2019.1645091

Douglas, K. A., Moore, T. J., & Merzdorf, H. E. (2018). Informed designers? Students' reflections on their engineering design process. *International Journal of Education in Mathematics, Science and Technology*, *6*(4), 443–459. https://doi.org/10.18404/ijemst.440347

Dyba, T., Kitchenham, B. A., & Jorgensen, M. (2005). Evidence-based software engineering for practitioners. *IEEE Software*, *22*(1), 58–65. https://doi.org/10.1109/MS.2005.6

Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, *14*(2), 1–16. https://doi.org/10.1187/cbe-14-06-0095

Estapa, A. T., & Tank, K. M. (2017). Supporting integrated STEM in the elementary classroom: A professional development approach centered on an engineering design challenge. *International Journal of STEM Education*, *4*(1), 1–16. https://doi.org/10.1186/s40594-017-0058-3

Fortus, D., Dershimer, R. C., Krajcik, J., Marx, R. W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. *Journal of Research in Science Teaching*, *41*(10), 1081–1110. https://doi.org/10.1002/tea.20040

Frykholm, J., & Glasson, G. (2005). Connecting science and mathematics instruction: Pedagogical context knowledge for teachers. *School Science and Mathematics*, *105*(3), 127–141. https://doi.org/10.1111/j.1949-8594.2005.tb18047.x

Gabriel, R. (2017). Rubrics and reflection: A discursive analysis of observation debrief conversations between novice teach for America teachers and mentors. *Action in Teacher Education*, *39*(1), 85–102. https://doi.org/10.1080/01626620.2016.1245636

Gleason, J., Livers, S., & Zelkowski, J. (2017). Mathematics classroom observation protocol for practices (MCOP2): A validation study. *Investigations in Mathematics Learning*, *9*(3), 111–129. https://doi.org/10.1080/19477503.2017.1308697

Guzey, S. S., Caskurlu, S., & Kozan, K. (2020). Integrated STEM pedagogies and student learning. In C. JohnsonM. Mohr-Schroeder, T. Moore, & L. English (Eds.), *Handbook of research on STEM education* (pp. 65–75). Routledge.

Guzey, S. S., Harwell, M., Moreno, M., Peralta, Y., & Moore, T. J. (2017). The impact of design-based STEM integration curricula on student achievement in engineering, science, and mathematics. *Journal of Science Education and Technology*, *26*(2), 207–222. https://doi.org/10.1007/s10956-016-9673-x

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. https://doi.org/10.1348/000711006X126600

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/

Horizons Research Inc. (2000). *Validity and reliability information for the LSC Classroom Observation Protocol*. http://www.horizon-research.com/reports/

Householder, D. L., & Hailey, C. E. (2012). *Incorporating engineering design challenges into STEM courses*. National Center for Engineering and Technology Education. http://ncete.org/flash/pdfs/NCETECaucusReport.pdf

Hurley, M. (2001). Reviewing integrated science and mathematics: The search for evidence and definitions from new perspectives. *School Science and Mathematics*, *101*(5), 259–268. https://doi.org/10.1111/j.1949-8594.2001.tb18028.x

Judson, E. (2013). Development of an instrument to assess and deliberate on the integration of mathematics into student-centered science learning. *School Science and Mathematics*, *113*(2), 56–68. https://doi.org/10.1111/ssm.12004

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1), 1–11. https://doi.org/10.1186/s40594-016-0046-z

Kurt, M., & Benzer, S. (2020). An investigation on the effect of STEM practices on sixth grade students' academic achievement, problem solving skills, and attitudes towards STEM. *Journal of Science Learning*, *3*(2), 79–88. https://doi.org/10.17509/jsl.v3i2.21419

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lie, R., Guzey, S., & Moore, T. J. (2019). Implementing engineering in diverse upper elementary and middle school science classrooms: Student learning and attitudes. *Journal of Science Education and Technology*, *28*(2), 104–117. https://doi.org/10.1007/s10956-018-9751-3

Massachusetts DOE. (2006). *Massachusetts science and technology/engineering curriculum framework*. Massachusetts Department of Education.

Moore, T. J. (2008). STEM integration: Crossing disciplinary borders to promote learning and engagement. *Invited presentation to the faculty and graduate students of the UTeachEngineering, UTeachNatural Sciences, and STEM Education program area at University of Texas at Austin* (pp. 1–13).

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., Smith, K. A., & Stohlmann, M. S. (2014a). A framework for quality K-12 engineering education: Research and development. *Journal of Pre-college Engineering Education Research (J-PEER)*, *4*(1), Article 2. https://doi.org/10.7771/2157-9288.1069

Moore, T. J., Guzey, S. S., & Brown, A. (2014b). Greenhouse design to increase habitable land: An engineering unit. *Science Scope*, *37*(7), 51–57.

Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). STEM integration: A synthesis of conceptual frameworks and definitions. In *Handbook of research on STEM education* (pp. 3–16). Routledge.

Moore, T. J., & Smith, K. A. (2014). Advancing the state of the art of STEM integration. *Journal of STEM Education: Innovations and Research*, *15*(1), 5–10.

Moore, T. J., Stohlmann, M. S., Wang, H. H., Tank, K. M., Glancy, A. W., & Roehrig, G. H. (2014c). Implementation and integration of engineering in K-12 STEM education. In *Engineering in pre-college settings: Synthesizing research, policy, and practices* (pp. 35–60). Purdue University Press.

Morrison, J. S. (2006). Attributes of STEM education: The student, the school, the classroom. *TIES (Teaching Institute for Excellence in STEM)*, *20*, 2–7. Retrieved January 20, 2023, from http://www.leadingpbl.org/f/Jans%20pdf%20Attributes_of_STEM_Education-1.pdf

Mustafa, N., Ismail, Z., Tasir, Z., Said, M., & Haruzuan, M. N. (2016). A meta-analysis on effective strategies for integrated STEM education. *Advanced Science Letters*, *22*(12), 4225–4228. https://doi.org/10.1166/asl.2016.8111

Nadelson, L. S., & Seifert, A. L. (2017). Integrated STEM defined: Contexts, challenges, and the future. *Journal of Educational Research*, *110*(3), 221–223. https://doi.org/10.1080/00220671.2017.1289775

National Academies of Sciences, Engineering, and Medicine. (2019). *Science and engineering for grades 6-12: Investigation and design at the center*. National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Research Council. (2013). *Next generation science standards: For states, by states*. National Academies Press.

National Research Council. (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press.

Riskowski, J. L., Todd, C. D., Wee, B., Dark, M., & Harbor, J. (2009). Exploring the effectiveness of an interdisciplinary water resources engineering module in an eighth grade science course. *International Journal of Engineering Education*, *25*(1), 181–195. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.580.2126&rep=rep1&type=pdf

Roehrig, G. H., Dare, E. A., Ring-Whalen, E., & Wieselmann, J. R. (2021). Understanding coherence and integration in integrated STEM curriculum. *International Journal of STEM Education*, *8*, Article 2. https://doi.org/10.1186/s40594-020-00259-8

Roehrig, G. H., Moore, T. J., Wang, H. H., & Park, M. S. (2012). Is adding the E enough? Investigating the impact of K-12 engineering standards on the implementation of STEM integration. *School Science and Mathematics*, *112*(1), 31–44. https://doi.org/10.1111/j.1949-8594.2011.00112.x

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245–253. https://doi.org/10.1111/j.1949-8594.2002.tb17883.x

Shekhar, P., Demonbrun, M., Borrego, M., Finelli, C., Prince, M., Henderson, C., & Waters, C. (2015). Development of an observation protocol to study undergraduate engineering student resistance to active learning. *International Journal of Engineering Education*, *31*(2), 597–609.

Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, *12*(4), 618–627. https://doi.org/10.1187/cbe.13-08-0154

Stohlmann, M., Moore, T. J., McClelland, J., & Roehrig, G. H. (2011). Impressions of a middle grades STEM integration program: Educators share lessons learned from the implementation of a middle grades STEM curriculum model. *Middle School Journal*, *43*(1), 32–40. https://doi.org/10.1080/00940771.2011.11461791

Tong, F., Tang, S., Irby, B. J., Lara-Alecio, R., Guerrero, C., & Lopez, T. (2019). A process for establishing and maintaining inter-rater reliability for two observation instruments as a fidelity of implementation measure: A large-scale randomized controlled trial perspective. *Studies in Educational Evaluation*, *62*, 18–29. https://doi.org/10.1016/j.stueduc.2019.04.008

Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press. (Original work published 1958).

Van Epps, A. S. (2013). Educating for evidence based decisions in engineering: The view as librarian and instructor. *ASEE Annual Conference and Exposition, Conference Proceedings*.

Walkington, C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L., & Marder, M. (2012). *Development of the UTeach observation protocol: A classroom observation instrument to evaluate mathematics and science teachers from the UTeach preparation program*. Southern Methodist University. https://www.researchgate.net/profile/Candace-Walkington/publication/265478444_Development_of_the_UTeach_Observation_Protocol_A_Classroom_Observation_Instrument_to_Evaluate_Mathematics_and_Science_Teachers_from_the_UTeach_Preparation_Program/links/543952900cf2d6698be16304/Development-of-the-UTeach-Observation-Protocol-A-Classroom-Observation-Instrument-to-Evaluate-Mathematics-and-Science-Teachers-from-the-UTeach-Preparation-Program.pdf

Wang, H. H., Moore, T. J., Roehrig, G. H., & Park, M. S. (2011). STEM integration: Teacher perceptions and practice. *Journal of Pre-College Engineering Education Research*, *1*(2), Article 2. https://doi.org/10.5703/1288284314636

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. https://doi.org/10.1002/tea.21557

Zhou, S. Q., Dong, Z. H., Wang, H. H., & Chiu, M. M. (2023). A meta-analysis of STEM integration and student achievement: an engineering-design approach [Manuscript under review]. *Journal of Science Education and Technology*.

Appendix

Integrated STEM Teacher Classroom Observation Protocol

		Rubric				
	Pedagogical content goals	**Objectives**	**Sub-objectives**			**Notes**
Define	Teacher presents the engineering challenge and introduces/reviews the design process. Teacher facilitates students to understand the scope of the problem by asking questions and interacting with the client. Teacher encourages students to consider what is the fundamental need of stakeholders in the problem.	Teacher explains the design process as a decision-making process based on evidence that results in technology.	Covers key aspects of the design process.	yes	no	
			Asks students to describe the whole process.	yes	no	
			The result of process is technology.	yes	no	
		Teacher asks students questions to prompt their thinking about what the problem is from the perspective of stakeholders.	Instructs students to further discover aspects of the problem.	yes	no	
			Asks questions to students about their understanding of the problems.	yes	no	
			Asks students what each stakeholder needs.	yes	no	
			Asks students why the problem is important to them.	yes	no	
		Teacher asks students to identify criteria/constraints, and trade-offs of the problem.	Asks students to identify the criteria/constraints of the problem.	yes	no	
			Asks students to identify trade-offs of the problem.	yes	no	
		Teacher helps students to tell identify the end-user and the client.	Introduces the end-user.	yes	no	
			Asks students to define the end-user.	yes	no	
			Introduces the client, if there is a client.	yes	no	
			Asks students to define the client, if there is a client.	yes	no	
			Asks students to differentiate between an end-user and a client, if there is a client.	yes	no	
		Teacher challenges students to learn more about the problem.	Questions what content students will need to learn about the problem.	yes	no	
			Questions how students could find/learn information.	yes	no	
			Instructs students to identify what criteria and constraints are important.	yes	no	
			Challenges students' assumptions about the problem (why?).	yes	no	
		Teacher allows students to explain why the problem is important, using information from the client and stakeholders to support the reasons.	Ask students whether the problem is important.	yes	no	
			Instructs students to explain why the problem is important.	yes	no	
			Instructs students to provide rationale from client information.	yes	no	
Learn	Teacher gives students the opportunity to build their knowledge for the design through research for the purpose of developing design solutions. The research informs their design and allows them to further scope the problems. This research can include mathematics and science concepts, and other strategies.	Teacher emphasizes the fundamental science/mathematics concepts related to the engineering design problem while making connections to the engineering design problem.	Discusses relevant science/mathematics concepts.	yes	no	
			Connects scientific or mathematics concepts to design problems.	yes	no	
			Questions students to identify what about the concepts can inform design.	yes	no	
		Teacher gives students opportunities to learn background information to inform their design and understanding of the problem.	Instructs students to do research about the problem, e.g., from written sources, surveys, interviews, etc.	yes	no	
			Asks what existing design, product, or solution fills a similar need.	yes	no	
			Asks students to identify what informational knowledge they need to learn.	yes	no	
			Facilitates students to learn informational knowledge.	yes	no	

Appendix (Continued)

Appendix
(Continued)

		Rubric				
	Pedagogical content goals	Objectives	Sub-objectives			Notes
Plan	During idea generation, teacher supports students to continue to define the problem as they generate a wide variety of solution paths and ideas. Teacher does not judge student ideas at this point and supports students to openly consider many ideas.	Teacher expects students to generate a variety of potential solution ideas.	No judgements of potential ideas.	yes	no	
			Encourages a variety of solution approaches.	yes	no	
			Instructs students to consider simplifying assumptions in order to help generate ideas. (This does not fit for K-2.)	yes	no	
		Teacher encourages students to continue to define the problem as they generate ideas.	Recalls or points students to look back at criteria and constraints.	yes	no	
			Challenges students to think about the problem differently.	yes	no	
	During idea selection, teacher instructs students to select potential solutions from multiple design ideas based on evidence learned about the problem. Teacher instructs students to communicate the details of their design ideas, the decisions that students made, and how they justify their decisions with evidence.	Teacher instructs students to select a design from multiple solutions based on their evidence.	Instructs students to compare multiple solutions.	yes	no	
			Instructs students to think about the trade-offs of each of their design plans.	yes	no	
			Instructs students to make decisions using mathematics concepts as evidence.	yes	no	
			Instructs students to make decisions using science concepts as evidence.	yes	no	
			Provides students the opportunity to state which solution was chosen.	yes	no	
			Instructs students to consider how well their design meets clients' needs.	yes	no	
			Accepts evidence that might also come from other sources/concepts beyond mathematics or science.	yes	no	
		Teacher instructs students to communicate how they make their design decisions based on the evidence.	Provides students the opportunity to communicate why design choices were made.	yes	no	
			Provides students the opportunity to communicate how parts of the design work.	yes	no	
			Provides students the opportunity to communicate how the design meets the criteria and constraints.	yes	no	
			Has students communicate how mathematics and science concepts inform their design.	yes	no	
			Has students communicate what design decisions they make and why.	yes	no	
		Teacher instructs students to work out details of their design ideas for the design they select.	Instructs students to represent their design idea.	yes	no	
			Instructs students to label important parts of the design.	yes	no	
			Instructs students to describe how their design works.	yes	no	
Try	Teacher facilitates students to put the plan into action by building a testable representation of a solution (e.g., building prototype, process). Teacher instructs students to make modifications of their prototype in order to better meet criteria/constraints if needed.	Teacher instructs students to build a prototype, model, or product for the design they selected from plan.	Provides student materials that they needed.	yes	no	
			Has students keep in mind criteria/constraints when they build their design.	yes	no	
			Has students keep in mind clients' need when they build their design.	yes	no	
			Has students keep in mind science/mathematics concepts from Learn stage.	yes	no	
		Teacher instructs students to optimize their prototype when they develop the design.	Instructs students to consider weaknesses of a prototype.	yes	no	
			Instructs students to modify their prototype to follow their plan if needed.	yes	no	
			Requires students to consider the trade-offs.	yes	no	
Test	Teacher instructs students to consider testable questions about the performance of their design prototype, model or product. Teacher has students collect specific results from the test and instructs students to analyze their collected results for evaluating solution performance.	Teacher has students follow up test procedures to answer testable questions of problems to collect data.	Instructs students to consider testable questions of problems during testing.	yes	no	
			Asks students to make predictions about their test results.	yes	no	
			Instructs students to consider the criteria and constraints of the problem during testing.	yes	no	
			Gives students opportunities to carry out the experiments.	yes	no	
		Teacher requires students to analyze and interpret collected data.	Asks students to record the experiment data.	yes	no	
			Asks students to analyze the test data.	yes	no	
			Requires students to explain the advantages of design found from data analysis.	yes	no	
			Requires students to explain drawbacks of design found from data analysis.	yes	no	

Appendix (Continued)

Appendix
(Continued)

	Pedagogical content goals	Rubric				Notes
		Objectives	Sub-objectives			
Decide	Teacher asks students to decide the overall performance of their design. Teacher instructs students to communicate and justify whether their design decision is appropriate based on evidence.	Teacher gives students opportunities to communicate and justify their design.	Provides students the opportunity to communicate their design as a synthesis.	yes	no	
			Encourages to include the evidence from the testing result as a justification.	yes	no	
			Teacher encourages to justify based on a correct understanding of mathematics/science concepts.	yes	no	
			Encourages students to discuss what was successful about their design.	yes	no	
		Teacher requires students to use evidence to evaluate their design.	Instructs students to apply results (e.g., data) gathered from testing to evaluate their design.	yes	no	
			Instructs students to apply learned core mathematics/science concepts to evaluate their design.	yes	no	
			Asks students to consider whether their design stays within criteria/constraints.	yes	no	
			Asks students to consider how their design balances trade-offs.	yes	no	
Improve	Teacher gives students the opportunity to revisit the solution, and make changes and improvements based on an understanding of the problem and results. Teacher instructs students to improve the performance of chosen solution (i.e., redesign) and provides students with opportunities to learn from failure.	Teacher asks students to decide whether their design needs improvement and communicates how their design could be improved if needed.	Requires students to use test results as evidence for improvement or no improvement.	yes	no	
			Encourages students to discuss how their design could be improved based on their test results and feedback from the client/user.	yes	no	
			Requires students to explain the rationale for improvements based on test results.	yes	no	
			Requires students to explain the rationale for improvements based on a correct understanding of mathematics/science concepts.	yes	no	
			Requires students to list planned improvements.	yes	no	
		Teacher asks students to redesign their solution based on the performance of the prototype.	Instructs students to redefine the problem.	yes	no	
			Instructs students to replan.	yes	no	
			Instructs students to retry.	yes	no	
			Instructs students to retest.	yes	no	
		Teacher provides students with opportunities to learn from failure.	Facilitates student conversations about failure for purpose of improvement.	yes	no	
			Teacher gives students feedback on planned improvements, i.e., redesign, but without directly telling what students can do.	yes	no	

Appendix (Continued)

Appendix
(Continued)

		Rubric				
	Pedagogical content goals	**Objectives**	**Sub-objectives**			**Notes**
Instructional Strategies	Teacher gives students the opportunity to construct their learning and promotes teamwork. Teacher is the facilitator of students' learning, and teacher emphasizes dialogues with students and teacher responds in a way that meets their learning needs.	Teacher promotes collaborative and cooperative learning in teamwork.	Instructs students to identify their roles during teamwork.	yes	no	
			Encourages students to contribute to group work.			
			Encourage students to help team members confront problems.	yes	no	
			Instructs students to have discussions with each other to solve problems.	yes	no	
			Helps students to consider different viewpoints from each teammate.	yes	no	
			Promotes students to give peer feedback.			
		Teacher uses scaffolding instruction, which allows students to be responsible for their own learning.	Communicates students' responsibility during the instruction.	yes	no	
			Fades his/her own support during the instruction.	yes	no	
		Teacher gives students constructive feedback.	Provides specific feedback as necessary.	yes	no	
			Provides forward thinking feedback.	yes	no	
			The feedback teacher provides is worded as a question sometimes.	yes	no	
			Gives frequent feedback to students.	yes	no	
	Teacher promotes a student-centered approach. Teacher encourages students to look at problems in multiple ways and probes students' conceptual understanding.	Teacher gives students a voice in the class, seeks students' contribution, and adapts instruction according to students' needs.	Respects students' ideas and invites students to share ideas.	yes	no	
			Actively solicits student ideas and builds on these conceptions and ideas during instruction.	yes	no	
			Demonstrates consideration for the pacing of a lesson based on students' progression from whole unit perspectives.	yes	no	
		Teacher uses multiple forms of representations to help students to learn concepts.	Uses at least two forms of representations to help students explain the concept.	yes	no	
			Translates between representations to help students explain the concept.	yes	no	
			Encourages students to use multiple forms of representations.	yes	no	
			Encourages students to translate between representations to explain the concept.	yes	no	
		Teacher promotes the development and application of models for developing deep thinking skills.	Instructs students to develop models for representing the problems.	yes	no	
			Instructs students to apply models to guide their design.	yes	no	