ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #36801

Current Practices in K-12 Integrated STEM Education: A Comparison Across Science Content Areas and Grade-Levels (Fundamental)

Emily Anna Dare (Dr.)

Dr. Emily Dare is an Associate Professor of Science Education at Florida International University. Dr. Dare's research interests focus on K-12 STEM education. In particular, she is interested in supporting science teachers' pedagogy while also exploring their beliefs about teaching and learning. As science classrooms shift towards integrated STEM approaches that include engineering design as a central component, this is especially critical. Additionally, Dr. Dare has a passion for working with K-12 students to understand how changes in classroom instruction towards these integrated STEM approaches impact their attitudes towards and beliefs about STEM fields. In particular, she examines methods that positively impact girls, which may increase the number of women pursuing careers in STEM-related fields where they are currently underrepresented.

Joshua Alexander Ellis (Associate Professor of STEM Education)

Associate Professor of STEM Education

Mark Rouleau

Associate professor in the Social Sciences department at Michigan Technological University with a background in computational social science, research design, and social science research methods for assessment and evaluation purposes

Gillian Roehrig (Professor)

Elizabeth Ring-whalen

Elizabeth A. Ring-Whalen is an Assistant Professor of Education at St. Catherine University in St. Paul, MN and the Director of the National Center for STEM in Elementary Education (NCSEE). She received her PhD in Curriculum and Instruction - STEM Education from the University of Minnesota in 2017. Her research focuses on STEM education and what this looks like in PreK-12 classrooms and explores teachers' beliefs of integrated STEM as well as how these beliefs influence teachers' practices and student achievement in the classroom. Alongside this research, she has worked to explore the attitudes and beliefs teachers hold about cultural diversity and teaching culturally diverse students. She recently secured and acts as a PI on an NSF grant whose project outcomes aim to develop a K-12 STEM observation protocol that can be used in a variety of educational contexts through an online platform.

Current Practices in K-12 Integrated STEM Education: A Comparison Across Science Content Areas and Grade-Levels (Fundamental)

Abstract
Despite the popularization of integrated approaches to teaching science, technology, engineering, and mathematics (STEM) in policy documents, standards, and classrooms over the past several years, research related to the teaching of K-12 integrated STEM education continues to be impeded by the lack of observational tools available to education researchers. The work presented here uses a new observation protocol – the STEM Observation Protocol (STEM-OP) – designed for measuring the degree of integrated STEM teaching in K-12 science and engineering classrooms. 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, 3) 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. In this study, we used the STEM-OP to explore current practices in integrated STEM education, including comparisons across science content areas (Physical, Earth, and Life Science) and grade levels (elementary, middle, and high school). Our data set included a total of 2,030 video-recorded classroom observations from K-12 science classrooms where integrated STEM teaching was enacted through the use of an engineering design challenge to learn and apply science and mathematics content. Our results suggest that current K-12 science teachers miss opportunities in their lessons to relate content to students' lives, develop multiple solutions, use evidence-based reasoning, engage students in technology practices, and promote STEM career awareness. However, results from crosstab and non-parametric analyses reveal that various components of integrated STEM education occur more frequently and at higher levels in Physical Science and elementary classrooms compared to Life/Earth Science and middle/high school classrooms, respectively. Our work illustrates various places where integrated STEM education could be focused, including a better representation of engineering through developing multiple solutions and using evidence-based reasoning. Our work also highlights the importance of providing K-12 teachers with more opportunities to engage in professional development related to integrated STEM education. Implications for this work include those for K-12 teachers, teacher educators, classroom coaches, and administrators.

Introduction

Current policy documents have called for changes in K-12 science classrooms to employ integrated STEM strategies to provide a more authentic learning environment for a diverse population of students (National Academy of Engineering [NAE] and National Research Council [NRC], 2014). Integrated approaches to teaching science, technology, engineering, and mathematics (commonly referred to as STEM education) have been theorized to improve student learning (Jong et al., 2020) and better prepare students to address 21st century problems that are by nature interdisciplinary (e.g., Moore et al., 2020; NAE and NRC, 2014). As a result, a growing number of K-12 teachers have begun to incorporate engineering in their science classrooms. Such changes are the result of shifts in science standards to include engineering as evidenced by *A Framework for K-12 Science Education* (NRC, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013). Although national policy documents strongly support integrated approaches, which include mathematical and computational thinking in science education (NRC, 2012), there remains disagreement on models and effective approaches for its implementation. A recent review by

Moore et al. (2020) notes inconsistencies surrounding which and how many of the STEM disciplines should be integrated and puts into question the nature of the integration among disciplines; currently, there is no "correct" answer to this and some argue that a prescribed definition could be problematic (e.g., Bybee, 2010). However, it is clear that integrated STEM education is a multidimensional construct (e.g., Kelley & Knowles, 2016; Moore et al., 2020; Roehrig et al., 2021a), requiring considerations for both content and pedagogy; these complexities have made it challenging to define for practice.

Beyond the lack of an agreed-upon definition of integrated STEM for practice, an additional challenge lies in the dearth of observation protocols sensitive to measuring integrated STEM teaching and learning. Previously established protocols, such as the Reformed Teaching Observation Protocol (RTOP) (Sawada et al., 2002), the Classroom Observation Protocol for Undergraduate STEM (COPUS) (Smith et al., 2013), the Science and Engineering Learning Observation Protocol (Dringenberg et al., 2012) and the Classroom Observation Protocol for Engineering Design (COPED) (Wheeler et al., 2019), were not designed to measure integrated STEM teaching and focus on other elements, like teacher quality in a single discipline or engineering design. Although these instruments each excel in their intended areas, they were not designed with an integrated STEM education framework in mind, thus missing key characteristics of integrated STEM. This lack of an integrated STEM observation protocol for K-12 education has delayed progress with respect to enacting integrated STEM in K-12 classrooms, which inherently impedes our ability to measure related student outcomes as a result of integrated STEM teaching. This has also been problematic for educational research that seeks to understand and improve integrated STEM education. However, our team recently created the STEM Observation Protocol (STEM-OP) intentionally designed to measure the extent to which integrated STEM is practiced in K-12 science and engineering classrooms (Dare et al., 2021b); this has allowed us and other educational researchers to begin to ask and answer important questions about the implementation of integrated STEM education.

Because of the delay in establishing an observation protocol suitable for integrated STEM education, combined with the abundance of definitions that exist, little is currently known about how integrated STEM is currently implemented in classrooms. For instance, how do considerations such as different science content areas and grade-levels compare with one another? Although Roehrig et al. (2021b) examined K-12 integrated STEM curriculum units and found Physical Science curriculum units tended to be more integrated and conceptually coherent compared to Life and Earth Science units, that research did not focus on the enactment or implementation of the written curriculum. Guzey et al. (2016) similarly concluded that K-12 Physical Science curriculum units were better contextualized and more engaging for students than Life and Earth Science units. To contrast this, Pleasants et al. (2021) analyzed elementary curriculum units that followed an Engineering Design-based Science Teaching (EDST) model, reporting no differences across Physical, Earth/Space, or Life Science units with respect to conceptual connections. Research that has focused on implementation of integrated STEM in science classrooms has not included comparisons of different science content, focusing solely on one science content area (e.g., Dare et al., 2018, Guzey et al., 2017). Although several researchers have postulated that Physical Science content is better positioned for STEM integration due to the abundance of mathematical concepts and close relationship to mechanical engineering in particular (e.g., Dare et al., 2014; Guzey et al., 2016), this has not been confirmed through direct comparisons with other science disciplines.

The work of Roehrig et al. (2021b), Guzey et al. (2016), and Pleasants et al. (2021) illuminate another question with respect to integrated STEM education surrounding differences in grade-levels. While Pleasants et al. (2021) exclusively examined elementary units, both Roehrig et al. (2021b) and Guzey et al. (2016) did not report differences across grade levels. This highlights the fact that little is known about how integrated STEM unfolds in elementary versus middle versus high school teaching and learning contexts. These types of comparisons would allow educators and researchers to better understand how to support pre-service and in-service teachers across these various contexts, especially since teaching at the middle and high school levels has traditionally focused on siloed disciplines. To our knowledge, no published work has intentionally compared implementation of integrated STEM education across science content areas and grade-levels using a reliable observation protocol designed for integrated STEM education.

In response to this gap in the knowledge base, the work presented here shares how we used the STEM-OP (Dare et al., 2021) to measure the degree of integrated STEM within K-12 teaching, drawing upon a large dataset of video-recorded classroom observations, collected from a previous project in which teachers co-created integrated STEM curriculum units for use in a science classroom. This work attempts to address questions that educators have about integrated STEM in science classrooms concerning science content areas and grade-levels. With this in mind, we sought to answer the following research questions: 1) *To what extent is integrated STEM education being implemented in K-12 science classrooms as evidenced by our integrated STEM protocol?*; 2) *What differences in practice as measured by protocol scores, if any, exist across different science content?*; and 3) *What differences in practices as measured by protocol scores, if any, exist between protocol scores across grade levels?*

Theoretical Framework

The work presented here utilizes a framework comprised of seven central characteristics of integrated STEM education (Roehrig et al., 2021a): 1) a focus on real-world problems, 2) the centrality of engineering, 3) context integration, 4) content integration, 5) engagement in STEM practices, 6) 21st century skills, and 7) informing students about STEM careers. The framework arose out of the extant literature, drawing primarily from Kelley and Knowles' (2016) definition that broadly defines STEM education as "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). To expand upon and provide more specifics to that definition, our framework centralizes engineering design in which students are presented with an authentic problem to solve. While a full description of our framework can be found in Roehrig et al. (2021a), the following section summarizes the central characteristics.

To begin, we situate our framework by discussing engineering as a discipline within K-12 settings, most notably in conjunction with science as denoted by current K-12 science standards (e.g., NGSS Lead States, 2013; NRC, 2012). The literature has noted that within integrated STEM curriculum, real-world problems are often presented in the context of engineering design challenges that ask students to design solutions to a real-world problem (Berland & Steingut, 2016; Moore, Stohlmann, et al., 2014). Here, engineering also acts as a specific form of a real-world problem as a context for learning. These problems, however, should relate to or be aligned with students' interest and experiences to provide motivation for student learning (Djonko-Moore et al., 2018; Kelley & Knowles, 2016).

Students are expected to develop solutions to these real-world engineering problems by engaging in practices and 21st century skills used by STEM professionals. The development of design solutions relies on students using and developing an understanding of content from multiple disciplines (e.g., Thibaut et al., 2018). Further, students must engage in engineering practices (Berland & Steingut, 2016; NAE and NRC, 2014). Throughout the iterative design process, students are expected to assess and reflect upon how well their design addresses and responds to the problem at hand. This includes reflecting on meeting criteria and constraints, but it also requires that students use their knowledge of STEM content and data from iterative testing to refine their solutions (Siverling et al., 2019). It is critical that K-12 students have opportunities to fully engage in the iterative engineering design process, completing at least one cycle of redesign (Wendell et al., 2017), so that they can meaningfully participate in this reflective process.

Beyond placing real-world problems and engineering as central features (Roehrig et al., 2021a), the framework highlights two forms of integration: context and content. *Context* integration refers to explicit connections between STEM concepts and practices to those real-world problems (Berland & Steingut, 2016) such that there is clear alignment between the problem, the design challenge, and the STEM content. This contrasts with *content* integration, wherein integration across STEM content areas is included in the integrated STEM lessons, deviating from the siloed approach that has traditionally been used in K-12; it should be noted that in our framework we consider interdisciplinary science to be one content area (science). The intention of this is to help students "build knowledge and skill both within the disciplines and across disciplines" (NAE and NRC, 2014, p. 5). However, to make the most impact, teachers should make these connections clear and explicit to students (English, 2016; Kelley & Knowles, 2016), as students often do not spontaneously recognize these connections on their own (Tran & Nathan, 2010).

Our framework also addresses the need for students to engage in STEM practices and develop 21st century skills within integrated STEM learning. Broadly speaking, their inclusion is meant to present a more accurate depiction of the work of STEM professionals (Reyante et al., 2020), which relies on engaging in scientific and engineering practices (NRC, 2012), evidence-based reasoning (Siverling et al., 2019), and the creation, collection, manipulation, analysis, and visualization of data (Weintrop et al., 2016). This enables students to learn how to "do STEM" in a realistic context that shifts the pedagogy towards a more student-centered exploration in which they exercise agency (Berland & Steingut, 2016; Miller et al., 2018). Engaging in these STEM practices reinforces the development of 21st century skills, which are critical for students "to adapt and thrive in an ever-changing world" (Stehle & Peters-Burton, 2019, p. 2). With a focus on engineering, lessons that engage students in developing solutions to real-world problems and engineering design challenges inherently incorporate creativity and critical thinking, as there is no single correct solution to these complex problems (Simpson et al., 2018; Stretch & Roehrig, 2021). Iterative testing and learning from failure lead to stronger designs and innovation through the application of creativity and critical thinking (Simpson et al., 2018). Within K-12 settings, students are expected to work collaboratively within small groups to co-construct knowledge of STEM content and design solutions to real-world problems (e.g., Moore, Glancy, et al., 2014).

The last characteristic of our framework attends to the promotion of STEM career awareness within integrated STEM education. While engaging in STEM practices and developing 21st century skills help to promote positive identities towards STEM (Kitchen et al., 2018), students should also be exposed to details about STEM careers (Jahn & Myers, 2014; Luo et al., 2021). This reflects the current goals of promoting future participation in STEM careers to encourage a diverse population of future STEM professionals, as found in policy documents (NRC, 2012).

Methods

Context

Our project team had access to over 2000 video-recorded classroom observations of integrated STEM teaching in K-12 science classrooms that were collected as part of a prior, federally funded 5-year project. This prior project included professional development for the observed teachers that used a design-based framework for integrated STEM education drawing from Moore, Glancy et al. (2014) and Moore, Stohlmann et al. (2014) to develop and implement integrated STEM curriculum. These two frameworks centralize the role of engineering design in integrated STEM education by using an engineering design challenge as a context to teach STEM content while solving a real-world problem; these design challenges, which required students to learn and/or apply science content and practices as well as mathematics (primarily through data analysis), were often shared with students through fictitious client letters that asked the students to help them with their problem. These frameworks additionally guided the development of teacher team-created integrated STEM curriculum units. Participating teachers individually implemented their co-created curriculum unit in their classrooms during which the observations were video-recorded each day of implementation. These observations (typically ~50 minutes of instruction) represent a variety of classroom settings, including different grade levels, teachers, student demographics, science content, and engineering design challenges. Specifically, the data set includes 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 (grade 9 in particular) are represented to a lesser extent. The science content covered in these units spans several topics in Physical Science (e.g., force and motion), Life Science (e.g., ecosystems), and Earth Science (e.g., plate tectonics). A total of 48 unique curriculum units ranging in length from several days to several weeks were designed as part of this prior project. Table 1 provides a breakdown of the curriculum units by science content area, disciplinary topics, and grade band.

Table 1. *Summary table of observed curriculum units available.*

	Disciplinary Topics	**Curricula by Grade Band***
Physical Science	Heat Transfer and States of Matter Force and Motion Waves and Electromagnetism	3 Elem, 3 MS 4 Elem, 1 MS, 1 HS 5 Elem, 4 MS, 1 HS
Life Science	Ecosystems Natural Selection and Evolution Genetics	4 Elem, 3 MS 2 Elem, 1 Elem/MS, 3 MS 1 Elem, 1 Elem/MS, 1 MS
Earth Science	Plate Tectonics and Landforms Weather and Water Cycle Rocks and Soil & Renewable Energy	4 Elem, 3 MS 2 Elem, 2 MS 4 Elem (1 pre-K), 1 MS

* Elem = grades K-5, MS = grades 6-8, HS = grades 9-12

Data Collection

To answer our research questions, we used a new observation protocol for K-12 integrated STEM education (Dare et al., 2021) that we developed based on the Roehrig et al. (2021) framework (Table 2); a full description of the instrument development process can be found in Dare et al. (2021). 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. All items on the instruction met our selected acceptability threshold for inter-rater reliability of Krippendorff's $\alpha \geq 0.6$ with the exception of Item 5 (Integrating STEM content) that achieved $\alpha \geq 0.58$ (Krippendorff, 2004; 2011). In addition to the four descriptive levels, each item includes a brief description of the item and a set of user guidelines that provide more details about the item and how to score it, including non-exhaustive examples.

Table 2. *Alignment of Protocol Items with Integrated STEM Framework (Roehrig et al., 2021a).*

Item	*Item Name*	*Item Description*	*Alignment to Framework*
1	Relating Content to Students' Lives	Students' everyday and personal experiences from outside the classroom should be activated, meaningfully incorporated into the lesson.	Context Integration
2	Contextualizing Students Learning	Learning should be contextualized within an appropriate real-world problem or design challenge that connects to the content of the lesson.	A Focus on Real-World Problems Centrality of Engineering
3	Developing Multiple Solutions	Students should be encouraged to develop multiple solutions and evaluate them, identifying the relative advantages and disadvantages of each possible solution.	STEM Practices Centrality of Engineering
4	Cognitive Engagement in STEM	Students engage in learning within a STEM lesson at different cognitive levels. Including applying concepts in new situations and evaluating and analyzing concepts.	21st Century Skills
5	Integrating STEM Content	Within the lesson, multiple content areas are represented that cut across two or more STEM disciplines.	Content Integration
6	Student Agency	Epistemic agency refers to students' ability to shape and evaluate knowledge and knowledge-building practices in the classroom.	STEM Practices
7	Student Collaboration	Students should be encouraged to consider ideas from multiple individuals, critiquing these ideas and integrating new ideas into their existing understanding to co-construct a deeper understanding of STEM content.	STEM Practices
8	Evidence-Based Reasoning	Students should use and evaluate evidence to support their claims about phenomena and/or justify design decisions.	21^{s} Century Skills
9	Technology Practices in STEM	Students should engage in technology practices that are analogous to those used by practitioners of science, mathematics, and engineering.	STEM Practices 21st Century Skills
10	STEM Career Awareness	Students should be made aware of STEM careers at age-appropriate levels. These careers should directly relate to the lesson and expose students to future STEM career options.	Informing Students about STEM Careers

After establishing inter-rater reliability, the remaining video observations were scored using the STEM-OP. Throughout this process, some video observations were removed for various logistical reasons, including video and/or audio issues or incomplete observations that were significantly shorter than a class period; this accounted for a small percentage of total video observations available (<10%). Each video recorded observation served as our unit of analysis, representing one class period of instruction (~ 50 minutes). At the end of this process, we were left with a total of 2,030 scored video observations. These included 999 Physical Science, 434 Earth Science, and 597 Life Science observations. These videos also represented 885 elementary (K-5), 1071 middle school (6-8), and 74 high school (9-12) classrooms.

Data Analysis

To determine the extent to which integrated STEM education occurred in our observed K-12 classrooms, we first examined the mean, median, and overall distribution of item scores across the entire data set. We then used two complementary methods to determine the extent to which integrated STEM instruction varied across grade level and science content type. In our initial method, we used crosstab analysis (Field, 2009) to compare the absence or presence of each item in a binary fashion; this allowed us to first understand in what contexts items were statistically more present, disregarding the extent or rigor to which it was included (i.e., the level scored on each item). In preparation for this, item scores for all observations were transformed into one of two categories: "absent" if an observation score was 0 for that item (item code remained 0) or "present" if the item score was non-zero (item was then re-coded to 1). This allowed us to see variation in how often items were observed ("present") versus not observed ("absent"). A chi-squared test was then used to determine if any differences in the observed presence of our items were statistically significant across content area and grade-level, respectively. In other words, we wanted to know if certain items occurred more or less frequently in any capacity in some types of classrooms compared to others. Due to the non-normal distribution of most of our item scores, we then used a Kruskal-Wallis test (Siegel & Castellan, 1988) to analyze the means of our original item scores (scaled 0-3) across science content area and grade-level to determine the extent to which these different classroom types differed in terms of the degree (or level of rigor) to which a given item was implemented. This second analysis allowed us to expand our understanding from the crosstab analysis to consider the degree to which each item was present in a given context. We also conducted post-hoc Dunn tests (Siegel & Castellan, 1988) when item mean differences were found to be statistically significant to identify the item level(s) in which these differences occurred.

Findings and Discussion

Due to the exploratory nature of our work, the following sections combine findings from our analysis alongside a discussion related to each of our three research questions: 1) *To what extent is integrated STEM education being implemented in K-12 science classrooms as evidenced by our integrated STEM protocol?*; 2) *What differences in practice as measured by protocol scores, if any, exist across different science content?*; and 3) *What differences in practices as measured by protocol scores, if any, exist between protocol scores across grade levels?*

Overall

The means and medians of our item scores were relatively low across all ten items observed in our 2,030 classroom videos (Table 3). This suggests that current classroom practices related to integrated STEM education are not necessarily aligned with the aspirations of the instrument and the theory supporting it. This is most notable for Items 1, 3, 8, 9, and 10. The frequency distributions for each item further demonstrate not only low scores in general, but a non-normal distribution of scores for most items (Table 4), typically skewed towards lower values. While some items, such as Item 2, are rather normally distributed, several items stand out as skewed towards scores of 0 (Items 1, 9, and 10), indicating a complete absence of those features of integrated STEM instruction in many of our observed classroom videos.

Table 3. *Table of Item Protocol Means, Scored 0-3.*

	Item 1	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**Item 7**	**Item 8**	**Item 9**	**Item 10**
Mean	0.55	1.88	0.83	1.95	0.98	1.16	1.70	0.86	0.24	0.50
Median	0	2	0	2	1	1	1	0	0	0

Table 4. *Frequency Distributions of Protocol Items (percentage).*

Score	**Item 1**	**Item 2**	**Item 3**	**Item 4**	**Item 5**	**Item 6**	**Item 7**	**Item 8**	**Item 9**	**Item 10**
0	63.17	25.02	41.53	7.88	33.05	18.76	20.99	56.45	80.54	75.12
1	16.01	17.04	27.29	36.45	39.85	45.47	34.09	21.23	13.55	2.96
2	8.42	18.23	16.21	26.35	19.01	29.06	13.05	8.13	4.33	18.37
3	12.41	39.70	14.98	29.31	8.08	6.70	31.87	14.19	1.58	3.55

Discussion: Overall Extent of Integration

The low means and medians of most of our protocol item scores (Table 3) reveals that, overall, the practices observed in our data set do not yet reflect the ambitious goals of integrated STEM education as outlined in our theoretical framework and instrument; for one, we would have expected more normal distributions. Interestingly, the item with the highest mean score was Item 4 ($M_{\text{Item 4}} = 1.95$, $Mdn_{\text{Item 4}} = 2$), which relates to student cognitive engagement. This is perhaps not entirely surprising, as inquiry-based and student-centered learning that allows students to engage in higher cognitive levels has been a focus of science teaching for several decades (NRC, 2000). The second highest scoring item was Item 2 ($M_{\text{Item 2}} = 1.88$, $Mdn_{\text{Item 2}} = 2$), which relates to the context of the lesson as framed by an overarching real-world problem. Much like the literature reports (e.g., Berland & Steingut; Moore, Stohlmana, 2014), these curricula were framed by a central engineering design challenge. The third highest scoring item, Item 7 ($M_{\text{Item 7}} = 1.70$, $Mdn_{\text{Item 7}} = 1$), relates to student collaboration, another feature of integrated STEM education that has been a dominant component of science teaching practice for several years (Baines et al., 2003). These three features of integrated STEM education were attended to at the highest levels by teachers within our data set.

Our item mean scores (Table 3) and frequency distributions (Table 4) also indicate that there are several items that – while they align to theory related to integrated STEM education – are not currently routinely enacted in the classrooms that comprise our data set. In particular, Items 1 (Relating content to students' lives), 9 (Technology practices in STEM), and 10 (STEM career awareness) were frequently absent (i.e., received scores of 0) in our observations. Although we have identified these as key to integrated STEM education (NAE & NRC, 2014; Roehrig et al., 2021a) they are clearly areas in need of attention when it comes to teacher learning of integrated STEM. For example, knowing where technology fits into integrated STEM education has been challenging (e.g. Cullen & Gou, 2020; Ellis et al., 2020). Further, some visions of integrated STEM education do not attend to lived experiences (e.g., Kelley & Knowles, 2016), which has been pointed out by critics of integrated STEM education (e.g., Gunkel & Tolbert, 2018). The lack of attention to STEM careers may reflect that some teachers have little knowledge of STEM careers and thus do not intentionally include it in their teaching (Cohen et al., 2013). It should also be noted that in the prior project from which these videos were obtained, these particular components were not emphasized in the associated professional development, although they were included to some extent.

Concerning the inclusion of engineering within integrated STEM education, we also observed low average scores related to engineering design and decision making as evidenced by Items 3 ($M_{Item\ 3} = 0.83$, $Mdn_{Item\ 3} = 0$) and 8 ($M_{Item\ 8} = 0.86$, $Mdn_{Item\ 8} = 0$), respectively. These two items collectively address the need for divergent thinking through the development of multiple solutions (Item 3) and evidence-based reasoning (Item 8). Without these two components, engineering design looks more like tinkering through trial and error rather than critically thinking about design solutions (e.g., Crismond & Adams, 2012; Siverling et al., 2019). What is surprising with this finding is that the contextualizing professional development heavily focused on these two components, so the low item scores could reflect how the curriculum units were designed and implemented. This is to say that within a two-week long curriculum unit, there may have been fewer days dedicated to the engineering design process and decision making compared to days dedicated to learning the science content needed to address the engineering design challenge; an examination of how these items were scored (including presence and rigor) throughout a given curriculum unit may help to elucidate this to a greater extent. In other words, examining a day-by-day implementation could shed light on how each of these components were addressed throughout the entire curriculum unit implementation. Although Item 2, which serves to measure the extent to which student learning is contextualized, was found to be one of our highest scoring items across the data set (Table 3), we can conclude that actual engagement in engineering design and engineering thinking may not account for much of the observed lessons.

Comparing Science Content

Our crosstab results (Table 5) show that the percentage of classrooms in which Items 7 (Student collaboration) and 10 (STEM career awareness) were observed to be present (i.e., non-zero scores) was essentially the same across classrooms regardless of science content. On the other hand, there were statistically significant observed differences in the percentage of observations in which Items 1, 2, 3, 4, 5, 6, 8, and 9 were observed to be present (non-zero scores) depending on the science content focus. In all these instances, the items in question were more frequently present in Physical Science classrooms than in Earth Science classrooms. Additionally, these items were also observed more frequently in Earth Science than in Life Science classrooms. We should note that, although these results were statistically significant, substantively speaking, the observed difference in the presence of Items 2 (Contextualizing student learning) and 4 (Cognitive engagement in STEM) between Physical Science and Earth Science was less than 2%. This was also the case for Items 4 (Cognitive engagement in STEM) and 8 (Evidence-based reasoning) between Earth Science and Life Science. With these caveats in mind, our results indicate that Physical Science seems to do the best in terms of including the practices present in our protocol items relative to Earth Science and Life Science, particularly for Items 1, 3, 5, 6, 8, and 9. Finally, we should note that the greatest room for improvement seems to be with Items 1, 9, and 10, considering how few classrooms of any science content type implemented these aspects of integrated STEM instruction, ranging from 10.2% to 34.8%.

Table 5. *Presence (Non-Zero Score) of Items (percentage) Broken Down by Science Content Area.*

Item	*Physical Science*	*Earth Science*	*Life Science*	χ^2	*p*
1	34.8%	31.1%	26.5%	12.171	0.002**
2	82.8%	81.8%	73.47%	20.424	<0.001***
3	51.9%	49.3%	41.7%	15.610	<0.001***
4	99.7%	98.4%	97.8%	12.891	0.002**
5	69.1%	62.4%	56.3%	26.990	<0.001***
6	85.1%	78.6%	72.0%	40.169	<0.001***
7	86.9%	87.8%	85.6%	1.108	0.575
8	50.3%	42.4%	42.2%	12.941	0.002**
9	27.2%	16.1%	10.2%	72.759	<0.001***
10	23.7%	27.9%	24.6%	2.826	0.243

$^{\sim}$ p <.1, *p<.05, **p<.01, ***p<.001

In comparing the item scores across the science content areas (Physical Science, Earth Science, and Life Science) using the Kruskal-Wallis test, several items (Items 1, 2, 3, 5, 6, 8, and 9) exhibited mean scores that were significantly different across these science content areas at the α = .95 level (Table 6); comparatively, Items 4 (Cognitive engagement in STEM), 7 (Student collaboration), and 10 (STEM career awareness) did not show significant differences. Post-hoc analyses via the Dunn method revealed statistically significant differences between Physical Science and Life Science mean scores for Items 1, 2, 3, 5, 6, and 9, with statistically significant differences also occurring between Physical Science and Earth Science scores for Items 5, 6, and 9; these results reflect the results from crosstab analysis, but also suggest that the rigor of implementation was also higher in Physical Science. For all items with statistically significant differences in mean scores, we observed that Physical Science observations scored higher on these items than both Earth Science and Life Science. Additionally, for Items 2 and 6, post-hoc analysis revealed statistically significant differences between Earth Science and Life Science mean scores, with Earth Science outperforming Life Science.

Table 6. *Results of Kruskal-Wallis Test Comparing Physical, Earth, and Life Science.*

Item	M_{PS} (SD)	M_{ES} (SD)	M_{LS} (SD)	χ^2	*p*
1	0.63 (0.99)	0.50 (0.86)	0.46 (0.88)	13.06	0.002**
2	1.95 (1.14)	1.95 (1.17)	1.72 (1.24)	14.50	<0.001***
3	0.86 (1.01)	0.85 (1.03)	0.77 (1.06)	7.352	0.025*
4	1.97 (0.83)	1.96 (0.89)	1.91 (0.88)	1.553	0.460
5	1.14 (1.00)	0.82 (0.78)	0.83 (0.91)	48.60	<0.001***
6	1.24 (0.72)	1.15 (0.79)	1.03 (0.79)	26.22	<0.001***
7	1.70 (1.08)	1.76 (1.08)	1.68 (1.09)	1.518	0.469
8	0.90 (1.09)	0.83 (1.15)	0.82 (1.13)	6.721	0.035*
9	0.33 (0.61)	0.20 (0.51)	0.13 (0.41)	71.62	<0.001***
10	0.46 (0.85)	0.59 (0.99)	0.51 (0.94)	4.250	0.119

$^{\sim}$ p <.1, *p<.05, **p<.01, ***p<.001

Discussion: How do Science Content Areas Compare?

When comparing science content areas, we learn slightly different information from our two analyses concerning both the presence of items and the rigor in which they were implemented in the observed classrooms. Both sets of analyses indicate no substantive variation of Items 7 (Student collaboration) and 10 (STEM career awareness) across Physical Science, Life Science, and Earth Science with respect to presence or rigor. Although the crosstab analysis found that the presence of Item 4 (Cognitive engagement in STEM) varied, this was not substantial; considering the

Kruskal-Wallis test also did not show differences in rigor (i.e., mean item scores), we can conclude that Item 4 also does not substantively vary by science content area. The lack of variation in Items 4 and 7 is interesting to note as these two items were among the top-scoring items in general (Tables 2 and 3). This further suggests that science teachers, regardless of what content they are teaching, are equally likely to cognitively engage students in their learning (Item 4) and to include student collaboration (Item 7), likely because these two pedagogical ideas have been staples in science teaching for several decades (Baines et al., 2003; NRC, 2000). The lack of variation in Item 10 across content areas may be due to its generally low presence in our data set at large.

In comparing the two types of analysis, we found that Items 1, 2, 3, 5, 6, and 9 were not only significantly more frequently observed in Physical Science, but were also implemented more rigorously compared to Earth Science and Life Science. Although Item 8 (Evidence-based reasoning) was present more frequently in Physical Science classrooms, the comparison of the mean scores, while statistically significant, were not substantively so as the post-hoc test did not reveal significance. In general, Physical Science observations appear to be better connected to students' lives (Item 1), include more context through real-world problems or engineering design challenges (Item 2), require the development of multiple solutions (Item 3), better integrate STEM content (Item 5), provide increased student agency (Item 6), and incorporate more technology practices (Item 9) than Life Science observations. Further, Physical Science observations showed stronger content integration (Item 5) and used more technology practices (Item 9) than Earth Science observations.

Our findings suggest that Life Science in particular may be disadvantaged when it comes to integrated STEM education as measured by the STEM-OP, which reflects other research noting the complexity and challenges of integrating Life Science with engineering (Guzey et al., 2017) and how Physical Science may be better positioned for the integration of engineering in particular (Dare et al., 2018; Guzey et al., 2016). This finding elucidates the idea that Physical Science curriculum units may be integrated and more conceptually coherent as indicated by Items 2 and 5 in particular. We also found that Earth Science lessons were more frequently and better connected to real-world problems or design challenges (Item 2) and provided increased student agency (Item 6) compared to Life Science lessons. In particular, the finding related to Item 2 reflects the work of Roehrig et al. (2021b), who found that more Earth Science curriculum units were integrated and coherent compared to Life Science units, of which none fell into this category. Unlike Pleasants et al. (2020), our results show clear differences in integrated STEM education among science content areas. However, these analyses above were conducted across all grade levels. Broadly speaking, then, it appears that across K-12 the science content area has some effect on the degree to which integrated STEM education occurs in practice.

Comparing Grade-Level

In reviewing the results of the crosstab analysis with respect to grade-level (Table 7), we found that the presence of Items 1 (Relating content to student lives), 4 (Cognitive engagement in STEM), 5 (Integrating STEM content), and 9 (Technology practices in STEM) were essentially the same regardless of grade-level. Across the remaining items, elementary classrooms tended to outperform both middle and high school classrooms in terms of the percentage of observations in which these items were present. The exception here was with Item 6 (Student agency) in which the item was observed more frequently in high school observations compared to both elementary and middle school. Two other patterns emerged from these findings. The first was the decrease in present (non-zero) scores on Items 2 (Contextualizing student learning), 7 (Student collaboration), and 10 (STEM career awareness) as grade-level increased. The second pattern revealed that Items 3 (Developing multiple solutions), 6 (Student agency), and 8 (Evidence-based reasoning) tended to be less present in middle school compared to either elementary or high school.

Table 7. *Presence of Items (percentage) Broken Down by Grade-Level.*

Item	*Elementary (K-5)*	*Middle (6-8)*	*High (9-12)*	χ^2	*p*
1	31.6%	31.2%	36.5%	0.903	0.637
2	83.4%	78.8%	54.1%	38.294	$<0.001^{***}$
3	52.8%	44.4%	51.4%	13.727	0.001^{**}
4	99.3%	98.5%	98.6%	2.913	0.233
5	65.6%	62.9%	56.8%	3.246	0.197
6	83.7%	76.3%	85.1%	18.027	$<0.001^{***}$
7	88.9%	85.3%	79.7%	8.638	0.013^{*}
8	49.4%	43.5%	47.3%	6.749	0.034^{*}
9	19.9%	19.2%	28.41%	3.639	0.162
10	28.7%	22.6%	12.2%	16.307	$<0.001^{***}$

$^{\sim}$ p <.1, *p<.05, **p<.01, ***p<.001

The Kruskal-Wallis comparison helped to provide further insight about the level to which these practices were incorporated into instruction (Table 8). Comparisons across grade levels (elementary, middle, and high) revealed differences in mean scores for Items 2, 3, 4, 6, 7, 8, and 10 with mean elementary scores higher than both middle and high school scores. In all cases, post-hoc Dunn tests revealed statistically significant differences between elementary and middle school mean scores. Additionally, we observed statistically significant differences in the mean scores for Items 2 (Contextualizing student learning) and 10 (STEM career awareness) between elementary and high school observations. For Item 2, we also observed a statistically significant difference in mean scores between middle and high school observations.

Table 8. *Results of Kruskal-Wallis Test Comparing Elementary, Middle, and High School.*

Item	M_{Elem} (SD)	M_{Middle} (SD)	M_{High} (SD)	χ^2	*p*
1	0.53 (0.90)	0.56 (0.96)	0.59 (0.94)	0.575	0.750
2	2.00 (1.14)	1.84 (1.19)	1.11 (1.21)	37.51	$<0.001^{***}$
3	0.90 (1.03)	0.78 (1.03)	0.77 (0.94)	9.993	0.007^{**}
4	2.03 (0.85)	1.88 (0.87)	1.93 (0.73)	14.21	$<0.001^{***}$
5	0.97 (0.89)	0.99 (0.97)	0.93 (1.04)	0.803	0.669
6	1.23 (0.75)	1.09 (0.77)	1.31 (0.74)	19.16	$<0.001^{***}$
7	1.78 (1.07)	1.64 (1.09)	1.69 (0.19)	7.787	0.020^{*}
8	0.94 (1.13)	0.81 (1.10)	0.77 (1.00)	7.528	0.023^{*}
9	0.24 (0.53)	0.23 (0.52)	0.47 (0.88)	4.996	$0.082^{\sim}$
10	0.59 (0.97)	0.45 (0.87)	0.26 (0.70)	17.18	$<0.001^{***}$

$^{\sim}$ p <.1, *p<.05, **p<.01, ***p<.001

Discussion: How do Grade-Levels Compare?

The results above show that three items (1, 5, and 9) demonstrated no difference by grade-level in presence or mean score (Tables 7 and 8). The lack of variation across Items 1 (Relating content to students' lives) and 9 (Technology practices in STEM) may be due to their overall lack of presence in all observed lessons (Tables 3 and 4), likely an effect of the contextualizing professional development having not specifically focused on these elements. However, importantly, the lack of statistical difference in the presence of and mean scores for Item 5 (Integrating STEM content) suggests that there are equal opportunities to integrate content from multiple STEM disciplines across elementary, middle, and high school observations.

Several items did demonstrate significant differences with respect to presence and rigor (Items 2, 3, 6, 7, 8, and 10), with elementary classrooms outperforming both middle school and high school. While Item 4 (Cognitive engagement in STEM) did not vary in terms of presence in lessons at different grade levels (Table 7), the mean item score was statistically higher in elementary observations compared to middle and high school (Table 8). This suggests that although teachers at all grade levels appear to incorporate Item 4 at any level equally in their classrooms, elementary classrooms are reaching higher levels of cognitive engagement for students compared to the other grade levels. Further, our crosstab analysis showed that the presence of Items 2, 7, and 10 *decreased* as grade-level *increased*. This may suggest that elementary teachers are more intentional and explicit in their guidance of student learning related to integrated STEM compared to middle and high school teachers, especially with respect to contextualizing student learning (Item 2), engaging students in collaboration (Item 7), and promoting STEM career awareness (Item 10). This may contribute to why Pleasants et al. (2020) saw no differences across elementary teacher's curriculum units across science disciplines; it is possible that elementary curricula are equally attentive to particular aspects of integrated STEM regardless of science discipline.

Our comparison of grade-levels with respect to the presence of Items 3, 6, and 8 "dip" in middle school, although the difference in mean item scores was only between elementary and middle school observations. This suggests that something odd may be going on in these middle school observations, although we cannot ignore the lower number of observations in high school potentially being a factor. We also cannot ignore the fact that all of the high school observations were in Physical Science classrooms (Table 2); knowing that Physical Science observations tended to outperform the other disciplines could be a factor here. Since Items 3 and 8 are more engineering-forward, this raises questions about how middle school students are presented with opportunities to engage in engineering design (Item 3) and evidence-based reasoning (Item 8). The dip for Item 6 is also noteworthy as it relates to student agency, and this seems at odds with how we traditionally think about student agency within science; we might expect that middle and high school students would be given less direction (and thus provided with more agency) from a teacher compared to elementary students. It is possible that, within the scope of this project, middle school teachers faced additional barriers from their students who had not yet been exposed to engineering previously. This could have made the learning curve more intense for these students who came from having a more traditional and teacher-directed science education compared to elementary students who had fewer years of experience with teacher-directed science instruction (i.e., a fresh start to addressing open-ended problems rather than trying to change students' perceptions of science class). It is not only the teacher who is responsible for updating their pedagogy when it comes to integrated STEM, but students must adapt to these changes in ways of learning as well.

Item 2 also offers interesting insight when comparing across grade-levels. For this item, we see the decrease in presence and mean item scores as grade-level increases. This is the only item in which for each grade-level increase, the means scores significantly decrease. Given the fact that the high school observations were all Physical Science, this is noteworthy as Tables 5 and 6 indicate Item 2 as more present and rigorous than in Life Science classroom; this strongly suggests that high school integrated STEM may not be as contextualized as elementary and middle school. As noted above, this may be related to elementary teachers being more direct and explicit about contextualizing student learning, recognizing the importance of getting students to “buy into” an engineering design challenge to situate their learning of science content.

Limitations

This work is primarily limited by the data set that we had access to, which limits our ability to make claims about all K-12 integrated STEM instruction. The data set reflects the work of teachers who participated in the same professional development that utilized a design-based framework (Moore, Glancy et al., 2014; Moore, Stohlmann, et al., 2014) for integrated STEM education. The frameworks used in that professional development are similar to our own theoretical framework (Roehrig et al., 2021a), which guided the design of the STEM-OP (Dare et al., 2021). However, the STEM-OP expands on those in a few notable ways that reflect the underrepresentation of Items 1, 9, and 10; none of the concepts evidenced by these items were explicitly addressed or emphasized in the professional development that contextualized the development of curricula that were video-recorded (and thus used in this study).

The video-recorded observations, although large in number, are also limited. First, lower elementary (i.e., K-2) and high school (i.e., grade 9-12) classrooms are underrepresented compared to other grade bands; therefore, our ability to make broad claims about these grade levels is limited. Second, the video-recorded observations are limited to the camera’s position within the classroom; this restricts our analysis of classroom events to only what the camera can capture. This limits our ability to use additional classroom artifacts (e.g., student work) to better understand the full lesson. In future work with this instrument, this may be overcome with conducting live observations and/or collecting other pieces of evidence of integrated STEM teaching and learning.

Implications and Conclusions

The overall low mean item scores suggest that the STEM-OP may be ahead of current teaching practice, but that it can be used as a new target for where K-12 integrated STEM education practice should be heading. Of note, it appears that teachers may benefit from targeted support in relating content to students’ lives (Item 1), technology practices in STEM (Item 9), and STEM career awareness (Item 10). Further, the low presence and lower mean item scores of Items 3 (Developing multiple solutions) and 8 (Evidence-based reasoning) suggest that science teachers may benefit from learning more about engineering and engineering education. Teacher educators should rethink teacher professional learning, especially around engineering; the inclusion of this within pre-service spaces is also necessary.

Because the science and STEM education communities’ understanding of integrated STEM education has not yet been well-defined when it comes to implementation and practice, this work helps us and others in the field better understand the current landscape of integrated STEM education in practice. Our work suggests that Physical Science may be more conducive for not only including these elements of integrated STEM, but also achieving higher scores on the protocol, possibly due to the naturally strong connections between science and engineering and science and mathematics. Additionally, this work suggests that elementary teachers may be more adept at enacting integrated STEM in their classrooms as measured by our protocol. Further examination of elementary classrooms may help to identify specific practices used, which could then be implemented in middle and high school classrooms; middle school teachers in particular may need to overcome additional barriers when considering the inclusion of engineering into their science classrooms. A deeper, qualitative analysis is warranted in order to better understand these overall patterns, and other measures may be necessary to address these concerns.

Acknowledgements

This work was supported by the National Science Foundation under Award Numbers 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.

References

Baines, E., Blatchford, P., & Kutnick, P. (2003). Changes in grouping practices over primary and secondary school. *International Journal of Educational Research*, *39*(1–2), 9–34.

Berland, L. K., & Steingut, R. (2016). Explaining variation in student efforts towards using math and science knowledge in engineering contexts. *International Journal of Science Education*, *38*(18), 2742–2761.

Bybee, R. E. (2013). *A case for STEM education*. National Science Teachers' Association Press.

Cohen, C., Patterson, D. G., Kovarik, D. N., & Chowning, J. T. (2013). Fostering STEM career awareness: Emerging opportunities for teachers. *Washington State Kappan*, *6,* 12–17.

Crismond, D. P., & Adams, R. S. (2012). The informed design teaching and learning matrix. *Journal of Engineering Education*, *101*(4), 738–797.

Cullen, T., & Guo, M. (2020). The nature of technology. In V. L. Akerson, V.L. & G. A. Buck (Eds.), *Critical Questions in STEM Education* (pp. 21-32). Springer.

Dare, E. A., Ellis, J. A., & Roehrig, G. H. (2014). Driven by beliefs: Understanding challenges physical science teachers face when integrating engineering and physics. *Journal of Pre-College Engineering Education Research, 4*(2), 47-61

Dare, E. A., Ellis, J. A., & Roehrig, G. H. (2018). Understanding science teachers' implementations of integrated STEM curricular units through a phenomenological multiple case study. *International Journal of STEM Education, 5*(4), 1-19.

Dare, E. A., Hiwatig, B., Keratithamkul, 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, July). 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.

Djonko-Moore, C., Leonard, J., Holifield, Q., Bailey, E., & Almughyirah, S. (2018). Using culturally relevant experiential education to enhance urban children's knowledge and engagement in science. *The Journal of Experimental Education*, *41*(2), 137–153.

Dringenberg, E., Wertz, R., Purzer, S., & Strobel, J. (2012). Development of the science and engineering classroom learning observation protocol. In *Proc. 2012 ASEE Annual Conf. & Expo*, pp.25.4651-25.465.8

Ellis, J. A., Wieselmann, J. R., Sivaraj, R., Roehrig, G. H., Dare, E. A., & Ring-Whalen, E. A. (2020). Toward a productive definition of technology in science and STEM education. *Contemporary Issues in Technology and Teacher Education – Science, 20*(3), 472-496.

English, L. D. (2016). STEM education K-12: Perspectives on integration. *International Journal of STEM Education*, *3*(1), 1–8.

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex, drugs and rock 'n' roll)* (3rd ed.). SAGE Publications.

Gunckel, K. L., & Tolbert, S. (2018). The imperative to move toward a dimension of care in engineering education. *Journal of Research in Science Teaching*, *55*(7), 938–961.

Guzey, S. S., Moore, T. J., & Harwell, M. (2016). Building up STEM: An analysis of teacher-developed engineering design-based STEM integration curricular materials. *Journal of Pre-College Engineering Education Research*, *6*(1), 11–29.

Guzey, S. S., Ring-Whalen, E. A., Harwell, M., & Peralta, Y. (2017). Life STEM: A case study of life science learning through engineering design. *International Journal of Science and Mathematics Education*, *17*(1), 23-42.

Jahn, J. L. S., & Myers, K. K. (2014). Vocational anticipatory socialization of adolescents: Messages, sources, and frameworks that influence interest in STEM careers. *Journal of Applied Communication Research*, *42*(1), 85–106.

Jong, C., Priddie, C., Roberts, T., & Museus, S. D. (2020). Race-related factors in STEM: A review of research on educational experiences and outcomes for racial and ethnic minorities. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L. D. English (Eds). *Handbook of research in STEM education* (pp. 278-288). Routledge.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(1), 1–11.

Kitchen, J. A., Sonnert, G., & Sadler, P. M. (2018). The impact of college-and university-run high school summer programs on students' end of high school STEM career aspirations. *Science Education*, *102*(3), 529–547.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30* (3), 411-433.

Krippendorff, K. (2011). *Computing Krippendorff's alpha-reliability*. Retrieved from https://repository.upenn.edu/asc_papers/43

Luo, T., So, W. W. M., Wan, Z. H., & Li, W. C. (2021). STEM stereotypes predict students' STEM career interest via self-efficacy and outcome expectations. *International Journal of STEM Education*, *8*(36).

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. *Journal of Research in Science Teaching*, *55*(7), 1053–1075.

Moore, T. J., Glancy, A. W., Tank, K. M., Kersten, J. A., & Smith, K. A. (2014). A framework for quality K-12 engineering education: Research and development. *Journal of Pre-College Engineering Education Research, 4*(1), 1-13.

Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). STEM integration: A synthesis of conceptual frameworks and definitions. In C. C. Johnson, M. J. Mohr-Schroeder, T. J. Moore, & L. D. English (Eds). *Handbook of research on STEM education* (pp. 3–16). Routledge.

Moore, T. J., Stohlmann, M. S., Wang, H.-H., Tank, K. M., Glancy, A., & Roehrig, G. H. (2014). Implementation and integration of engineering in K-12 STEM education. In J. Strobel, S. Purzer, & M. Cardella (Eds.), Engineering in precollege settings: Research into practice. Rotterdam, the Netherlands: Sense Publishers

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

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning.* The National Academies Press.
National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. The National Academies Press.
NGSS Lead States. (2013). *Next generation science standards: For states, by states*. National Academies Press.
Pleasants, J. Tank, K. M., & Olson, J. K. (2021). Conceptual connections between science and engineering in elementary teachers' unit plans. *International Journal of STEM Education*, *8*, 16.
Reynante, B. M., Selbach-Allen, M. E., & Pimentel, D. R. (2020). Exploring the promises and perils of integrated STEM, through disciplinary practices and epistemologies. *Science & Education*, *29*(4), 785–803.
Roehrig, G., Dare, E. A., Ellis, J. A., & Ring-Whalen, E. (2021). Beyond the basics: A detailed conceptual framework of integrated STEM. *Disciplinary and Interdisciplinary Science Education Research*, *3*(11).
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).
Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics*, *102*(6), 245-253
Siegel, S., & Castellan, N. J. (1988). *Nonparametic statistics for the behavioral sciences* (2^{nd} ed.). Mc-Graw-Hill.
Simpson, E., Bradley, D., & O'Keeffe, J. (2018). Failure is an option: An innovative engineering curriculum. *International Journal of Building Pathology and Adaptation*, *36*(3), 268–282.
Siverling, E. A., Suazo-Flores, A., Mathis, C. A., & Moore, T. J. (2019). Students' use of STEM content in design justifications during engineering design-based STEM integration. *School Science and Mathematics*, *119*(8), 457–474.
Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education*, *12*(4), 618-627.
Stehle, S. M., & Peters-Burton, E. E. (2019). Developing student 21st century skills in selected exemplary inclusive STEM high schools. *International Journal of STEM Education*, *6*(1), 6.
Stretch, E. J. & Roehrig, G. H. (2021). Framing Failure: Leveraging Uncertainty to Launch Creativity in STEM Education. *International Journal of Learning and Teaching, 7(2),* 123-133.
Thibaut, L., Knipprath, H., Dehaene, W., & Depaepe, F. (2018). How school context and personal factors relate to teachers' attitudes toward teaching integrated STEM. *International Journal of Technology & Design Education*, *28*(3), 631–651.
Tran, N. A., & Nathan, M. J. (2010). Pre-college engineering studies: An investigation of the relationship between pre-college engineering studies and student achievement in science and mathematics. *Journal of Engineering Education*, *99*(2), 143–157.
Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147.

Wendell, K. B., Wright, C. G., & Paugh, P. (2017). Reflective decision-making in elementary students' engineering design. *Journal of Engineering Education*, *106*(3), 356–397.
Wheeler, L. B., Navy, S. L., Maeng, J. L., & Whitworth, B. A. (2019). Development and validation of the classroom observation protocol for engineering design (COPED). *Journal of Research in Science Teaching*, *56*(9), 1285-1305.