

RESEARCH

Open Access



Investigating the presence of mathematics and the levels of cognitively demanding mathematical tasks in integrated STEM units

Elizabeth N. Forde^{1*} , Latanya Robinson², Joshua A. Ellis² and Emily A. Dare²

Abstract

Effective K-12 integrated STEM education should reflect an intentional effort to adequately represent and facilitate each of its component disciplines in a meaningful way. However, most research in this space has been conducted within the context of science classrooms, ignoring mathematics. Also missing from the literature is research that examines the level of cognitive demand required from mathematical tasks present within integrated STEM lessons. In order to seek insight pertaining to this gap in the literature, we sought to better understand how science teachers use mathematics within K-12 integrated STEM instruction. We used an explanatory sequential mixed methods research design to explore the enactment of mathematics in integrated STEM lessons that focus on physical, earth, and life science content. We first examined 2030 sets of video-recorded classroom observation scores generated from the 10-item STEM Observation Protocol (STEM-OP) designed for observing integrated STEM education in K-12 classrooms. We compared the STEM-OP scores of classroom observations that included mathematics with those that did not. This quantitative analysis was followed by a closer, more in-depth qualitative examination of how mathematics was employed, focusing on the degree of cognitive demand. To do this, we coded and analyzed transcripts from video-recorded classroom observations in which mathematical content was present. Our study yielded two main findings about mathematics in integrated STEM lessons: (1) the presence of mathematical content resulted in higher STEM-OP scores on nearly all items, and (2) mathematical tasks within these lessons were categorized as requiring mainly low levels of cognitive demand from students. This study highlights the need for the increased inclusion of mathematical tasks in integrated STEM teaching. Implications for including higher-order mathematical thinking within integrated STEM teaching are discussed.

Keywords Integrated STEM, Mathematical tasks, Cognitive demand

Introduction

With an increased desire to draw individuals into the science, technology, engineering, and mathematics (STEM) fields to meet current workforce demands and compete in the global economy (Wang et al., 2011), many countries

have heightened their focus on STEM education within recent years. In the United States, there is a projected increase in STEM occupations of 10.5% as compared to non-STEM occupations of 7.5% for the period 2020 to 2030 (U.S. Bureau of Labor Statistics, 2021). Wang et al. (2011) noted that the problems being faced in our changing global society are multidisciplinary and require the integration of STEM-related knowledge and skills to create feasible solutions. Although an explicit, clear definition for integrated STEM education has not been agreed upon (Angier, 2010; Dare et al., 2019; Bybee, 2013; Moore et al., 2020), many scholars share an increased interest

*Correspondence:
Elizabeth N. Forde
fordee@newpaltz.edu

¹ State University of New York (SUNY), New Paltz, USA

² Department of Teaching and Learning, Florida International University, Miami, USA

in its importance and inclusion in education, specifically within K-12 spaces.

A range of STEM competencies is developed through integrated STEM education according to the varying requirements posed by a particular situation or a problem (McLoughlin et al., 2020). Bybee (2013) suggested that these competencies include a combination of conceptual understanding as well as procedural skills and abilities for individuals to address STEM-related social and global issues. He further explained that these concepts and skills embody the integration of STEM disciplines. Hence, a quality integrated STEM education experience ought to reflect an approach that seeks to effectively integrate the four disciplines of science, technology, engineering, and mathematics (Johnson et al., 2020). A joint statement from The National Council of Supervisors of Mathematics (NCSM) and the National Council of Teachers of Mathematics (NCTM) (NCSM & NCTM, 2018) purports that any integrated STEM activity that aims to address the effective incorporation of mathematics should do so with integrity and target the grade level's mathematics content and appropriate mathematical practices. Currently, though, the majority of research related to integrated STEM education has taken place within the context of science classrooms (e.g., English, 2016), which offer a space for reasoning abstractly and quantitatively, as described in *A Framework for K-12 Science Education* (National Research Council, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013). It is clear that additional research is needed to determine the ways in which learning across disciplines can support integration so that one discipline does not dominate the others (English, 2016).

Although Fitzallen (2015) and Maass et al. (2019) examined the role of mathematics in integrated STEM education, other scholars have noted an under-representation of mathematics activities and practices used in STEM education (English, 2016; Marginson et al., 2013; Shaughnessy, 2013). English (2016) and Stohlmann (2018) noted that despite the inclusion of mathematical content and concepts within integrated STEM curricula, the significance and level of mathematical thinking required from students remain unclear. This is problematic as Maass et al. (2019) acknowledged that although mathematics undoubtedly supports the other three STEM disciplines, its understated and underrepresented role in integrated STEM education cannot go unnoticed. For instance, Ring et al. (2017) and Ring-Whalen et al. (2018) found that science teachers' conceptualizations of integrated STEM often relegated mathematics (along with technology) as a tool to be used in support of science and engineering learning. Consequently, elevating the role of mathematics

within the broader context of integrated STEM teaching and learning is needed. In order to make that happen, we first must understand how exactly mathematics is being utilized, especially with respect to the level of mathematical thinking and cognitive demand.

This dearth of literature related to mathematics within integrated STEM teaching has not gone unnoticed. English (2016) noted that of the four STEM disciplines, student outcomes in mathematics were the lowest when compared to the other three disciplines. Hence, English (2016) suggested that this anomaly was worth further scrutiny. Additionally, Stohlmann (2018) called for a focus on how the content of mathematics included within integrated STEM ties to the other STEM content to make the learning of mathematics more explicit. In response to these calls for research on mathematics in integrated STEM education, the current study sought to address the following research questions: (1) *How do integrated STEM lessons that include mathematics perform on an integrated STEM observational protocol compared to lessons without mathematics?* and (2) *What levels of mathematical cognitive demand are represented in physical, earth, and life science integrated STEM units?*

Literature review

Mathematics and STEM integration

In a call for greater emphasis on STEM disciplinary integration, English (2016) reiterated that mathematics, along with engineering are notably underrepresented in studies of STEM education and subsequently called for "lifting the profile of mathematics in STEM integration" (p. 4). Although research has made some initial progress along these lines and attempts to understand mathematics inclusion within integrated STEM educational contexts are present in the literature, it is primarily based on few empirical studies (e.g., Becker & Park, 2011; Hurley, 2001) that investigated how different integrated approaches can support mathematics learning. The different approaches can include coordination across disciplines or complementary overlapping across disciplines (Bybee, 2013), and the distinct characteristics of the specific approach make mathematics achievement challenging (National Academy of Engineering and National Research Council [NAE and NRC], 2014). Becker and Park (2011) meta-analysis focused on the effects of integrative approaches among STEM subjects while considering eight different combinations of STEM disciplines (e.g., science and mathematics; engineering and technology; science, technology, and mathematics). Mathematics achievement showed the smallest effect size when paired with another discipline (Becker & Park, 2011). Similarly, Hurley's (2001) meta-analysis of the effects of five different integrated teaching approaches for mathematics

and science (i.e., sequenced, parallel, partial, enhanced, and total) found similar results concerning mathematics achievement. However, Becker & Park's (2011) work also revealed differences in effect sizes of mathematics achievement when comparing across the types of integration. For example, in the sequence integration approach, the effect size of mathematics achievement was large when compared to that of the parallel and total integration approaches. These findings suggest that when mathematics is included in integrated STEM that attention needs to be paid to *how* mathematics is paired with other disciplines and the integrated teaching approach(es) to support mathematics achievement.

Whereas these two studies focused on discipline combination and integrated approaches, there is research that looks more specifically at how mathematical content and practices can be incorporated into discipline integration. For instance, Baldinger et al.'s (2020) literature review of 32 published studies from 2013 to 2018 focused on mathematical topics, proficiencies, and practices to determine how mathematics is integrated with other disciplines. Baldinger et al. (2020) then noted that currently within discipline integration practices, mathematics serves as a supporting role for science, technology, or engineering learning goals, while the associated conceptual mathematics learning goals are essentially overlooked. Furthermore, Stohlmann's (2018) analysis of 21 studies from 2008 to 2018 examined the outcome of mathematics learning in discipline integration by considering content integration. Stohlmann (2018) suggested the fact that mathematics is not emphasized in integrated STEM may partially be a result of the "perception that mathematics achievement is difficult to promote through STEM integration" (p. 317). In terms of content integration, Moore, Stohlmann, et al., (2014) stated that integrated STEM lessons with an emphasis on content integration with learning objectives from mathematics and another discipline should focus on the central idea that connects the disciplines. Additionally, for effective development of mathematics concepts and skills, Stohlmann (2018) reiterated that it is essential to attend to the natural connections between mathematics and other STEM disciplines in integrated lessons.

When addressing content integration as proposed above, the NAE and NRC (2014) noted the challenges of leveraging similarities between overlapping content presented in the different disciplines' standards to develop discipline-specific knowledge. One challenge is that the shared content is presented in different ways for the subjects in their respective fields. For example, according to the Standards for Technological and Engineering Literacy (2020), geometry is necessary within science, engineering, and technology. The mathematics topic of geometry

involves determining congruence and similarity in mathematics using physical models, transparencies, or geometry software. Particularly, this mathematics topic ties engineering and technology to the design solutions and supports the concept of developing models in science (NAE and NRC, 2014). The challenge lies in how to effectively convey the underpinning mathematical ideas of the content when the disciplines are combined. Notably, in mathematics instruction, the focus is on the development of concepts whereas when it is integrated with science, engineering, and technology, mathematics is used to support the development of the concepts within these respective disciplines. Hence, this requires a close examination of the nature of integration with a focus on the connected concepts within disciplines.

Integrated STEM and higher order thinking

Integrated STEM education ties together practices of scientific inquiry and mathematical analysis, which aligns with the interdisciplinary format of STEM standards in science and mathematics (Bybee, 2013; Sanders, 2012). Today's K-12 educational systems have made efforts to reform STEM discipline standards making provision for students to think critically and experience meaningful integration of the STEM disciplines within the context of authentic, real-world challenges (Council of Chief State School Officers & National Governors Association Center for Best Practices, 2010; NGSS Lead States, 2013). The content standards in mathematics (e.g., Common Core) and science (e.g., NGSS) now emphasize developing students' abilities to think deeply and understand relationships about their respective disciplinary concepts and practices (Achieve, 2010; NAE and NRC, 2014). Students should engage in disciplinary tasks that require interpretation and construction of meaning to arrive at answers or solutions that are not obvious at the onset of assigned tasks (Tekkumru-Kisa et al., 2020). These tasks can be in the form of an instructional unit of interdisciplinary work or a classroom activity that is assigned to students by the teacher and directly requires students to intellectually engage in science and/or mathematics thinking (Tekkumru-Kisa et al., 2015). For instance, a task that requires students to memorize or reproduce procedures is considered to be a low-level task, whereas a task that requires students to construct arguments or use evidence-based reasoning to support ideas is a high-level task. Students should spend a significant amount of time working on tasks that are considered high-level because such tasks improve students' abilities to learn content at a deeper level and to think more critically (Tekkumru-Kisa et al., 2015).

Categorizing activities that require students to complete high-level or low-level tasks is not new. This method

has been used in education since the 1980s. For example, Doyle's (1983) work on the mental processes students engage in to complete academic tasks is often credited with the start of classifying tasks as low-level or high-level; this type of categorization has been used in both mathematics and science education communities. Further, Smith and Stein's (1998) task-analysis guide for the evaluation of Characteristics of Mathematical Instructional Tasks has been used by NCTM and the Stanford Center for Professional Development in mathematics professional development workshops that focus on the quality of mathematical tasks. This guide can be found in mathematics methods course textbooks. Notably, mathematics education journals report its use in supporting the development of mathematical reasoning and problem-solving skills (Boston & Smith, 2011; Dempsey & O'Shea, 2020). The lowest level demand ("memorization") involves reproducing previously learned rules or formulas, for example, requiring students to recall the formula for two- and three-dimensional shapes. The highest-level demand ("doing mathematics") requires non-algorithmic thinking and an understanding of mathematical relationships between concepts, for example, presenting students with opportunities to apply more than one solution strategy and producing explanations and justifications for these. The intention of this guide, though, was not on observing instruction, but on classifying mathematical tasks as "good" (Smith & Stein, 1998, p. 344). Building on Smith and Stein's (1998) work, Matsumura et al. (2006) developed a series of protocols collectively referred to as the Instructional Quality Assessment (IQA), which were designed to assess observed classroom instruction and the quality of work in mathematics (and comprehension) that teachers assign to students. One of these protocols is the Academic Rigor - Implementation of the Task rubric (Rubric 2), which rates the quality and cognitive demand of tasks that students are engaged in during instruction using five levels (Levels 0–4). The rubric mirrors the design of the original Smith and Stein (1998) task-analysis guide in Levels 1–4 but includes a new level, Level 0, in which the task does not require mathematical knowledge on the part of students, only the teacher.

There are parallels among levels in these instruments. For example, the *memorization* level is common in the works of Smith and Stein (1998) and Matsumura et al. (2006). It is also present in Tekkumru-Kisa et al.'s (2015) Task Analysis Guide in Science, a similar guide used in the science education community. In the Tekkumru-Kisa et al.'s (2015) guide, the lowest level is *memorization*, and the highest level is *doing science*. These guides and rubrics are useful to consider because the aims of integrated STEM education require students to engage in higher-order thinking as they consider real-world, open-ended

problems (Kelley & Knowles, 2016). The openness of an authentic, real-world problem allows students to use complex, higher-order thinking in which they analyze data and determine patterns within information to make informed decisions while simultaneously drawing upon knowledge from the disciplines of mathematics and/or science knowledge.

As students are presented with opportunities to draw on knowledge from multiple STEM disciplines, the types of activities they are expected to engage in will look different across classrooms that incorporate integrated STEM. Because K-12 content standards in mathematics and science require students to reason and engage in critical thinking both mathematically *and* scientifically (Council of Chief State School Officers & National Governors Association Center for Best Practices, 2010; NGSS Lead States, 2013), one might expect that most activities would be at a cognitively high-level. Classroom activities within integrated STEM lessons that include mathematics or science should inherently allow students to: (a) construct meaning of the content, (b) make sense of the underlying disciplinary idea, and (c) engage in complex thinking (Moore, Stohlmann, et al., 2014). The alignment between the levels of cognitive demand in the science and mathematics rubrics noted above makes it possible to analyze classroom activities across different settings. This is critical when the objective is to understand the degree to which mathematics or science cognitive demand tasks are present within integrated STEM classroom activities. Moreover, this is fundamentally important considering that a key feature of integrated STEM education is crossing disciplinary boundaries by presenting tasks in a real-world context requiring students to think critically and effectively use knowledge from several disciplines (NAE and NRC, 2014).

Conceptual Framework

Currently, scholars use no standard definition for the term *integrated STEM education*, which has been reported in numerous editorial comments and reviews of the literature (e.g., English, 2016; Li et al., 2020; Martín-Páez et al., 2019; Moore et al., 2020). For example, Martín-Páez et al. (2019) in their review of the literature shared STEM education as "a teaching approach that integrates content and skills specific to science, technology, engineering, and mathematics" (p. 815). Li et al. (2020) also suggested an understanding of STEM education that reaches beyond the simple integration of the disciplines' content but rather one that is seen as "a broad and inclusive perspective to include education in the individual disciplines of STEM, i.e., science education, technology education, engineering education, and mathematics education" (p. 2). This interdisciplinary

perspective was also reflected in Vasquez et al.'s (2013) definition of integrated STEM education as an approach that spans the four disciplines of science, technology, engineering, and mathematics and integrates them to provide relevant and rigorous learning experiences for a diverse range of learners. Additionally, Johnson et al. (2020) offer another definition that suggests using disciplinary knowledge and practices from engineering and technology to teach and learn specific science and/or mathematics knowledge. Likewise, Moore, Stohlmann, et al.'s (2014) definition proposes the use of engineering design to develop an understanding of technologies that require the application of mathematics and/or science content. Notably, a commonality across these different definitions of integrated STEM education is the combination of disciplines that is far reaching and promotes the learning of content and skills from different disciplines.

For the purpose of this study, we broadly define and conceptualize integrated STEM education as an approach that focuses on the interconnectedness between the content and skills of the STEM disciplines; science, technology, engineering, and mathematics. In this study, we focus on the inclusion of mathematics to support student learning and how it is positioned with respect to cognitive demand within integrated STEM activities. This aligns best with the definition provided by Kelley and Knowles (2016) who suggested that STEM education is, "The approach to teaching the STEM content of two or more STEM domains, bound by STEM practices within an authentic context for the purpose of connecting these subjects to enhance student learning" (p. 3). The purpose of intentionally combining content from these disciplines is to ultimately develop concepts and skills from the disciplines and improve student learning outcomes in a meaningful, interconnected way as opposed to a compartmentalized, siloed way that has historically been practiced in K-12 education.

Methods and findings

The aim of this study was to better understand the use of mathematics within integrated STEM contexts. To reach this aim, we first explored how integrated STEM lessons that include mathematics perform on an integrated STEM observational protocol compared to lessons without mathematics. We then investigated the levels of mathematical cognitive demand present in physical, earth, and life science integrated STEM units. The following sections first present the research design and context for the study. This is followed by the quantitative and

qualitative phases along with the respective findings for each phase.

Research design

This study used an explanatory sequential mixed methods research design in which the quantitative data were collected and analyzed prior to conducting qualitative analysis (Teddlie & Tashakkori, 2009). This design allowed us to first investigate how integrated STEM lessons with and without mathematics performed on an integrated STEM observational protocol. For the second phase of the research, we qualitatively analyzed video-recorded classroom observations to identify the level of mathematical cognitive demand represented in a sample of physical, earth, and life science integrated STEM units. The qualitative analysis provided more detail about the presence and cognitive demand of mathematics within different science domains. The quantitative and qualitative results were later synthesized to understand our findings at a more detailed level.

Context

The raw data used in this study (in the form of video-recorded integrated STEM observations) were collected as part of a previously funded 5-year project. During the first 3 years of the project, K-12 science teachers participated in professional development related to integrated STEM education, co-created integrated STEM curriculum units, and implemented these units into their classrooms. In the fourth and fifth years of the project, teachers participated in a similar professional development but field-tested several of the previously created curriculum units instead of writing new ones. The professional development utilized a design-based vision of integrated STEM based on two frameworks that centralized learning activities within an engineering design challenge (Moore, Glancy, et al., 2014, Moore, Stohlmann, et al., 2014). In the professional development, an emphasis was placed on the inclusion of data analysis as one method of engaging in evidence-based reasoning. Participating teachers were video recorded during the implementation of their integrated STEM units, with each video corresponding to one class period (~45 minutes). These video-recorded classroom observations represent a variety of classroom settings, including different grade levels, teachers, student demographics, science content, and engineering design challenges. Specifically, the data used in this study include 2030 video-recorded observations from 106 unique teachers' classrooms from five school districts that include urban, inner-ring suburban, and outer-ring suburban K-12 settings in the Midwestern United States. Most of the observations focus on grades

4–8, although early elementary (K-3) and high school are represented to a lesser extent. The science content covered in the 48 unique curriculum units from the first 3 years of the project spans several topics in the physical (e.g., force and motion), life (e.g., ecosystems), and earth sciences (e.g., plate tectonics); a total of 13 of those curriculum units were field-tested in the fourth and fifth years. Our data set includes 999 physical science, 434 earth science, and 597 life science observations. These video observations also represent 885 elementary (K-5), 1071 middle school (6–8), and 74 high school (9–12) classrooms.

Quantitative phase

Data collection

In the quantitative phase, we compared integrated STEM video observations that included mathematics to those that did not. To do this, we used the STEM Observation Protocol (STEM-OP) (Dare et al., 2021) to score the 2030 classroom observation videos described above. This observational protocol (Dare et al., 2021) measures the degree to which integrated STEM takes place in K-12 science and engineering classrooms; the instrument does not attend to pedagogical quality as other instruments that attend to this are already available. All items on the instrument have demonstrated acceptable Krippendorff's alpha levels ($\alpha > .6$) for interrater reliability with the exception of Item 5 ($\alpha = .58$), which approached our selected threshold. Further, we have also established the structure and reliability of the instrument through principal component analysis (PCA) (Roehrig et al., 2023). The PCA work revealed two core dimensions of integrated STEM education when using our instrument: 1) real-world problem solving and 2) the nature of integrated STEM. The STEM-OP includes 10 items with four descriptive levels for each item (scored 0–3): 1) relating content to students' lives, 2) contextualizing student

learning, 3) developing multiple solutions, 4) cognitive engagement in STEM, 5) integrating STEM content, 6) student agency, 7) student collaboration, 8) evidence-based reasoning, 9) technology practices in STEM, and 10) STEM career awareness. After completing rigorous training with the STEM-OP and establishing interrater reliability, our coding team scored all of the video recordings made available from the previously described project. During this process, the coding team noted whether a given observation included science, technology, engineering, and/or mathematics content. We then used these indicators to subdivide the data into two categories: observations with mathematics ($n = 637$) and observations without mathematics ($n = 1393$).

Data analysis

To determine differences in STEM-OP scores between observations that included or did not include mathematics, we used the Mann-Whitney-Wilcoxon test (Mann & Whitney, 1947; Wilcoxon, 1945). This nonparametric test is used to compare the outcome between two independent groups from the same or identical sample (Hahs-Vaughn & Lomax, 2020). This test does not require that the difference between the samples is normally distributed or that the variances of the two populations are equal (Hahs-Vaughn & Lomax, 2020). An additional advantage of the Mann-Whitney-Wilcoxon test is that the two samples can have an unequal number of observations, which was true for our data. This nonparametric test helped answer the first research question: *How do integrated STEM lessons that include mathematics perform on an integrated STEM observational protocol compared to lessons without mathematics?*

Quantitative findings

Findings from the Mann-Whitney-Wilcoxon tests reveal that there are statistically significant differences in mean

Table 1 Mann-Whitney-Wilcoxon test comparing lessons that include mathematics and lessons that do not include mathematics

STEM-OP Item	Mean _{Math}	Mean _{NoMath}	Med _{Math}	Med _{NoMath}	Diff in Mean	p
1: Relating content to students' lives	0.46	0.59	0	0	-0.13	.003
2: Contextualizing student learning	1.90	1.87	2	2	0.03	.398
3: Developing multiple solutions	0.91	0.79	1	0	0.12	.006
4: Cognitive engagement in STEM	2.22	1.83	2	2	0.39	<.001
5: Integrating STEM content	1.31	0.83	1	1	0.48	<.001
6: Student agency	1.32	1.09	1	1	0.23	<.001
7: Student collaboration	1.82	1.65	1	1	0.17	<.001
8: Evidence-based reasoning	1.01	0.80	1	0	0.21	<.001
9: Technology practices in STEM	0.49	0.13	0	0	0.36	<.001
10: STEM career awareness	0.41	0.54	0	0	-0.13	.006

rank order for most of the STEM-OP's 10 items when comparing lessons that included mathematics ($n=637$) to those that did not ($n=1393$) (Table 1). These results suggest that the presence of mathematics in an observed integrated STEM lesson leads to higher scores on Items 3–9. There is also a negative effect for Items 1 and 10 and no statistically significant effect for Item 2.

The positive effect of the presence of mathematics for Items 3–9 suggests multiple ways in which mathematics may improve integrated STEM instruction. For instance, Item 3 measures to what extent students are provided with opportunities to develop multiple design solutions to an engineering challenge. These design challenges oftentimes require students to use, for example, geometrical concepts in designing their multiple solutions, the incorporation of these mathematical concepts in such engineering aspects of the lessons appeared to have a positive effect on integrated STEM instruction.

Item 4 on the STEM-OP evaluates the level of cognitive engagement in STEM disciplines within the lessons. Mathematics inclusion in these integrated lessons suggests that students were engaged in higher levels of cognitive thinking, for example, in addition to doing calculations related to budgets and cost of production, they used the results of these calculations to make decisions about their design challenges. The positive effect of the presence of mathematics in integrated STEM lessons was also observed in our findings for Item 5. This item measures how teachers integrate content from multiple STEM disciplines, inclusive of mathematics. It should be noted that our comparison used mathematics to sort our scores; however, even within non-mathematics observations, multiple STEM disciplines could still have been integrated (e.g., an observation with science and engineering). The importance in this finding is that including mathematics appears to make a significant difference in the integration of content. Notably, Items 4 and 5, which both examine how STEM content is presented in the observed lessons, displayed the highest mean differences with 0.39 and 0.48, respectively.

Similarly, we observe differences between the two groups for Item 6, which assesses the degree to which students have agency over their learning. It would appear that when mathematics is included that there is more student agency. Somewhat unexpected, we see differences for Item 7, which focuses on student collaboration. Including mathematics appears to result in more complex collaborative activities compared to integrated STEM activities that do not include mathematics. Item 8, which measures evidence-based reasoning, also scored higher when mathematics is present, which may reflect that students often use mathematical evidence in their scientific claims or design decisions. The presence of mathematics

within integrated STEM lessons is also supportive in creating opportunities for students to engage in the appropriate use of technology in calculating, collecting, and/or analyzing data as they work on creating possible solutions for the design challenge; these technology-based opportunities are captured by Item 9. Thus, the results from these remaining items suggest that the inclusion of mathematics is correlated to higher scores in the development of multiple solutions (Item 3), cognitive engagement in STEM (Item 4), integrating STEM content (Item 5), student agency (Item 6), student collaboration (Item 7), evidence-based reasoning (Item 8), and technology practices in STEM (Item 9). For Item 1 (relating content to students' lives) and Item 10 (STEM career awareness), not included in this list, it is curious to note the negative effect of mathematics on both items. This may relate to the fact that these lessons were implemented by science teachers in which their knowledge of relating mathematical content to students' lives and careers may have been limited.

In conclusion, these quantitative results provide evidence that the inclusion of mathematical content within integrated STEM lessons is associated with overall higher scores on Items 3–9 of the STEM-OP. Considering that the STEM-OP was intentionally designed to measure the degree to which integrated STEM occurs, this is evidence that the inclusion of mathematics within integrated STEM lessons is notable.

Qualitative phase

Data collection and materials

Since the results of our initial quantitative phase revealed that observed lessons that included mathematics scored higher on the STEM-OP for most items, we sought to conduct a further investigation of the mathematical tasks occurring in these lessons. This was achieved by exploring the levels of mathematical cognitive demand required from the mathematical tasks within selected physical, earth, and life science integrated STEM units. Thus, we address our second research question: *What levels of mathematical cognitive demand are represented in physical, earth, and life science integrated STEM units?*

The following sections describe the qualitative methods used to investigate the levels of cognitive demand for the mathematical tasks in selected curriculum units. We used a multiple case study design that focused on developing an in-depth understanding of a case or bounded system focused on understanding how events occur and which ones may influence particular outcomes (Savin-Baden & Howell Major, 2012). In order to define the cases for our study, we took several steps. We first considered curriculum units wherein 50% or more of the daily observed lessons

included mathematics; this left us with 20 curriculum units, which cut across all three science disciplines (11 physical science, seven earth science, and two life science) implemented by multiple teachers due to the team-like nature of the original project. We then considered which grade level to examine for the study and which curriculum unit within each science domain to be selected. To make a decision on the grade level, we closely reviewed the Common Core State Mathematics standards which were contained in each curricular document that the research team had access to. The use of this secondary data source assisted us in narrowing in on the elementary level as we observed that these units covered a wider variety of mathematics domains including measurement and data, number and operations, and geometry. As a result, we selected units implemented in the elementary grades, of which there were nine. To decide which specific curriculum unit would best serve as the case for each science discipline, we again referred to the secondary data, this time simultaneously reviewing the stated mathematics standards and topics contained in each curricular document.

Based on this process, we selected a curriculum unit each for physical science, earth science, and life science. There were 53 video observations in total, which included 19 for the physical science unit, 17 for the earth science unit, and 17 for the life science unit. Both the physical science and earth science units were implemented by four teachers, while the life science unit was implemented by five teachers. All curriculum units were implemented at either the fourth or fifth-grade level and were centered on three different engineering design challenges. In addition to the mathematical content/topics, the written curriculum units also contained science content/topics, clear explanations of the intended engineering design challenge that students were expected to be engaged in, as well as technology and engineering connections. Table 2 presents the science discipline, a brief description of the

engineering design challenge, and science and mathematics topics for each curriculum unit.

Data analysis

Overall, a total of 53 classroom-recorded observations that included mathematics were considered and analyzed for the qualitative component of our study. In the first step of our analysis, we reviewed all 53 video observations, noting specific segments in each video for which mathematical tasks were either directly identified by the teacher and/or performed by the students. Through this process, we identified 153 unique segments as multiple segments were possible in the individual video observation. These segments were transcribed in preparation for coding of the mathematical tasks present.

Second, we used the Academic Rigor - Implementation of the Task rubric (Rubric 2) (Matsumura et al., 2006) to code the video segments and identify the level of mathematics cognitive demand required of students in the observed integrated STEM activities. This particular rubric was selected over other rubrics we explored because of its focus on specifically categorizing the cognitive demand levels of student engagement for the mathematical tasks examined. The rubric, which includes five levels, has been found to be valid and reliable with overall cognitive demand level agreement at 81.8% (Matsumura et al., 2006). The Cronbach's alpha calculated for the consistency of the rating results yielded an alpha of 0.92 (Matsumura et al., 2006). These reliability measures have been reported as good overall. The rubric's inter-rater agreement is reported at a moderate value of 76.3% across pairs of raters (Matsumura et al., 2006). The five levels of the rubric begin with lower cognitive demand and increase in engagement at each level. As previously noted, Levels 1–4 in particular are derived from Smith and Stein's (1998) Characteristics of Mathematical Tasks at Four Levels of Cognitive Demand, such that these levels can be broken into lower-level (Levels 1 and 2) and higher-level (Levels 3 and 4) demand tasks. The rubric

Table 2 Overview of the Fourth and Fifth-Grade STEM Curriculum Units

Science Discipline	Engineering Challenge	Science Topics	Mathematics Topics
Physical Science	create a vaccine container not requiring a power source for warm climates	vaccines, heat energy and heat transfer, conductors, insulators	interpret line graphs, use data tables, measure temperature, perform operations with decimals in real-world problems
Earth Science	design a device to extract resources from a mining site on an exoplanet	renewable and nonrenewable resources, environmental impact	operations with decimals in real-world cost problems, measure areas by counting unit squares, convert improper fractions to mixed numbers
Life Science	design a greenhouse to grow food in cold climates	plant needs and parts optimal habitat for a living organism	data collection and analysis, surface area, cost analysis, line graphs

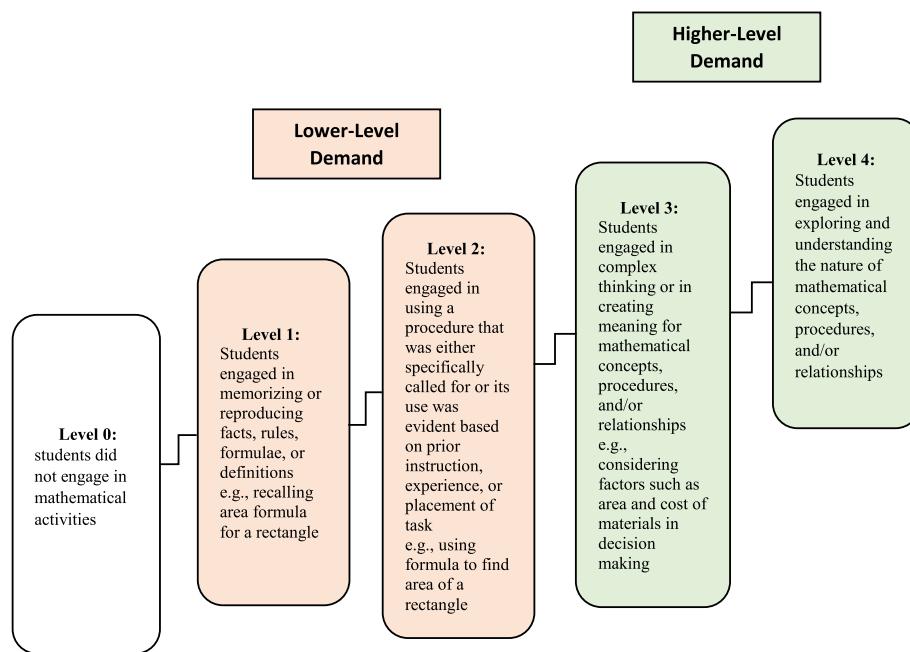


Fig. 1 The Characteristics of the Implementation of Mathematical Tasks at Five Levels of Cognitive Demand. Figure 1 Adapted from The Characteristics of the Implementation of Mathematical Tasks at Five Levels of Cognitive Demand (Adapted from Academic Rigor - Implementation of the Task rubric - Rubric 2 (Matsumura et al., 2006)

levels are presented visually in Fig. 1 with descriptions of each level in the section that follows.

The first level, Level 0, indicates that although a mathematical task might be explained to students by the teacher, this did not require cognitive engagement by the students. This may be represented by instances in which a teacher provides explanations, directions, instructions, and/or referred to mathematical objectives. The second level, Level 1, is marked by students engaging in mathematical tasks that focus on memorizing or reproducing facts, rules, formulae, or definitions without making connections to or meaning of the concepts at hand. Level 2 activities require that students engage in using a procedure that was either specifically called for or its use was evident based on prior instruction, experience, or placement of the task. In this, students follow a prescriptive method with little room to make connections to concepts or meaning underlying procedures used. For Level 2 mathematical tasks, students merely used procedure(s) that are specifically called for, requiring no effort by the students to use their initiative or make decisions.

Levels 3 and 4 are considered higher cognitive level tasks. Level 3 tasks are marked by students engaging in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships. These tasks require higher-order or complex thinking, but without obvious evidence of students' reasoning and

understanding. At this level, students engage in performing mathematical tasks or procedures with connections within mathematical concepts, however evidence of these connections is not explicit within the assigned tasks. At the highest level of cognitive engagement, Level 4, the mathematical tasks would engage students in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships (i.e., they used complex and non-algorithmic thinking). At this uppermost level, students are expected to use procedures with connections among mathematical concepts as they work on the assigned mathematical tasks.

The first and second authors used this rubric to independently code each of the 153 identified segments and established interrater reliability using Cohen's weighted kappa ($\kappa=0.80$). This demonstrates substantial to approaching almost perfect agreement (Cohen, 1960). During this phase of coding, it was necessary to further subdivide some of the identified segments as additional mathematics concepts or skills were embedded within them. Coders resolved disagreements through discussion until they reached a consensus for each identified segment. Once the codes were agreed upon, we were able to count the frequency of codes (rubric Levels) for each of our three cases. This allowed us to understand the frequency distribution of the cognitive demand in the mathematical tasks.

The final step in our analysis was to look for patterns in the segments that were coded at each level. This allowed us to understand not just the levels of the tasks, but what specifically students were doing while engaging in those tasks during an integrated STEM curriculum unit. This was first done within each case and then compared across the three cases in a cross-case comparison.

Qualitative findings

In this section, we first present the three cases (physical, earth, and life science) individually. Within each case, examples of mathematical tasks and how they were coded according to the different levels of cognitive demand required from students are explained. This is then followed by a cross-case comparison, addressing patterns and similarities across the cases. The science units served as the three cases for the study - Case 1: physical science, Case 2: earth science, and Case 3: life science.

Case 1: Physical science The physical science case consisted of four teachers who implemented the same curriculum unit. This curriculum unit was based on a design challenge in which students were asked to construct a container that does not require a power source to keep vaccines cool in warm climates. Therefore, students completed activities to determine which materials (e.g., metal, cotton fabric) were conductors or insulators, which would allow for heat energy to flow through quickly or slowly, respectively. Throughout this unit there were a total of 32 segments that included the presence of mathematics; this was the lowest frequency count of the three cases. An overview of the frequency of codes within this physical science unit is presented in Table 3.

When examining the presence of mathematics within each of the cognitive demand levels for this case, we found that of the 32 instances of mathematics, seven of these were coded at Level 0. These represent instances in which the teachers either outlined the objectives of the lessons as related to mathematics or they gave instructions or explanations related to mathematics. As a result, students were not directly engaged in any mathematical activity. For example, in one instance the teacher only explained to students that in this unit they were doing some data analysis, but data analysis took place in a later lesson.

Collectively, 16 of the segments in this case were coded at lower levels of cognitive demand (i.e., Level 1 and Level 2). These two levels accounted for 64% of the instances of the non-Level 0 mathematics in this physical science unit. One instance of a task that was coded at Level 1 was when students were asked to state how they will determine the mean melting time for three materials: metal, wood, and plastic. These instances were coded at this lower level of cognitive demand as students were tasked with reproducing a formula/fact without performing any mathematical procedures or calculations. Another instance of a task which was coded at this memorization level within this unit was when teachers displayed a graph for the cases of a medical condition (Pertussis) to the students. At this point, teachers called on students to recall vocabulary terms associated with constructing graphs. For example, "What do you call the horizontal axis of the graph? What letter is attached to it?" A student responded, "x-axis."

The number of mathematical tasks coded peaked at Level 2 (Table 3) such that students engaged in using a procedure, but the nature of the task did not allow them to make connections to the concepts or meaning underlying the procedure being used. These tasks ranged from the basic recall of mathematical facts to decision-making based on previously collected data. In this case, students were required to read digital thermometers, which is a step-by-step procedure that required limited cognitive demand for successful completion. In this instance, students were engaged in more than just memorizing or reproducing facts, rules, formulae, and definitions.

As students progressed along the unit, the level of cognitive demand for the mathematical tasks increased across all teachers. For instance, teachers provided students with opportunities to develop mathematical ideas related to graphs. Specifically, students analyzed and interpreted collected data to make the decisions needed to understand the science concepts. Particularly, one mathematical task required students to determine which of the materials (e.g., felt, bubble wrap, plastic wrap) they tested would be a good insulator based on the previously collected temperature-change data for these materials. This mathematical task engaged the students in making decisions about the best insulators needed for the specific engineering design challenge of creating a vaccine

Table 3 Distribution of Mathematics Cognitive Demand Codes within the Physical Science Case

Levels of Cognitive Demand	Level 0	Level 1	Level 2	Level 3	Level 4	Total
Number of codes	7	5	11	9	0	32

container and was coded at Level 3. Another instance that was coded at Level 3 was when students described trends and patterns in line graphs, they created that indicated temperature change over time when testing different vaccine containers. They subsequently interpreted these graphs and used the information to determine if their vaccine container met the criteria of the design challenge. Along with exploring the science concepts of heat conductors, students were required to use mathematical knowledge and skills simultaneously to assist in making an informed decision about the most suitable material to use. Hence, for both of these instances, students were engaged in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships.

Within this physical science case, there were no instances in which any mathematical tasks assigned by the teachers were coded at Level 4, the highest level of cognitive demand. In other words, none of the tasks required students to be engaged in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships.

Case 2: Earth science The second case was earth science, in which there were also four teachers. The design challenge for this curriculum unit was to create a mining tool that could be used to extract specific renewable and nonrenewable resources from different exoplanets mining sites. For this unit, there were a total of 78 instances of mathematics recorded and coded, which was the highest occurrence of the three cases. These instances included activities such as calculating costs of materials, profits, and/or area of surfaces. Table 4 provides the distribution for the frequency of the cognitive demand codes for the implementation of mathematical tasks for the earth science case. Based on this distribution, it was noted that most of these tasks were found to be at Level 2, with students engaged in utilizing previously taught procedures. The codes noted for Levels 1 and 3 were equal in number. Similar to the physical science case, there were no mathematical tasks coded at Level 4.

Of the 78 segments of mathematics in the earth science unit, 37 of them were coded at Level 0, which represents just less than half the total segments. With respect to Levels 1 and 2, a total of 32 instances were coded at

these levels, requiring lower cognitive demand thinking from students. The tasks within this unit were predominantly procedural in nature but were not directly connected to other mathematical concepts and hence could not be coded above Level 2. More than 78% of the occurrences of mathematics across Levels 1 through 4, in which students were engaged were at the lower levels (1 and 2) of cognitive demand. At Level 1 one task prevalent in the unit was related to calculating area. In particular, students needed the area formula to determine the area of two-dimensional figures. While all four teachers in this case presented mathematical tasks involving the concept of area, three specifically requested students to recall the area formulas. For example, one teacher stated, "We need to find the area," and then elicited from students how this could be accomplished. In one instance, a student responded by stating "multiply." Because this specific instance simply involved students reproducing a previously learned rule/fact rather than learning it, it was coded at Level 1. Students were also engaged in tasks related to using procedures previously taught. One example was when students were asked to convert improper fractions to mixed numbers by the teachers in this case. To complete this conversion, students used algorithmic steps and subsequently equated the whole numbers from the mixed numbers to the number of materials extracted from the mining site to the number of shipping container units filled.

Similarly, the instances within this earth science case that required students to calculate the total cost of mining the resources as well as extracting the resource from the mining sites were coded at Level 2. These calculations were categorized as such because coders considered that the addition and/or multiplication algorithms used to calculate total cost and finalize the budget would have been mathematics competencies/skills covered prior to these fourth and fifth-grade levels. Hence, they were considered below grade level as there was little ambiguity in these tasks, and the implementation of these tasks focused on students producing correct answers rather than developing mathematical understanding. However, in follow-up lessons, students calculated the profit in creating their proposed mining tool and cross referencing this with the area of environmental impact caused by using the tool. This cross-referencing activity enabled students to develop an understanding of maximizing

Table 4 Distribution of Mathematical Cognitive Demand Codes within the Earth Science Case

Levels of Cognitive Demand	Level 0	Level 1	Level 2	Level 3	Level 4	Total
Number of codes	37	9	23	9	0	78

profits, and they were also able to make connections between the purpose of the budget and its importance to the engineering design challenge. These calculations and connections required students to engage in complex thinking, and thus supported a deeper understanding of concepts and connected ideas instead of simply performing procedures. This requires some degree of cognitive effort and was accordingly coded at Level 3 as guided by the rubric. None of the mathematical tasks implemented within the earth science case required students' engagement in exploring and understanding the nature of mathematical concepts or procedures nor considerable cognitive effort from students; no instances were coded at Level 4 within this unit.

Case 3: Life science Five teachers implemented the life science curriculum unit. In this case, the curriculum unit called for students to design and construct a model greenhouse capable of maintaining an optimal temperature closest to 24°C (75.2°F) and maintaining a temperature between 18°C and 35°C (64.4°F and 95°F). As a result, students calculated the area of shapes, measured and recorded temperatures, and/or analyzed data on graphs. In total, there were 43 segments in which mathematics was observed. The distribution of the cognitive demand codes for this case is presented in Table 5. Within this case, the highest number of instances coded were at Level 2, signifying that students were mostly engaged in using procedures that they previously encountered whether it was in prior grades or within the said unit. Level 4 was not present in any of the instances of mathematics.

We noticed that 19 of the 43 segments with mathematics were coded at Level 0, slightly less than 50% of all total codes. On examining Levels 1 through 4, we observed that for Levels 1 and 2, there were 19 coded segments combined, which represents over 79% of the instances of mathematics. At Level 1 students were tasked with recalling the formula for finding the area of a triangle as they were to consider different possible shapes for greenhouse windows. As the lessons continued, there were instances where in addition to students recalling the area formula, teachers followed up such tasks by requiring students to then use formulas for calculating the areas of both triangles and squares for the two types of panels of their greenhouse designs. Since these mathematical tasks drew

on students' abilities to perform algorithmic calculations with limited cognitive demands, the tasks were coded at Level 2. The inclusion of this measurement concept of area with respect to also determining the window size for the greenhouse was similar across all five teachers' implementations of mathematical tasks.

To guide students along with the engineering design task for this curriculum unit, teachers-initiated class discussions about testing different materials that may be appropriate for covering the windows of students' greenhouse models so that the internal temperature can be kept within the required optimal range of 64.4°F and 95°F . In one such classroom discourse, the teacher shared, "So most of the temperature of the material dropped around 70° somewhere around there they kind of settled around the room temperature...based on that information, what have you learned about the material that would be best for your greenhouse?" A student responded, "I think the felt or tinfoil would work because the felt only went up to 86 [degrees], and the tinfoil only went up to 72 [degrees]." Such instances were coded at Level 3 because this type of higher-order questioning required some degree of complex thinking from students as they made connections with the results of the temperature data they previously collected and one of the criteria (optimal temperature range) for the engineering design challenge. Students were also entrusted with factoring in the cost of constructing the greenhouse and analyzing line graphs in their decision-making. As a result of teachers providing opportunities for students to acquire a deeper understanding and connection of concepts, all these instances were coded at Level 3. Notably, yet again there were no instances in which mathematical tasks required students to engage in considerable cognitive thinking and hence the uppermost Level 4 code was not applicable in this case.

Cross case comparison From the original 53 classroom observation videos identified as including mathematics content in some way, we found a total of 153 segments of mathematical tasks that spanned the first four levels of the cognitive demand rubric; there were no instances of Level 4 in any of our cases. However, the distribution of the codes in the levels revealed some similarities and variations across the cases.

Table 5 Distribution of Mathematics Cognitive Demand Codes within the Life Science Case

Levels of Cognitive Demand	Level 0	Level 1	Level 2	Level 3	Level 4	Total
Number of codes	19	8	11	5	0	43

Throughout the three cases, there were 63 instances coded at Level 0. At this initial level, students were not engaged in mathematical activity; instead, teachers across the units either provided directions, instructions, or lesson objectives related to concepts or procedures in mathematics. Of these 63 instances, the earth science case had the most occurrences of mathematics segments (37) coded at this lowest level, while physical science and life science were significantly less at seven and 19, respectively. This high occurrence of segments coded at Level 0 in the earth science unit was evident as to how instruction heavy these lessons were across all earth science teachers. In many instances throughout this unit, teachers explained to students what mathematical concepts were involved, for example, in outlining the design challenge teachers stated that it will be necessary to analyze the data or complete a material cost sheet.

At Level 1, students either recalled previously recorded temperature values as in the physical science case, or they were asked to reproduce previously learned area formulas as in the earth and life science cases. A comparison of these two cases indicated that the number of instances at Level 1 only slightly differed. The physical science unit had a total of five instances of mathematics observed at Level 1. Interestingly, across all three cases, this was the least represented cognitive engagement Level.

At the other lower-level cognitive demand phase, Level 2, the implementation of mathematical tasks for all three cases showed an increase in occurrences when compared to Level 1. This indicated that teachers assigned more mathematical tasks that required higher cognitive demand thinking from students. Specifically, across the three cases, there were more instances where teachers extended students' knowledge beyond recalling the area formulas and required students to calculate areas of triangles and squares. Just as noted for Level 1, the earth science case recorded the most instances across the three cases for Level 2. Students' activities at this level of cognitive demand included conversion between mixed numbers and improper fractions as well as calculating total costs. Whereas, for the physical science case, students read thermometers and for the life science case, they calculated areas of shapes. There were an equal number of implementations of mathematical tasks coded at Level 2 for the physical and life science units.

The implementation of mathematical tasks was equal in number at Level 3 for both physical science and earth science while life science recorded the least of the three cases. The mathematical tasks assigned to students in the life science unit that were coded at Level 3 drew upon

students' abilities to perform mathematical tasks involving procedures that were critical to decision-making for the engineering design challenge. Specifically, students were required to make decisions in relation to optimal temperatures, surface area, cost factor for the construction of the greenhouse, and analysis of data. Hence, students' decisions encompassed a combination of mathematical and science concepts along with engineering skills in their attempts to adhere to the criteria and constraints of building their greenhouse models. Students also made decisions in the physical science unit; however, these decisions were primarily based on the selection of the best insulating materials contingent on the change between initial and final temperatures. The decision-making, mathematics-related tasks that students were engaged in for the earth science unit were related to maximizing profits as well as comparing budgets and environmental impact.

In general, among the three cases, earth science had the greatest number of mathematical tasks despite there being more teachers implementing the life science unit. The total number of instances for the physical science unit was ten less than that for the life science unit and just less than half that for earth science. With respect to the overall distribution of codes throughout the three cases, over 70% of the codes for each science unit were at the lowest levels of cognitive demand. There is a remarkable absence of implementation of mathematical tasks that sought to promote the highest level of cognitive thinking from students. The lack of mathematical tasks at Level 4 in particular, meant students were not presented with opportunities to aptly understand and explore the nature of mathematical concepts, processes, or relationships. Such learning opportunities would be synonymous with using non-algorithmic thinking and procedures as well as exploring and extending students' thinking in relation to mathematical concepts and ideas within these science units.

Discussion

This study examined the presence of mathematics in integrated STEM instruction as well as the levels of cognitive demand required by mathematical tasks which were assigned to students within integrated lessons for physical, earth, and life science units. The examination of these two areas was driven by the notable under-representation of mathematical content within integrated STEM education (English, 2016; Marginson et al., 2013) and, more specifically, the level of mathematical thinking required from students in integrated STEM curricula (Baldinger et al., 2020; English, 2016; Stohlmann, 2018). Moreover,

we were drawn to the importance paid to a discipline-integrated approach in education (Li et al., 2020; Martín-Páez et al., 2019). It is clear from the literature that mathematical tasks ought to be intentionally included in integrated STEM lessons (Fitzallen, 2015; Maass et al., 2019; Shaughnessy, 2013). Moreover, attention must be given as to how the implementation of these mathematics tasks is combined with other disciplinary content to support mathematics concept development; this occurs when mathematics tasks are at higher levels of cognitive demand. These considerations in this study in relation to mathematics inclusion were fueled by the disparity that currently exists in comparison to the extent to which science is noticeably represented and emphasized in integrated STEM. As Ring et al. (2017) and Ring-Whalen et al. (2018) noted, teachers often position mathematics as “less than” in integrated STEM curricula, seeing it as a support for teaching science and engineering. As a result of this need, in this study, we explored both features.

Our initial quantitative analysis showed that the presence of mathematics in integrated STEM lessons resulted in statistically significant differences in mean rank order between integrated STEM lessons that included mathematics and without mathematics. From our findings, Items 4 and 5 on the STEM-OP were highlighted for two main reasons, their statistically significant differences in mean rank order as well as the direct presence and/or integration of multiple disciplines. Our finding that video-recorded observations that include mathematics scored higher on these items than those without mathematics reinforces the idea that including mathematical content within integrated STEM lessons correlates to overall increased cognitive engagement as well as the depth of content integration in a meaningful way. This can ultimately provide effective student learning opportunities. The results of the Mann-Whitney-Wilcoxon tests also showed that when teachers intentionally included mathematical content in integrated STEM lessons, there were significant statistical differences for other items among the two groups. In particular, we observed higher overall scores on Items 3–9, which indicates the vital role that mathematics has within integrated STEM education as measured by the STEM-OP (Dare et al., 2021).

In our second qualitative component, we observed that teachers in each science domain presented the mathematics and science content within an engineering design context that allowed students to develop an understanding of mathematical content (Moore, Stohlmann, et al., 2014) categorized at the different levels of cognitive demand. Even though mathematics was included in the integrated STEM lessons for this study, the mathematical tasks that students engaged in did not necessarily allow them to reach the highest level of cognitive

demand, Level 4. We were particularly drawn to the high number of instances coded at Level 0 across the cases - a total of 63 out of 153 instances. The rubric as proposed by Matsumura et al. (2006) outlines this initial level as the absence of direct student engagement in mathematical activities. There were many instances when teachers either described the lesson/unit objectives or gave explanations that were mathematical in nature. This signifies that although there was intent on the part of teachers to integrate mathematical activity within the integrated STEM units, the extent to which these activities engaged students was not always evident (i.e., they were not always necessarily reflected in actual implementation). With respect to the next two levels, our findings revealed that even though the earth science case contained the greatest number of mathematical tasks compared to the other two cases, the majority of these tasks were at the lower levels of cognitive demand: Levels 1 and 2. As noted in the cross-case comparison, of the three cases, the physical science unit recorded the fewest instances coded at Level 1, such that students engaged in mathematics beyond memorizing or recalling mathematical facts, formulas, or rules. Furthermore, both life science and physical science units reported the same number of instances coded at Level 2, thus indicating a similarity among teachers in these two units in requiring students to engage in using procedures when performing mathematical tasks. The implementation of these integrated STEM lessons demonstrated that the science teachers within this study possess mathematics subject-matter knowledge based on the mathematical tasks that they assigned to their students within the units. The issue, however, stands as to how science teachers can be taught or encouraged to successfully incorporate their mathematics knowledge with their science knowledge within integrated STEM lessons to effectively create tasks that facilitate high cognitive demand from their students.

For the instances in the integrated STEM lessons that were coded at a high level of cognitive demand, which were all at Level 3, we observed that the scientific content covered in the physical, earth, and life science units facilitated the use of mathematical reasoning skills by connecting the mathematics content to the engineering problem. For example, some tasks required that students engage in optimizing profits tied to creating a budget; representing and interpreting statistical data is necessary to make decisions about designing prototypes or selecting the appropriate materials for prototypes. These assigned tasks necessitated the meaningful integration of mathematical concepts within science and/or engineering contexts to address the engineering design challenge. This suggests that the teachers in this study could integrate content across the STEM disciplines and that

mathematics plays an integral role in this process, especially concerning the engineering design challenge. This notion was previously noted and supported by the quantitative findings of our study; the inclusion of mathematics in integrated STEM lessons positively affected how these lessons scored on the STEM-OP.

Despite this favorable effect of the presence of mathematics concepts within integrated STEM instruction, interestingly, our findings showed that of the 2030 STEM lessons observed, only 637 (31%) of them contained mathematical content. This low representation of mathematics within the integrated STEM lessons of this study as compared to the other disciplines of science, technology, and engineering, was no surprise as acknowledged by other researchers (e.g., English, 2016). Our closer examination of the mathematical tasks assigned within these integrated STEM lessons indicated that what is currently being included did not require the highest level of cognitive demand thinking, i.e. Level 4, from students. This mirrors the notion that when it comes to integrated STEM education, science teachers may inherently understand the importance of including mathematics, but do not always prioritize it in their curriculum design and implementation (Ring-Whalen et al., 2018; Roehrig et al., 2021). Our findings further suggest that when teachers do incorporate mathematics into their integrated STEM teaching, they may not consider the degree to which the mathematical tasks are cognitively demanding as their focus is more on creating opportunities for including mathematics to begin with (Ring-Whalen et al., 2018). It would appear that the science teachers' focus in this study was predominantly developing science and/or engineering concepts and practices, hence they incorporated mathematical skills/tasks which they felt were part of their students' prior knowledge or would have been covered in previous grades. For example, there were instances when teachers drew on mathematical concepts such as finding the area of two-dimensional shapes or calculating means. These topics were not taught in the integrated STEM lessons that were observed for this study: instead, teachers merely asked students to recall or perform those tasks, thus assuming that these concepts were previously learned by students.

Our exploratory work here suggests that new mathematical concepts ought to be introduced within STEM instruction if the goal of integrated STEM education is to ultimately develop content knowledge and skills across *all* STEM disciplines, not just science and engineering. Consequently, our work confirms what other researchers have called for in terms of a stronger emphasis on mathematics in integrated STEM education (e.g., English, 2016; Marginson et al., 2013; Shaughnessy, 2013). This will require energy and effort in terms of curriculum

design and implementation. For instance, including Level 4 tasks would likely require significant revisions to the curriculum, not to mention equipping the teachers (the curriculum designers) with in-depth knowledge of developing such higher-level cognitive demanding mathematical tasks.

Limitations

There are two main limitations to this study. First, the researchers did not conduct direct classroom observations of the teachers implementing the integrated STEM lessons. The primary data source was pre-recorded videos; therefore, it was challenging at times to capture all instances when students were directly engaged in mathematical tasks. This meant that we coded and analyzed the data based on the audio generated from teacher-student discussions and other classroom discourse and student activities captured by the camera located in the classroom. Second, most of the video observations from the original project in which the videos were collected focused on fourth through eighth-grade classrooms. We, however, analyzed a specific subset of these elementary-grade videos using the Academic Rigor - Implementation of the Task rubric (Rubric 2) (Matsumura et al., 2006). Analyzing the remaining fourth and fifth-grade video observations would allow for more comprehensive data collection and analysis processes. As a result, generalizability to other grade levels needs to be taken into consideration as the data collection and analysis were done for integrated STEM lessons taught by science teachers in grades at the elementary level. Applying the Implementation of the Task rubric at the kindergarten to second-grade level would allow for a comparison of the levels of cognitive demand for mathematical tasks between the elementary grades and the earlier grades. Additionally, future work should expand this study to middle and high school grade level curriculum units to understand how or if the level of mathematics cognitive demand changes given Baldinger et al.'s (2020) work indicating mathematics conceptual needs are not being met at the secondary level within discipline integrated settings.

Implications

In addition to science teachers' approaches to implementing integrated STEM education, the findings from this study also have implications for teachers in general who engage or who are considering engaging students in integrated STEM activities. There are also implications for professional development initiatives which are geared towards promoting integrated STEM teaching among teachers. Addressing the level of cognitive demand for assigned mathematical tasks in integrated STEM lessons needs more attention. In this study, science teachers

demonstrated that they assigned mathematical tasks at the higher cognitive demand, Level 3, however, unfortunately these were not implemented as frequently as lower-levels (Levels 1 and 2) mathematical tasks.

Providing additional support to science teachers while they design mathematical tasks alongside science and engineering content may assist them in creating more cognitively demanding mathematical tasks for their integrated STEM curricular units. This support afforded to science teachers should pay attention to how the constituent disciplines' concepts are interconnected. One beneficial implication of such support can assist these teachers in engaging students in engineering design challenges requiring mathematics and science knowledge within integrated STEM lessons. This is critical as generally mathematics and science standards are geared towards allowing students to develop a deeper understanding of the respective content concepts. Therefore, we recommend professional development that focuses on guiding teachers to intentionally consider the inclusion of higher-order mathematical tasks within integrated STEM teaching. One means to accomplish this is familiarizing teachers with rubrics such as the one employed in this study. This could help to ensure the designed tasks effectively meet the four levels of cognitive demand. The ideal professional development ought to be co-taught by a science and mathematics education expert to ensure both discipline content needs are being addressed. Additionally, at the school site level, science teachers should collaborate with their mathematics colleagues to ensure that grade-level appropriate mathematical content is adequately addressed/presented in STEM lessons and activities. This will require support from the school administration with respect to designating simultaneous planning times for teachers.

Our study implies that it is also imperative that teacher educators ignite the importance to teachers, both pre-service and in-service, of targeting mathematical tasks within integrated STEM lessons that require higher levels of cognitive thinking from students. Awareness of this can be instructive and beneficial for teachers in the planning and implementation of quality integrated STEM lessons. There is no doubt that teachers can strive to include opportunities for high-level thinking through cognitively demanding tasks by way of questioning, providing opportunities for students to make connections, and supporting their answers with explanations (Boston, 2012).

Conclusion

This study resulted in two significant findings for including mathematics in integrated STEM units. Using the STEM-OP that measures the degree to which integrated STEM is present, we found that adding

mathematics content to integrated lessons increases the degree of STEM integration as measured by our protocol. The second finding is that teachers in this study presented cognitively demanding mathematical tasks to students in integrated STEM lessons. However, these tasks mainly fell into the lower-level demand categories of Level 1 and Level 2, especially in the physical and life science units. These findings suggest that additional work in the area of inclusion of higher cognitively demanding mathematical tasks needs to be more specifically examined. Additionally, support and guidance for teachers with respect to effectively attending to extending students' mathematical learning within integrated STEM lessons.

An overarching goal of STEM integration is to ultimately provide experiences that build skills and concepts as equitably as possible within and across all its disciplines, therefore, addressing how mathematics tasks are being included is necessary (NAE and NRC, 2014). Our findings reiterate the call for more research that is needed to establish a better understanding of both the presence and quality of mathematics tasks in integrated STEM education.

Acknowledgements

This research was made possible by the National Science Foundation grants 1854801, 1812794, and 1813342. The findings, conclusions, and opinions herein represent the views of the authors and do not necessarily represent the view of personnel affiliated with the National Science Foundation.

Authors' contributions

ENF and LR conceptualized and designed the study based on previous work. ENF, LR, and JAE participated in the data analysis. ENF and LR worked on the initial draft of the manuscript. JAE and EAD led revisions and the final format of the manuscript. All authors have read and approved the final manuscript.

Funding

This research was made possible by the National Science Foundation grants 1854801, 1812794, and 1813342. The findings, conclusions, and opinions herein represent the views of the authors and do not necessarily represent the view of personnel affiliated with the National Science Foundation.

Availability of data and materials

Availability of data and materials Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 4 July 2022 Accepted: 28 December 2022

Published online: 01 February 2023

References

- Achieve. (2010). *Taking the lead in science education: Forging next-generation science standards international science benchmarking report*, (p. 66) <https://www.achievetestprep.org/files/InternationalScienceBenchmarkingReport.pdf>.
- Angier, N. (2010, October 4). *STEM education has little to do with flowers*. The New York Times. <https://www.nytimes.com/2010/10/05/science/05anger.html>.

Baldinger, E. E., Staats, S., Covington Clarkson, L. M., Gullickson, E. C., Norman, F., & Akoto, B. (2020). A review of conceptions of secondary mathematics in integrated STEM education: Returning voice to the silent M. In J. Anderson, & Y. Li (Eds.), *Integrated approaches to STEM education*, (pp. 67–90). Springer International Publishing. https://doi.org/10.1007/978-3-030-52229-2_5.

Becker, K., & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education*, 12(5/6), 23–37.

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, 113(1), 76–104. <https://doi.org/10.1086/666387>.

Boston, M. D., & Smith, M. S. (2011). A 'task-centric approach' to professional development: Enhancing and sustaining mathematics teachers' ability to implement cognitively challenging mathematical tasks. *ZDM*, 43(6–7), 965–977. <https://doi.org/10.1007/s11858-011-0353-2>.

Bybee, R. W. (2013). *The case for STEM education: Challenges and opportunities*. National Science Teachers Association.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>.

Council of Chief State School Officers, & National Governors Association Center for Best Practices. (2010). *Common core state standards for mathematics*. Council of Chief State School Officers.

Dare, E. A., Hiwatig, B., Kerathithamkul, K., Ellis, J. A., Roehrig, G. H., Ring-Whalen, E. A., Rouleau, M. D., 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). In *Proceedings of the 2021 ASEE Annual Conference & Exposition*. Virtual Conference. <https://peer.asee.org/37307>.

Dare, E. A., Ring-Whalen, E. A., & Roehrig, G. H. (2019). Creating a continuum of STEM models: Exploring how K-12 science teachers conceptualize STEM education. *International Journal of Science Education*, 41(12), 1701–1720. <https://doi.org/10.1080/09500693.2019.1638531>.

Dempsey, M., & O'Shea, A. (2020). The role of task classification and design in curriculum making for preservice teachers of mathematics. *Curriculum Journal*, 31(3), 436–453 Education Source.

Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159–199. <https://doi.org/10.3102/00346543053002159>.

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>.

Fitzallen, N. (2015). STEM Education: What does mathematics have to offer? In M. Marshman, V. Geiger, & A. Bennison (Eds.), *Mathematics Education in the Margins (Proceedings of the 38th Annual Conference of the Mathematics Education Research Group of Australasia)* (pp. 237–244).

Hahs-Vaughn, D. L., & Lomax, R. G. (2020). *An introduction to statistical concepts*, (4th ed.,). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781315624358>.

Hurley, M. 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>.

International Technology and Engineering Educators Association. (2020). Standards for technological and engineering literacy: The role of technology and engineering in STEM education.

Johnson, C. C., Mohr-Schroeder, M. J., Moore, T. J., & English, L. D. (2020). STEM integration a synthesis of conceptual frameworks and definitions. In *Handbook of research on STEM education*, (1st ed.,). Routledge. <https://doi.org/10.4324/9780429021381>.

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

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. <https://doi.org/10.1186/s40594-020-00207-6>.

Maass, K., Geiger, V., Ariza, M. R., & Goos, M. (2019). The role of mathematics in interdisciplinary STEM education. *ZDM*, 51(6), 869–884. <https://doi.org/10.1007/s11858-019-01100-5>.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50–60.

Margison, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM: Country comparisons: International comparisons of science, technology, engineering, and mathematics (STEM) education*. Australian Council of Learned Academics Final report.

Martín-Páez, T., Aguilera, D., Perales-Palacios, F. J., & Vilchez-Gonzalez, J. M. (2019). What are we talking about when we talk about STEM education? A review of literature. *Science Education*, 103(4), 799–822. <https://doi.org/10.1002/sce.21522>.

Matsumura, L., Slater, S., Junker, B., & Peterson, M. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the instructional quality assessment (CSE technical report 681)*.

McLoughlin, E., Butler, D., Kaya, S., & Costello, E. (2020). *STEM education in schools: What can we learn from the research?* (1.0). Dublin City University. <https://doi.org/10.5281/ZENODO.3673728>.

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

Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). A synthesis of conceptual frameworks and definitions. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & E. L. D. (Eds.), *Handbook of research on STEM education*, (pp. 3–16). Routledge. <https://doi.org/10.4324/9780429021381>.

Moore, T. J., Stohlmann, M. S., Wang, H.-H., Tank, K. M., Glancy, A. W., & Roehrig, G. H. (2014). Implementation and Integration of Engineering in K-12 STEM Education. In *Engineering in pre-college settings*, (pp. 35–59). Purdue University Press.

National Academy of Engineering and National Research Council. (2014). *Implications of the research for designing integrated STEM experiences. In STEM integration in K-12 education: Status, prospects, and an agenda for research*. National Academies Press. <https://doi.org/10.17226/18612>.

National Council of Supervisors of Mathematics (NCSM) and National Council of Teachers of Mathematics (NCTM). (2018). *Building STEM education on a sound mathematical foundation*. <https://www.nctm.org/Standards-and-Positions/Position-Statements/Building-STEM-Education-on-a-Sound-Mathematical-Foundation/>.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press. <https://doi.org/10.17226/13165>.

NGSS Lead States. (2013). *Next generation science standards: For states, by states: Vol. 1 (the standards) and Vol. 2 (appendices)*. National Academy Press.

Ring, E. A., Dare, E. A., Crotty, E. A., & Roehrig, G. H. (2017). The evolution of teacher conceptions of STEM education throughout an intensive professional development experience. *Journal of Science Teacher Education*, 28(5), 444–467. <https://doi.org/10.1080/1046560X.2017.1356671>.

Ring-Whalen, E., Dare, E., Roehrig, G., Titu, P., & Crotty, E. (2018). From conception to curricula: The role of science, technology, engineering, and mathematics in integrated STEM units. *International Journal of Education in Mathematics, Science and Technology*, 6(4), 343–362. <https://doi.org/10.18404/ijemst.440338>.

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

Roehrig, G. H., Rouleau, M. D., Dare, E. A., & Ring-Whalen, E. A. (2023). Uncovering core dimensions of K-12 integrated STEM. *Research in Integrated STEM Education*, 1, 1–25. <https://doi.org/10.1163/2726673-00101004>.

Sanders, M. E. (2012). *Integrative STEM education as "best practice"*. Griffith Institute for Educational Research, Queensland, Australia.

Savin-Baden, M., & Howell Major, C. (2012). *Qualitative research: The essential guide to theory and practice*. Routledge.

Shaughnessy, M. (2013). By way of introduction: Mathematics in a STEM context. *Mathematics Teaching in the Middle school*, 18(6), 324. <https://doi.org/10.5951/mathteamd.18.6.0324>.

Smith, M. S., & Stein, M. K. (1998). Reflections on practice: Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School*, 3(5), 344–350.

Stohlmann, M. (2018). A vision for future work to focus on the "M" in integrated STEM. *School Science and Mathematics*, 118(7), 310–319. <https://doi.org/10.1111/ssm.12301>.

Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. In *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE.

Tekkumru-Kisa, M., Stein, M. K., & Doyle, W. (2020). Theory and research on tasks revisited: Task as a context for students' thinking in the era of ambitious reforms in mathematics and science. *Educational Researcher*, 49(8), 606–617. <https://doi.org/10.3102/0013189X20932480>.

Tekkumru-Kisa, M., Stein, M. K., & Schunn, C. (2015). A framework for analyzing cognitive demand and content-practices integration: Task analysis guide in science: Task analysis guide in science. *Journal of Research in Science Teaching*, 52(5), 659–685. <https://doi.org/10.1002/tea.21208>.

U.S. Bureau of Labor Statistics. (2021, September 8). Employment in STEM occupations. www.bls.gov/emp/tables/stem-employment.htm.

Vasquez, J. A., Cary, S., & Comer, M. (2013). *Stem lesson essentials grades 3–8 integrating science technology engineering and mathematics*.

Wang, 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), 1–13. <https://doi.org/10.5703/1288284314636>.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1, 80–83.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com