Paper Under Review: Please do not disseminate.

Key Practices in P-6 Integrated STEM Education: Delphi Panel Insights

Leah R. Cheek[a]*, Vinson Carter[b], Michael K. Daugherty[c], and Christian Z. Goering[d]

[a-d]*Department of Curriculum and Instruction, University of Arkansas, Fayetteville, AR United States*

*lrcheek@uark.edu

Key Practices in P-6 Integrated STEM Education: Delphi Panel Insights

Over the past two decades, efforts have been made to improve STEM education for P-12 students, yet patterns of engagement remain persistently unchanged, potentially hindering interest in STEM careers (Archer et al., 2020; Emrey-Arras, 2018; Hodson, 2003; National Research Council (NRC), 2012; National Science Foundation, 2018). This research reports findings from a modified Delphi study involving expert teacher educators across the US. The research focuses on identifying crucial practices used by teacher educators to prepare candidates for integrated STEM (science, technology, engineering, and mathematics) teaching in P-6 classrooms. Additionally, the research highlights key attributes of integrated P-6 STEM education compared to single-discipline curricula. The study reveals 16 interconnected essential practices and 9 differentiating attributes that define integrated P-6 STEM education, offering insights for researchers, teacher practitioners, and policymakers.

Keywords: integrated STEM education, elementary teacher education, essential practices, delineating attributes, Delphi method

Paper Under Review

Introduction

Throughout the past two decades policymakers, practitioners and researchers have been summoned to pursue essential educational practices to address and implement improvements in science, technology, engineering, and mathematical (STEM) competencies for P-12 students (Archer et al., 2020; Emrey-Arras, 2018; Hodson, 2003; NRC, 2012). Moreover, the National Science Foundation, in partnership with the National Research Council, the National Science Teachers Association, Achieve, and the American Association for the Advancement of Science released the Next Generation Science Standards to not only provide a framework towards a common set of benchmarks, but to foster increased interest among students in STEM subjects (NGSS Lead States, 2013).

Correspondingly, the International Technology and Engineering Educators Association (ITEEA, 2000) developed Standards for Technological Literacy (STL) (ITEEA, 2000) followed by the updated Standards for Technological and Engineering Literacy (STEL) (ITEEA, 2020) as a resource for educators to detect how STEM constructs might align with lessons through the integration of curricula with other disciplines (Daugherty et al., 2021). Irrespective of the longstanding investment of time and resources towards STEM education, students' patterns of STEM involvement in schools continue to be persistently unchanged (Archer et al., 2020; Hutchison, 2012; Ro & Knight, 2016). This study aims to explore the foundational groundwork in P-6 STEM integration content and pedagogies within teacher candidate preparation programs by identifying a core set of integrated STEM essential practices and attributes.

Rationale for Study

A Delphi panel methodology was utilized in the study to harness diverse expertise and perspectives from stakeholders to form a consensus (Avella, 2016; Clayton, 1997; Custer, et al., 1999; Green, 2014; Hsu, et al., 2007). The design and purpose of this modified Delphi study was

Paper Under Review

to contribute to the emerging literature (Gonzalez & Kuenzi, 2012; Nowikowski, 2017; Rinke et al., 2016; Rose et al., 2017) by obtaining a consensus concerning essential practices teacher educators utilize to help prepare teacher candidates to effectively teach and implement integrated STEM (science, technology, engineering, and mathematics) in the P-6 classroom. Additionally, this study delineates the core attributes of integrated STEM education compared to single discipline curricula. This investigation may contribute to the growing body of research propelled by the prediction of future shortages in the workforce in STEM (U.S. Bureau of Labor Statistics, 2022). To attain these objectives, the following research questions were used to guide the study: According to a panel of experts, (1) what are the essential practices teacher educators use to prepare teacher candidates to effectively teach integrated STEM in the P-6 classroom? And (2) what are the delineating attributes that set integrated STEM education curricula apart from a single discipline curriculum?

Related Literature

As a society, we learn about the world and advance our understanding through science, technology, engineering, and mathematics practices and experiences (Kelley & Knowles, 2016) and policy advisors, researchers and educators continue to focus on enhancing STEM experiences and proficiencies for P-12 students. The National Science Foundation (NSF) has acknowledged that even fewer Americans are entering STEM disciplines in the workforce (NSF, 2018), and ostensibly, the concern about student engagement in STEM and the numbers pursuing advanced study and careers in STEM fields linger as a perplexing challenge. “Despite the national movement for K–12 STEM education and its corresponding push to develop STEM educators, comparatively little attention has been given to the content of STEM teacher preparation” (Rinke et al., 2016). Providing teacher candidates with foundational groundwork in STEM integration content and pedagogies in their teacher preparation programs may be vital to

Paper Under Review

improving the qualifications for these rising teachers and directly influencing choices for their future students (Hutchinson, 2012; Nowikowski, 2017).

Although the pedagogy of elementary STEM teachers has been recognized as a crucial initiative, originating with targeted teacher candidate preparation (Gonzalez & Kuenzi, 2012), elementary teacher candidate preparation programs have historically included few courses that emphasize STEM integration in teacher candidates' future classrooms (Rinke et al., 2016; Rose et al., 2017). U.S. governmental reports ranging from the directive of the Federal Coordination in STEM Education Task Force Committee (National Science and Technology Council (NSTC), 2012) to the Progress Report on the Implementation of the Federal STEM Education Strategic Plan (NSTC, 2021) prioritize improvement of content knowledge and the integration of STEM instructional proficiencies of teacher candidates. Unfortunately, there is an inconsequential consensus of what defines an adequate STEM elementary teacher candidate preparation program (Alexander et al., 2014; Bybee, 2013; National Academy of Engineering (NAE) & National Research Council (NRC), 2014). Approaches to STEM teacher candidate preparation varies greatly. While there are no commonly accepted attributes defining a STEM-focused elementary teacher candidate preparation program, elementary teacher candidates that attend a college or university which provides STEM teacher candidate preparation may have exposure to integrated STEM teaching and learning and may benefit from the instructional practices and resources.

Though the approaches to STEM teacher candidate preparation vary, elementary teacher candidates in these programs have experiences that provide a rationale for making correlations between STEM content and the feasibility of integrating those connections with other subjects such as English language arts and before they enter their own classrooms (Alexander, et al., 2014; Rinke, et al., 2016). Contemplating the national focus on STEM and the significance of

Paper Under Review

teachers of STEM education, defining the attributes and identifying the essential practices that advance STEM integration practices for teacher candidates may be a valuable piece of the framework puzzle to combat the current and persistent unchanged pattern of students' involvement in STEM.

Delphi Research Method

The principal objective underlying this mixed-methods Delphi study was for the researchers to comprehend "quantitative and qualitative research approaches, and the stakeholders representing the relevant communities of practice" (Collins, et al., 2021) enabling the description of the relationships among the attributes that define this phenomenon (Greene, 1987; Plano Clark & Ivankova, 2016). Clayton (1997) informs that Delphi panels have been utilized for decades by researchers across the social sciences to facilitate expert collaborate investigations. Evolving from experimental research conducted for the RAND Corporation in the 1950s, the Delphi panel approach has been applied to health care, education, and management to develop group consensus around specific questions and decisions. Groups of experts anonymously answer questionnaires, receive feedback that represents the "group response," discuss, and revise their answers to see whether they can approach expert consensus. Researchers often employ the traditional Delphi panel which includes three rounds of questionnaires providing an opportunity for researchers to achieve maximum reflection and input from the panel members. Typically, the purpose of the first round is to inform the theme, the purpose of the second round is to provide the panelists with feedback form the first round so they can rate the identified items, and the purpose of the third round is to provide feedback from the previous round to reach a final consensus (Green, 2014). Each of the three rounds were followed by summaries of the previous questionnaire data provided to participants to inform

Paper Under Review

decision making and the development of the next round's questions (Clayton, 1997; Green, 2014).

Delphi Panel Selection

The process of selecting the panel of experts is critical to the strength and success of the Delphi study (Clayton, 1997; Green, 2014; Koster et al., 2005). For this investigation, purposeful sampling was used to select members to comprise the panel of experts. Creswell (2013) indicates purposeful sampling, "will intentionally sample a group of people that can best inform the researchers about the research problem under examination" (p. 147). To determine the expert panel for this modified Delphi study, participants were selected from elementary teacher educators who teach methods of integrated STEM (science, technology, engineering, and mathematics) which culminate in teacher licensure for teacher candidates. Because a Delphi study depends on a group of experts arriving at a consensus and does not depend on statistical power, group size typically ranges between 10-30 panelists (Clayton, 1997; Okoli & Pawolowski, 2004). Based on national reputations for scholarly work in the field of elementary STEM education and membership(s) in one or more professional associations/organizations focused on elementary STEM education, 37 prospective panel members were selected and then invited by email to take part in the research and 13 of those experts agreed to participate in the modified Delphi survey. Each panel member held a doctorate degree, had been a teacher educator for at least five years, and was or had been a certified teacher with nine of the panel members reporting they had also been a teacher in grades P-6. The panel was representative of institutions of higher learning ranging in size from 5000 to more than 40,000 students from states in the Eastern, Southern, Mid-South, and Midwestern regions of the United States (see Table 1).

[Table 1. Key Demographics of Teacher Educator Delphi Panel near here]

Paper Under Review

Research Design and Procedures

A three round modified Delphi panel methodology was used in this investigation for iterative feedback based on a model conducted by Kloser (2014). The ethical aspects of this research were reviewed and approved by the Institutional Review Board (IRB) from the researchers' university in January 2023. Participants in the research were informed about the purpose of the study and provided voluntary informed consent before their involvement in the study. All three survey rounds were created and distributed to the expert panel through utilization of the Qualtrics™ Academic Survey Software.

Round 1

In the initial round of the Delphi panel, a survey was emailed during the first week of February 2023 consisting of background information about the study in an introduction, followed by three sections: (1) panelist ratings using a Likert scale and justification pertaining to integrated STEM education practices derived from literature; (2) panelists suggestions and descriptions to add to the list of provided practices; (3) anonymous demographic information.

The survey introduction included a brief overview of practices teacher educators use to prepare teacher candidates to effectively teach integrated STEM in the P-6 classroom and a definition of elementary STEM education complied from the literature:

> P-6 integrated STEM education is an approach to teaching and learning that blurs the division of traditional subject areas by integrating science, technology, engineering, and mathematics (STEM) with other curricula. In P-6 integrated STEM education, students construct authentic hands-on projects that involve applying knowledge and experiences from multiple STEM disciplines. By combining ideas, curiosity, and comprehension of the interconnected STEM subjects with reading and writing, students build critical

Paper Under Review

thinking and problem-solving skills relevant to the real-world and essential for success in STEM fields.

In Section 1, the panelists reviewed eight potential practices from the literature included to encourage the exchange as well as serve as a model for generation of additional practices. Panel members were asked to evaluate each practice and description, providing feedback by rating their level of agreement on a 5-point Likert scale ranging from "Not Important-Very Important." Following the rating, panelists were asked to provide an explanation and/or comment.

Section 2, was designed to foster input from the expert panel, generate additional essential practices, and uncover delineating attributes of integrated STEM education. After evaluating the provided practices in Section 1, the panel was given an open-ended opportunity to propose essential practice alternatives for integrated elementary STEM teacher education. Next, the panel was asked to identify the delineating attributes that set integrated STEM education curricula apart from a single discipline curriculum. The panelists were informed that their alternatives and attributes may be rated in the future as part of Round 2. Finally, the panel was presented with a section to input any additional comments. Round 1, remained open for three weeks.

Section 3 was devised to provide anonymous demographic data. Panelists were encouraged to select the best answer or answers, including an option that stated they preferred not to answer.

Round 1 Analysis

In this study, a Concurrent Quan + Qual parallel mixed methods model design (Plano Clark & Ivankova, 2016) was employed, enabling the simultaneous collection of both

Paper Under Review

quantitative and qualitative data. The primary objective of adopting this concurrent approach was to comprehensively investigate the essential practices employed by teacher educators in preparing teacher candidates to effectively teaching integrated STEM in P-6 classrooms and to identify the attributes that set integrated STEM education curricula apart from a single discipline curriculum.

Quantitative measures were utilized to assess the rating of these practices, while qualitative data provided deeper insights into the underlying reasons for these ratings. By combining and integrating these two types of data, a more comprehensive and robust understanding was drawn. Figure 1 illustrates the process of data blending and how it contributed to a more nuanced understanding of the subject under investigation.

To analyze the quantitative information, survey data from the provided eight potential practices from literature were reviewed and categorized using basic descriptive statistics including mean and mode as a measure of central tendency, and standard deviation as a measure of variability (Clayton, 1997). As the Delphi methodology is an iterative approach used to garner expert opinions on a particular subject, the results and recommendations regarding an appropriate cut score have not shown a consistent agreement in literature (Hsu & Sandford, 2007; Kloser, 2014). A mean responses score of ≥ 3 and a mode response score of ≥ 3 for Round 1 was selected at the beginning of the study as the most appropriate picture of the expertise of the panel due to the desire to not reject any essential practices too early in this process (Green, 2014). All eight provided practices remained for another evaluation in Round 2.

Furthermore, the researchers independently reviewed the qualitative data from the panel's suggested essential practices, identified attributes, and comments from Round 1 and then coded inductively using memos (Cresswell, 2013; Saldana, 2009). The input from the panel was

Paper Under Review

grouped into categories, reviewed, and then the inductive coding process was repeated. Finally, a consensus on the identified themes of the panel's suggestions and comments was reached, with both researchers indicating agreement. The collective Delphi panel's responses to the open-ended questions in Round 1 provided 14 additional or alternative considerations of essential practices in P-6 integrated STEM teacher education. Additionally, the researchers identified 11 attributes that set integrated STEM education curricula apart from a single discipline curriculum through the panel's qualitative contribution. The merging of both the quantitative data and the qualitative data in Round 1 informed the structure of Round 2 (see Figure 1).

[Figure 1. Concurrent Quan + Qual Parallel Mixed Methods Model Design near here]

Round 2

Round 2 was influenced and shaped based on the descriptive statistics and qualitative summaries of panelist's feedback during Round 1. The purpose of Round 2 was to establish consensus on topics that were identified as important practices and on a list of attributes that distinguish integrated STEM education curricula from a single-discipline curriculum. The second round of the modified Delphi study was distributed electronically during the first week of March 2023 and remained open for three weeks. Email reminders were sent to the participants once a week. This survey consisted of three sections: (1) rating results from Round 1 of the provided integrated STEM education practices derived from literature, and an illustrative summary of the panel's explanations; (2) panel originated additional or alternative considerations or solutions to essential practices in P-6 integrated STEM teacher education derived from Round 1; (3) panel originated attributes that set integrated STEM education

Paper Under Review

curricula apart from single discipline curricula derived from Round 1. At the conclusion of each section in Round 2, panelists were asked to provide an explanation and/or comment.

In Section 1, panel members were presented with a summary of the descriptive statistics generated in Round 1 for each of the provided eight potential practices identified in literature. They also received an illustrative summary of the qualitative data (see Figure 2). Panelists were asked to confirm their agreement with the results from Round 1 on a 5-point Likert scale ranging from “Not Important-Very Important.”

[Figure 2. Example Summary of Round 1 Data near here]

In Section 2, 14 newly identified essential practices were introduced, which were uncovered from the panel's responses to the open-ended questions in Round 1. Panel members were asked to evaluate each practice and its description, offering feedback by rating their level of agreement on a 5-point Likert scale ranging from "Not Important" to "Very Important."

Finally, in Section 3, the comprehensive responses of the Delphi panel to the open-ended questions in Round 1 were presented, resulting in a list of 11 attributes that distinguish integrated STEM education curricula from a single discipline curriculum. Panel members were asked to rate the importance of these 11 attributes in delineating integrated STEM education from a single discipline curriculum using a 5-point Likert scale, ranging from "Not Important" to "Very Important."

Round 2 Analysis

In the second round of the modified Delphi study, 10 out of the initial 13 experts who participated in Round 1 completed the Round 2 survey. The attrition rate of three experts may have been influenced by several unknown participant loss factors (Gamst et al., 2008), including

Paper Under Review

time constraints, loss of interest, or personal reasons. For instance, March is the month when many institutes of higher learning schedule spring breaks, which may have affected the availability and engagement of the experts during this period.

Survey data from Round 2 underwent analysis using the same basic descriptive statistical methods applied in Round 1. However, to achieve a consensus, the cut score criterion was raised to a mean response score of ≥4 and a mode response score of ≥4, warranting that over two-thirds of the expert panel rated the practice as either "important" or "very important" (Kloser, 2014). In Section 1, a consensus was reached on seven out of the eight provided practices from the literature (see Table 2). In Section 2, the panelists evaluated and rated the 14 newly identified essential practices. A consensus was reached on all the practices, apart from "teacher educators leverage informal learning opportunities so teacher candidates can practice in a low-stakes environment" (see Table 3). In Section 3, a consensus was reached on all 11 attributes that distinguish integrated STEM education curricula from a single discipline curriculum (see Table 4). Finally, the Delphi panel's illustrative comments and suggestions from Section 2 were reviewed and coded by the researchers, providing deeper insight to the descriptive statistical data. As in Round 1, the merging of both the quantitative data and the qualitative data informed the structure of Round 3 (see Figure 1).

[Table 2. Descriptive Statistics of Essential Practices, Round 2, Section 1 near here]

[Table 3. Descriptive Statistics of Essential Practices, Round 2, Section 2 near here]

[Table 4. Descriptive Statistics of Delineating Attributes Round 2, Section 3 near here]

Paper Under Review

Round 3

The distribution of the final round of this modified Delphi study took place in the second week of April 2023 and remained open for 3 weeks. The primary purpose of Round 3 was to determine a definitive rank order of the expert consensus concerning the identified defining characteristics and attributes, collected during Round 1 and 2. By comparing the descriptions of the essential practices identified in the literature with the practices inspired by the panel, several repetitive and similar elements were revealed. These repetitions were simplified to provide Round 3 with seven essential practices that were originally identified in literature, and nine essential practices uncovered by the panel.

While all identified essential practices were deemed valuable, based on the Delphi panel's assessment, the panelists were informed that ranking the practices may assist in prioritizing the practices in integrated STEM teacher education. The panelists were requested to rank the practices in Section 1 and 2 by assigning a ranking of 1 to indicate the attribute of greatest importance, followed by other attributes in descending order of importance.

Ranking the importance of the delineating attributes that set integrated STEM education curricula apart from a single discipline curriculum was an equally important focus of Round 3. Following Round 2, several repetitive and similar attributes that set integrated STEM education curricula apart from a single discipline curriculum were revealed. These repetitions were simplified to provide Round 3 with a list of 9 delineating attributes. In Section 3, the panelists were asked to rank the attributes that distinguish integrated STEM education curricula from a single-discipline curriculum. The panelists were assured that all attributes had been identified as valuable but ranking them may assist in prioritizing the attributes in integrated STEM teacher education. As in Sections 1 and 2, the panelists were requested to assign a ranking of 1 to

Paper Under Review

indicate the attribute of greatest importance, followed by other attributes in descending order of significance.

Round 3 Analysis

In the final round of the modified Delphi study, 9 out of the 10 experts who participated in Round 2 completed the Round 3 survey. Email reminders were sent to the participants once a week. One notable aspect that may have affected the participation rate is the timing of Round 3. Given that the survey was distributed during April, many institutions of higher learning typically conclude their semester's learning activities and prepare for final exams during this time.

Although basic descriptive statistics including mean and mode as a measure of central tendency, and standard deviation as a measure of variability were calculated for each round, a cut score for consensus related to mean or mode was not used because the final goal of Round 3 was to obtain a collective, ranking perspective. Instead, an Interquartile Range (IQR) score of ≤2 was selected as a descriptive statistic to provide insight into the level of agreement among the expert panel by indicating the spread of data (Buelin et al., 2016; Custer et al., 1999). A larger IQR suggests more variability in the rankings and a smaller IQR suggests greater agreement among the respondents. A smaller IQR of ≤2 suggested that the responses were more closely grouped around the mean, indicating a higher level of agreement among the experts. Conversely, a larger IQR indicated more variation in responses, suggesting less consensus. Additionally, the analysis incorporated the component of perceived importance for each item. To determine this, the items ranked on the scale were evenly divided into categories of high, medium, and low importance. Understanding that a smaller mean score suggests that the experts generally agree on the higher importance of the ranked items, as the ranking is clustered around the top positions, these

Paper Under Review

categorized values were then compared to the mean rank, providing valuable insights into the level of agreement among the experts on the relative significance of each item as shown in Tables 5, 6 and 7 (Custer et al., 1999).

[Table 5. Descriptive Statistics of Ranking of Essential Practices Round 3, Section 1 near here]

[Table 6. Descriptive Statistics of Ranking of Essential Practices Round 3, Section 2 near here]

[Table 7. Descriptive Statistics of Ranking of Delineating Attributes Round 3, Section 3 near here]

Paper Under Review

Results

The summary of the quantitative results, conveyed through descriptive statistics, is intended to offer a comprehensive understanding of the tendencies observed throughout the three rounds of the modified Delphi study. Additionally, the inclusion of both recurrent and distinct comments from the panel's qualitative data serves to complement and enrich the overall summary, providing a more in-depth perspective on the research findings (Plano Clark & Ivankova, 2016).

Essential Practices in Integrated Elementary STEM Education

Rounds 1 and 2

Following two rounds of ratings and responses, 20 essential practices teacher educators use to prepare teacher candidates to effectively teach integrated STEM in the P-6 classroom achieved a consensus, including seven provided practices identified from literature, and 13 panel suggested practices (see Tables 2 and 3). Three practices: (1) providing hands-on opportunities to make and build; (2) offering experiences in utilizing the engineering design process; (3) and fostering curiosity, creativity, and innovation reached agreement with a mean score of 5.00, a SD score of 0.00, and a mode score of 5 which were the highest scores possible.

Moreover, four major themes emerged from the qualitative analysis across both rounds including imperative, experiential learning, engineering design, and authentic. For instance, many of the panelists' comments were coded by the researchers as "imperative." A panelist emphasized that "hands-on learning by doing is the hallmark of quality integrative STEM education and is key to developing a deeper understanding." Other comments highlighted the need for teacher candidates to have opportunities for "experiential learning" to foster deeper understanding by planning, building, reflecting, and revising. A panelist explained, "teacher candidates need to build deeper understandings of integrated STEM concepts as part of the process."

Panelists also emphasized the centrality of the engineering design process to integrated STEM education and the essential role of teacher educators to place the engineering design process at the forefront of integrated STEM instruction. One panelist identified that, "the engineering design process is integrated STEM education - the design process provides the perfect vehicle for integration." Additionally, panelists emphasized the importance of authentically cultivating curiosity, creativity, and innovation in teacher candidates for them to effectively teach integrated STEM education to their future students. Another panelist explained, "fostering curiosity, creativity, and innovation skills, understandings, and perspectives is critical to STEM education."

Furthermore, 17 practices surpassed the cut score mean criterion of ≥4 and a mode response score of ≥4, supporting that over two-thirds of the expert panel rated the practice as either "important" or "very important" (Kloser, 2014). For example, "solve real-world problems

Paper Under Review

related to helping others or society" reached a consensus with a mean score of 4.90, a SD score of 0.30, and a mode score of 5. Panelists' comments included: (1) "real-world problems allow people to see the applicability of STEM;" (2) "empathy is also an important component to give purpose to a design challenge;" and (3) "connecting to the real world is important for students to find value in what they are learning and doing. Solving real-world problems makes STEM education more authentic and meaningful."

In another example, "draw meaningful connections between STEM disciplines and children's literature" also exceeded the cut score with a mean score of 4.60, a SD score of 0.49, and a mode score of 5. One expert noted, "connections to children's literature are highly valuable, particularly since reading and literature are so central to what occurs in elementary schools. Making connections through literature helps to build and reinforce connections to things that teachers are familiar with." Another echoed, "children's literature helps ground engineering design - developing empathy for characters - providing purpose for design - and providing a comfortable setting - i.e., children's literature."

Finally, panelist comments also informed by indicating an explanation as to why a practice did not achieve a consensus. For example, "align STEM standards with educational textbooks" did not reach a consensus with a mean score of 2.40, a SD score of 1.36, and a mode score of 2. Panelists described textbooks as a supplemental source and encouraged teachers to look beyond textbooks for meaningful integrated STEM experiences for their students. A panelist explained rating this practice as "not important" by stating:

> Educational textbooks make me think of canned curricula products and I feel very strongly that this is not good practice. I believe that STEM teacher educators should help teacher candidates understand the importance of being able to develop their own work

Paper Under Review

that is aligned to content standards and meaningful in the context of their school and community.

Round 3

Ultimately, the primary purpose of Round 3 was to determine a definitive rank order of the expert consensus concerning the defining characteristics collected during Rounds 1 and 2. In Sections 1 and 2 of Round 3 panelists ranked the identified practices in order of importance, assigning a ranking of 1 for the greatest importance. A smaller mean score suggested that the panelists generally agreed on the higher importance of the ranked item. The additional categorized descriptive statistical values were then compared to the mean rank, providing insights into the level of agreement among the experts and the relative significance of each item. A consensus ranking was achieved if the Interquartile Range (IQR) score was ≤2 and there was agreement in the perceived importance (see Tables 5 and 6).

In Sections 1 and 2 of Round 3, a mere 38% of the 16 essential practices identified in Round 2 managed to reach a consensus ranking. For instance, among the three essential practices from Round 2 that obtained the highest possible agreement scores with a mean score of 5.00, a SD score of 0.00, and a mode score of 5, only two of the three achieved a consensus ranking in Round 3. "Providing hands-on opportunities to make and build" and "offering experiences in utilizing the engineering design process" received an IQR rating of 2 and obtained a high perceived importance ranking. During earlier rounds of the study, panelists highlighted the importance of acknowledging the interconnection between science, creativity, and curiosity, as well as the necessity for teacher candidates to gain experiences in cultivating these traits to foster them in their future students. However, despite these supporting comments, the ranking of the essential practice, "fostering curiosity, creativity, and innovation" obtained an IQR rating of

Paper Under Review

4, indicating a significant variation in responses. Additionally, the practice received only a medium perceived importance ranking, and did not reach consensus.

Additionally, “draw meaningful connections between STEM disciplines and children's literature” achieved a consensus in Round 2 by exceeding the cut score with a mean score of 4.60, a SD score of 0.49 and a mode score of 5. Many comments in Round 2 provided a deeper understanding by identifying the importance of launching integrated elementary STEM with children’s literature as opposed to beginning with commercial curricula and textbooks. A panelist added, framing integrated STEM through children’s literature is “a great practice. One of the best ways to integrate reading/comprehension and STEM.” In contrast, in Round 3, the practice of "draw meaningful connections between STEM disciplines and children's literature" also failed to achieve a consensus in rank ordering. The essential practice showed a lack of agreement with an IQR score of 4, a SD score of 2.51, a medium perceived importance position, and a bimodal score of 4 and 9.

Similar to the previous examples, these findings did not translate into a consensus ranking in Round 3. Despite its identified importance, the practice of teacher educators “using a variety of resources from different disciplines to connect content and help build purposeful understanding” did not achieve ranking consensus. It indicated a lack of agreement with an IQR score of 5, a SD score of 2.40, a medium perceived importance position, and a mode score of 3.

Providing clarity, the qualitative comments offered by the panelists presented valuable insights that balanced the quantitative data in Sections 1 and 2 of Round 3. While the quantitative analysis conveyed a numerical representation of the rankings and agreement levels among the essential practices, the qualitative comments elucidated the thought processes, perspectives, and considerations of the experts involved in the modified Delphi study.

Paper Under Review

Furthermore, the insights gleaned from the panelists' comments underscored the absence of a definitive consensus. Panelists poignantly noted reaching a decision was a challenging task due to the inherent significance of all the identified practices and attributes. One panelist specifically remarked, "this was difficult. If I did this again, they would probably be in a different order—they are all important." Additional remarks echoed this sentiment, illuminating the struggle to prioritize one essential practice over another, given their collective importance. These comments highlighted the challenges and complexities involved in ranking these essential practices, offering a deeper understanding of the decision-making process and the perspectives of the experts.

Delineating Attributes that Set Integrated STEM Education Curricula Apart from a Single Discipline Curriculum

Rounds 1 and 2

Following Rounds 1 and 2, the panel successfully reached consensus and identified 11 delineating attributes that distinguish integrated STEM education curricula from a single-discipline curriculum. These attributes were informed from the collective Delphi's responses to the open-ended question posed in Round 1 (see Table 4). The attributes ranged from integrated, hands-on teaching and learning with real-world connections to the centrality of the engineering design process. Input from the panel helped define and explain that quality integrative STEM curricula may meaningfully contribute to students' connectiveness to education by "building students' ownership of their learning which is the ultimate goal of all schooling." Panelists asserted STEM curricula is most often incorrectly viewed as either "just science," or a category of careers with STEM education simply being the list of the classes necessary for graduation or

Paper Under Review

employment. Instead, one panelist asserted, “I see it as an educational philosophy with aligned pedagogy.”

In addition, the four major qualitative themes identified in the essential practices of elementary STEM education (imperative, experiential learning, engineering design, and authentic) were also identified as pertaining to the delineating attributes of STEM education curricula.   The theme of imperative was identified as panelists underscored the significance of curricula targeted towards student curiosity, interest, and involvement, emphasizing that when students are genuinely engaged, they perceive STEM subjects and future possibilities in the field as positive and attainable rather than daunting and laborious. A panelist insisted, “it is extremely important to begin this process at the youngest age. All students can think about solving problems.”

The theme of imperative was also identified in the delineating attribute of “empowering students to take ownership of their learning by using content knowledge to creatively solve problems and challenges.” This attribute reached consensus with a mean score of 5.00, a SD score of 0.00, and a mode score of 5 which were the highest scores possible. A panelist explained, “knowledge is power for those who are becoming teachers. The opportunity to practice this brings credibility to how this kind of teaching works.”

Additionally, ten other delineating attributes were coded as imperative and surpassed the cut score mean criterion of ≥ 4 and a mode response score of ≥ 4, warranting that over two-thirds of the expert panel rated the attributes as either “important” or “very important” (Kloser, 2014). For example, the attribute delineating integrated STEM education as “providing experiential learning and problem solving that gives students opportunities to build using a variety of tools, materials, and processes” reached a consensus by exceeding the cut score with a mean score of

Paper Under Review

4.90, a SD score of 0.30 and a mode score of 5. Also recognizing that the experiential component is a key factor of integrated STEM curricula, panelists described tools, materials, and processes as the foundation of this pedagogy.

Round 3

The primary objective of Round 3, Section 3 was to establish a conclusive rank order based on the expert consensus of the 9 attributes that distinguish integrated STEM education curricula from a single-discipline curriculum. The panelists were instructed to rank the previously identified attributes based on their perceived importance, where a ranking of 1 represented the highest level of importance. Although a lower mean score indicated greater agreement among the panelists regarding the importance of a ranked item, additional descriptive statistical values were categorized and compared with the mean rank, illustrating both the level of agreement among the experts and the relative significance of each attribute. Specifically, a consensus was reached if the Interquartile Range (IQR) score was ≤2, accompanied by agreement in the perceived importance (see Table 7).

In Round 3, only one-third of the 9 panel-identified attributes that distinguish integrated STEM education curricula from a single-discipline curriculum achieved a consensus. Interestingly, none of the three attributes that obtained a consensus were perceived to be of high importance. “Empowering students to take ownership of their learning by using content knowledge to creatively solve problems and challenges” and “intentionally allowing students to see the world as a connected entity by mimicking the real world where disciplinary concepts are interconnected, facilitating natural and authentic teaching, and learning experiences for students” achieved a ranking consensus with a medium level of perceived importance. “Helping students

Paper Under Review

believe future STEM careers are achievable" achieved a ranking consensus with a low level of perceived importance.

Although the attribute "empowering students to take ownership of their learning by using content knowledge to creatively solve problems and challenges" reached a consensus with the highest possible scores in Round 2, this attribute did not achieve a rank order consensus in Round 3 with an IQR score of 3.00, indicating increased variation in the level of agreement. This was further corroborated by the mean score of 2.67 and a SD score of 1.56.

Additionally, this lack of congruency was repeated for five other attributes, including "providing experiential learning and problem solving that gives students opportunities to build using a variety of tools, materials, and processes." This attribute reached a consensus in Round 2 by exceeding the cut score with a mean score of 4.90, a SD score of 0.30 and a mode score of 5. In Round 3, the same attribute did not reach a ranking consensus with an IQR score of 3.00, a mean score of 4.22, and a SD score of 2.10.

Insights into the panelists' decisions about the attribute ranking offered a deeper understanding. Whereas the quantitative analysis conveyed a numerical representation of the rankings of the attributes, the qualitative comments expounded on the perceptions and ranking choices of the experts in the study. Panelist comments drew attention to the problem of identifying a clear ranking consensus. For example, panelists indicated the difficulty in selecting one attribute to be more important than another, indicating that all 9 attributes were important. One panelist succinctly explained, "it's very hard to rank these. So many of them, in my opinion, are interrelated." These remarks provided valuable insights into the decision-making process and the perspectives of the experts, underscoring the challenges and complexities involved in ranking the 9 attributes.

Paper Under Review

Discussion

The purpose of this modified Delphi study was to leverage the experiential and theoretical insights of an expert panel composed of elementary teacher educators who teach methods of integrated STEM (science, technology, engineering, and mathematics) to teacher candidates, with intention of contributing to the growing body of literature concerning integrated STEM education (Gonzalez & Kuenzi, 2012; Kelley & Knowles, 2016; Nowikowski, 2017; Rinke et al., 2016; & Rose et al., 2017). This study intended to benefit from the knowledge of the expert panelists by attaining (1) a consensus regarding essential practices teacher educators utilize to help prepare teacher candidates to effectively teach and implement integrated STEM in the P-6 classroom, and by attaining (2) a consensus of attributes that delineate the core attributes of integrated STEM education compared to a single discipline curriculum.

Ultimately, the panel of experts identified 16 essential practices teacher educators use to prepare teacher candidates to effectively teach integrated STEM in the P-6 classroom and 9 delineating attributes that set integrated STEM education curricula apart from a single discipline curriculum. While the quantitative analysis offered a numerical representation of the rankings and agreement levels among the essential practices and delineating attributes, the qualitative comments revealed the thought processes, perspectives, and considerations of the experts uncovering four major themes of integrated STEM education including imperative, experiential learning, engineering design, and authentic. Additionally, the panelists provided a common vocabulary and a workable list of essential practices and attributes that set integrated STEM education apart, offering valuable insights for educators, researchers, and policymakers.

Conclusion

While the significance of elementary STEM teachers' pedagogy has been acknowledged as a vital undertaking, stemming from focused teacher candidate preparation (Gonzalez &

Paper Under Review

Kuenzi, 2012), few higher education institutions indicate they provide coursework or learning experiences designed to deliver meaningful education opportunities which support STEM integration experiences within P-6 classrooms (Nesmith, et al., 2017; Rinke et al., 2016; Rose et al., 2017). This may trigger speculation that the institutions without integrated elementary STEM education programs may be attributing to the phenomenon of the missing talent in STEM.

The mixing of the quantitative and qualitative methods provided deeper understanding at the intersection when the results from both study strands were interpreted together (Plano Clark et al., 2016). It was projected by the researchers that the ranking process would help to prioritize and highlight the most significant and influential practices and attributes uncovered by the panel, delivering a clearer understanding of the relative importance of each essential practice and delineating attribute. In fact, the opposite occurred which added unexpected depth and structure to the data and enabled the researchers to draw more reliable conclusions. Each identified practice and attribute appear to hold a high degree of importance, making it nearly impossible for the expert panelists to prioritize one over another.

Despite recognizing their significance, the panelists' difficulties in ranking the practices indicated that they may hold equal importance in the context of integrated STEM teacher education. This insight is crucial to understanding why some practices did not reach a consensus ranking in Round 3 despite strong quantitative consensus scores in the previous rounds. The lack of a clear rank order coupled with the qualitative comments expressing the difficulty in ranking the practices and attributes emphasizes the complexity and richness of the data, with each identified practice and attribute interdependently contributing to the overall understanding of P-6 integrated STEM education. Within traditional curricula content areas such as science, technology, engineering, and math the subjects are viewed as independent practices, having

Paper Under Review

impassable walls that separate each individual subject (Bybee, 2013). Integrated STEM education transcends the confines of a siloed subject hierarchy, embracing instead the interconnected interplay among science, technology, engineering, and mathematics. Likewise, the study's findings, which revealed a challenge in achieving a definitive ranking order consensus, may similarly be attributed to the integral and collective significance of all 16 identified practices and the 9 delineating attributes recognized by the expert panel in the Delphi process. Attempting to impose a strict ranking hierarchy may overlook their interconnectedness and combined impact, just as this interconnectedness stands as the foundational cornerstone of integrated STEM education (Kelley & Knowles, 2016).

In addition, further research is needed to optimize the results of this study by exploring ways to transform the essential practices and delineating attributes of P-6 integrated STEM education employed in teacher education to into actionable methodologies for practitioners within P-6 classrooms. Additional research would also offer practitioners the chance not only to familiarize themselves with the identified essential practices and attributes, but also to provide the practitioners with strategies to foster their own confidence in successfully implementing these elements into their instructional practices (Herro et al., 2019). Moreover, this list of identified essential practices and delineating attributes of P-6 integrated STEM education is not definitive and continued discourse, analysis, and research may benefit both educators and researchers.

The validity of this study hinges not on the size of the population sample, but rather on the significance of the Delphi methodology and the adeptness of the panel's expertise. This underscores the necessity to assemble a panel comprising of individuals possessing complex levels of prowess and comprehension in both theoretical knowledge and practical application.

Paper Under Review

Through collaborative and iterative rounds, the expert panelists were afforded opportunities to leverage their existing theoretical knowledge, extensive academic experience, and the relevant literature to assess, rate, and elucidate the essential practices and delineating attributes of P-6 integrated STEM education. The results may offer invaluable guidance for researchers, practitioners, and policymakers striving to design effective programing to impact the missing talent in the STEM pipeline.

Identifying the interconnectedness of the 16 essential practices teacher educators use to prepare teacher candidates to effectively teach integrated STEM in the P-6 classroom, and the 9 delineating attributes that set integrated STEM education curricula apart from a single discipline curriculum may be a valuable piece of the framework puzzle to combat the current and persistent unchanged pattern of students' involvement in STEM.

References

Alexander, C., Knezek, G., Christensen, R., Tyler-Wood, T., & Bull, G. (2014). The impact of project-based learning on pre-service teachers' technology attitudes and skills. *Journal of Computers in Mathematics and Science Teaching, 33*(3), 257-282.

Archer, L., Moote, J., MacLeod, E., Francis, B., & DeWitt, J. (2020). ASPIRES 2: *Young people's science and career aspirations*.

Avella, J. R. (2016). Delphi panels: Research design, procedures, advantages, and challenges. *International Journal of Doctoral Studies*, *11*, 305.

Berlin, D. F., & White, A. L. (2012). A longitudinal look at attitudes and perceptions related to the integration of mathematics, science, and technology education. *School Science & Mathematics, 11*2(1), 20-30

Paper Under Review

Paper Under Review

Buelin, J., Clark, A. C., & Ernst, J. V. (2016). Engineering's grand challenges: Priorities and integration recommendations for technology education curriculum development. *Journal of Technology Education, 28*(1), 37-52.

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

Clayton, M. (1997). Delphi: a technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, *17*(4), 373-386.

Collins, K. M. T., Onwuegbuzie, A. J., & Johnson, R. B. (2012). Securing a place at the table: A review and extension of legitimation criteria for the conduct of mixed research. *American Behavioral Scientist, 56*(6), 849–865.

Cresswell, J.W. (2013). *Qualitative inquiry and research design* (3rd ed.). Thousand Oaks, California, Sage.

Custer, R. L., Scarcella, J. A., & Stewart, B. R. (1999). The modified Delphi technique -a rotational modification. *Journal of Vocational and Technical Education, 15*(2), 50–58.

Daugherty, M, Carter, V., & Sumner, A. (2021). Standards for technological engineering literacy and STEM education. *Technology & Engineering Teacher, 76*(1), 32-37.

Eckman, E., Williams, M., & Silver-Thorn, M., (2016). An integrated model for STEM teacher preparation: The value of a teaching cooperative educational experience. *Journal of STEM Teacher Education*, *51*(1), 71-81.

Emrey-Arras, M. (2018). *Science, technology, engineering, and mathematics education: Actions needed to better assess the federal investment*. United States Government Accountability Office.

Gamst, G., Meyers, L., & Guarino, A. J. (2008). *Analysis of Variance Designs*. New York, New York: Cambridge University Press.

Glass, G., & Hopkins, K. (1996). *Statistical Methods in Education and Psychology.* Needham Heights, Massachusetts: Simon & Schuster.

Green, R. (2014). The Delphi technique in educational research. *Sage Open, 4*(2), 2158244014529773.

Greene, J. C. (1987). Stakeholder participation in evaluation design: Is it worth the effort?. *Evaluation and Program Planning*, 10(4), 379-394.

Gonzalez, H. B. & Kuenzi, J. J. (2012). *Science, technology, engineering, and mathematics (STEM) education: A primer*. Washington, DC: Congressional Research Service, Library of Congress.

Herro, D., Quigley, C., & Cian, H. (2019). The challenges of STEAM instruction: Lessons from the field. *Action in Teacher Educations, 41*(2), 172-190.

Hodson, D. (2003). Time for action: science education for an alternative future. *International Journal of Science Education*, *25*(6), 645-670.

Hsu, C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical Assessment Research & Evaluation, 12*(10), 1-8.

Hutchison, L. F. (2012). Addressing the STEM teacher shortage in American schools: Ways to recruit and retain effective STEM teachers. *Action in Teacher Education, 34*(5-6), 541-550.

International Technology and Engineering Educators Association (ITEEA). (2020). *Standards for technological and engineering literacy: The role of technology and engineering in STEM education.* Reston, VA: Author.

Paper Under Review

International Technology Education Association (ITEA). (2000). *Standards for technological literacy: Content for the study of technology.* Reston, VA: Author.

Kelley, T. & Knowles, G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(11), 6-11.

Kloser, M. (2014). Identifying a core set of science teaching practices: A Delphi expert panel approach. *Journal of Research in Science Teaching, 51*(9), 1185-1217.

Koster, B., Brekelmans, M., Korthagen, F., & Wubbels, T. (2005). Quality requirements for teacher educators. *Teaching and Teacher Education, 21*(2), 157-176.

National Academy of Engineering (NAE) & National Research Council (NRC). (2014). *STEM integration in P-12 education: Status, prospects, and an agenda for research*. Washington, DC: The National Academies Press.

National Research Council (NRC). (2012). A framework for P-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a conceptual framework for new P-12 science education standards. *Board on Science Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.

National Science Foundation (NSF). (2018). *Mathematics and science education: Enrollment in postsecondary education*. Arlington, VA. https://www.nsf.gov/statistics/2018/nsb20181/report/sections/elementary-and-secondary-mathematics-and-science-education/transition-to-higher-education#enrollment-in-postsecondary-education

National Science and Technology Council (NSTC). (2012). Coordinating federal science, technology, engineering, and mathematics (STEM) education investments: Progress report. *A Report from the Federal Coordination in STEM Education Task Force*

Paper Under Review

Committee on STEM Education. Washington, DC: Congressional Research Service, Library of Congress.

National Science and Technology Council (NSTC). (2021). *Progress Report on the Implementation of the Federal STEM Education Plan*. Washington, DC: Congressional Research Service, Library of Congress.

Nesmith, S., Purdum-Cassidy, B., Cooper, S. & Rogers, R. (2017). Love it, like it, or leave it- elementary preservice teachers' field-based perspectives toward the integration of literature in mathematics. *Action in Teacher Education, 39*(3), 321-339,

NGSS Lead States. (2013). Next Generation Science Standards: For State; by States. In *Next Generation Science Standards: For State; by States: Vol. Appendix J*. National Academies Press.

Nowikowski, S. H. (2017). Successful with STEM? A qualitative case study of pre-service teacher perceptions. *The Qualitative Report, 22*(9), 2312-2333.

Okoli, C. & Pawolowski, S. (2004). The delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15-29.

Plano Clark, V. & Ivankova, N. (2016). *Mixed methods research: a guide to the field*. Thousand Oaks, CA: Sage.

Rinke, C. R., Gladstone-Brown, W., Kinlaw, C. R., & Cappiello, J. (2016). Characterizing STEM teacher education: Affordances and constraints of explicit STEM preparation for elementary teachers. *School Science and Mathematics, 116*(6), 300-309.

Ro, H. K., & Knight, D. B. (2016). Gender differences in learning outcomes from the college experiences of engineering students. *Journal of Engineering Education, 105*(3), 478-507

Paper Under Review

Rose, M. A., Carter, V., Brown, J., & Shumway, S. (2017). Status of elementary teacher development: Preparing elementary teachers to deliver technology and engineering experiences. *Journal of Technology Education*, *28*(2), 2–18.

Saldana, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage.

U.S. Bureau of Labor Statistics. (2022). Employment in STEM occupations. Employment projections. Table 1.11. *Employment in STEM Occupations, 2020 and projected 2030*. https://www.bls.gov/emp/tables/stem-employment.htm#2

Paper Under Review