




International Journal of Educational Methodology

Volume 8, Issue 1, 11 - 27.

ISSN: 2469-9632
<https://www.ijem.com/>

Sometimes Finding Nothing is Something: Shrinking the Gap between Emerging Bilingual Learners and English Fluent Students (Case in Point)

Michael W. Corrigan* 
Marshall University, USA

Douglas Grove 
Concordia International School Hanoi, VIETNAM

Sage Andersen 
University of Texas at Austin, USA

Joseph T. Wong 
University of California at Irvine, USA

Bradley S. Hughes 
University of California at Irvine, USA

Received: September 28, 2021 • Revised: December 17, 2021 • Accepted: January 21, 2022

Abstract: For United States of America (USA) and other developed countries, science achievement gaps begin to emerge in elementary and primary school. Such gaps between USA student groups typically are connected to socio-economic status (SES) and issues such as students still learning the English language. Through an experimental design, this National Science Foundation funded study explores how integrating the arts into science, technology, engineering, and mathematics (STEM) curriculum and leading with a more STEAM-first approach (e.g., curriculum which integrates science, technology, engineering, arts, and mathematics) might provide more equitable science learning opportunities for elementary or primary grade level students. More specifically, the project's research efforts seek to also examine how integrating the arts into science instruction might help emerging bilingual (EB) students who are simultaneously learning the English language and science. Although results provide somewhat conflicting findings of statistical significance with small to moderate effect sizes, outcomes provide initial evidence that leading with STEAM science instruction before STEM efforts can be beneficial to early readers, and for EB students this benefit is magnified. As the title of this study suggest, sometimes finding nothing is something.

Keywords: *Emerging bilingual, instruction, science, STEAM, STEM.*

To cite this article: Corrigan, M. W., Grove, D., Andersen, S., Wong, J. T., & Hughes, B. S. (2022). Sometimes finding nothing is something: Shrinking the gap between emerging bilingual learners and English fluent students (Case in point). *International Journal of Educational Methodology*, 8(1), 11-27. <https://doi.org/10.12973/ijem.8.1.11>

Introduction

Part of a large-scale longitudinal National Science Foundation (NSF) funded research initiative in the United States of America (USA), this study investigates the efficacy of integrating the arts into instruction of next generation science standards (NGSS, 2013). Through an experimental design, the study explores how integrating the arts into science, technology, engineering, and mathematics (STEM) curriculum and leading with a more STEAM-first approach (e.g., curriculum which integrates science, technology, engineering, arts, and mathematics) might provide more equitable science learning opportunities for elementary or primary grade level students. More specifically, the project's research efforts seek to examine how integrating the arts into science instruction might help younger challenged readers such as emerging bilingual (EB) students who are simultaneously learning the English language and science.

Please note, historically in the USA, emerging bilingual (EB) students in the process of learning the English language in addition to one or more other languages, have typically been referred to as English language learners (ELL) or English learners (EL). To some, and rightly so, ELL or EL is a deficit-oriented label with a negative connotation. A growing number of educators, however, have seen the need for replacing this wording with a more asset-oriented terminology that does not presume the need for English fluency to authentically engage in science (González-Howard & Suárez, 2021; Poza, 2018; Suárez, 2020; Ünsal et al., 2018; Wilmes & Siry, 2020). Thus, while sometimes using the EL or ELL acronyms in the literature review to represent more accurately what others have reported on in their research, in this paper (whenever

*Corresponding author:

Michael W. Corrigan, College of Education and Professional Development, Education Building, One John Marshall Drive, Huntington, USA.
✉ corrigan@marshall.edu

possible) we refer to such students as emerging bilingual (EB) learners or students. Because we acknowledge the challenge and accomplishment of learning more than one language and recognize 'bilingualism' as an asset.

This specific study sought to test if leading with a STEAM approach to earth science instruction could better benefit emerging bilingual (EB) learners (and their often equally English reading challenged lower-socio-economic status (SES) classmates), more than relying upon or specifically starting with a STEM inquiry-based learning approach to science. This study sought to test a different approach to teaching elementary students the *big* words, somewhat intimidating terminology, and often peculiar conundrums associated with science. The research team for this initiative set out to bring science to life for elementary students, by pulling the hard-to-understand concepts and words out of textbook pages, and using the arts through dance, song, digital media, design, poem, or performance to more clearly illuminate such ideas via a more social and cultural focused language all can better understand.

A more fitting title might have been, "*how can you learn science when you're still learning the language*". The study explores how language challenges (e.g., whether learners were fluent or not in the language of the textbook and what the teacher spoke) might be one of the more likely reasons behind why so many EB students perform lower than English fluent (EF) students. We wanted to test what might help such students find a bypass around this language barrier.

As a result, the study itself involved providing teachers with extensive professional development training to utilize both NGSS STEM and STEAM approaches for implementing a fourth grade NGSS earth science curriculum via a novel arts-enhanced teaching method. By supporting educators in providing students equally designed and effective lessons via NGSS STEM and STEAM instruction, our intent was to better ensure the efficacy of the STEAM-first and STEM-first approaches to assess order effects possibly emanating from the differing sequence of instructional approaches received.

The findings to follow provide insights as to how the arts could be an answer to improving instruction with children still learning to read the language with which their books are written, and their teachers teach. And for EB students worldwide, as well as their lower-SES counterparts in need of such efforts to better help them get an equitable education, we present the following insights, efforts, and findings.

Literature Review

The Importance of Science Instruction

Historically, and like many other countries, educating students in the USA in science has been and continues to be a national priority for both social and economic reasons (Bybee, 2014; Olson & Riordan, 2012). Learning science begins in the elementary or primary grades, which is a critical timeframe for students as they are introduced to foundational and crosscutting concepts necessary for later successes in science achievement (National Research Council, 2012). Previous research has established science content knowledge as a predictor of an individual's future success, and much of the USA economy depends on a science-educated population (Olson & Riordan, 2012).

Despite numerous calls for and efforts implemented to improve science education policy and curricula since at least the 1960's (Brown et al., 2012), however, the USA has trailed other developed nations in science-related skills for several decades (Corrigan et al., 2013; National Center for Education Statistics [NCES], 2000). Such shortcomings started to become evident in 2003 when the USA science literacy scores were below average when compared internationally (Lemke et al., 2004). Then, in 2006, the USA placed 23rd out of 57 industrialised and developing countries on science literacy according to the Organization for Economic Cooperation and Development's (OECD) Programme of International Student Assessment (PISA) (NCES, 2007).

The PISA is a test given every three years to students at age 15 (OECD, 2007a). In a recent round of PISA results (OECD, 2019), the USA fell further to a 25th ranking in science. The trend of declining performance in science is a cause for concern, and with Next Generation Science Standards movement in the USA stepping up more rigorous approaches to science instruction, there is a paramount need for more experimentation in innovations to improve science learning.

Furthermore, in the early grades is when science achievement gaps begin to emerge – both between the USA and other developed countries, as well as between student groups within the USA (Provasnik et al., 2012). Such gaps between USA student groups typically are connected to socio-economic status (SES) and issues such as students still learning the English language (Ormrod, 2011). There is an apparent qualitative challenge for young students trying to learn science. The text and new vocabulary involved with such science instruction often requires students to have proficient levels of reading comprehension. And for young students, and more specifically emerging bilingual (EB) students in the elementary or primary grades, reading comprehension is often a challenge.

Evidence suggests, however, that arts-based teaching methods are designed to decrease cognitive load by making abstract concepts more concrete and accessible (Fillmore, 2007). The possibility exists that incorporating the arts into STEM initiatives (i.e., STEAM) holds the potential for decreasing or shrinking the achievement gap. With the number of EB students rising in schools across the USA, there also is a need for exploring evidence-based approaches to elementary science instruction that may have the potential to boost science knowledge for all children regardless of the English language fluency level.

Addressing Learning Challenges for Emerging Bilingual Students

The achievement of English language learners in general continues to be low and this same challenge continues within STEM education (Bravo & Cervetti, 2014; Huerta & Jackson, 2010). Findings from the National Assessment of Educational Progress (NAEP), a long-standing standardized testing effort in the USA, reveal that EL/EB students score lower at all grade levels and are more likely to score below basic proficiencies (NCES, 2014). The need for support in STEM education is greater for EL students than for the non-EL students (Goldenberg, 2013). As a result, without research exploring and identifying the promise of more structured and strategic instructional methods, the Next Generation Science Standards most likely will continue to be a challenge to these EB students, as the bar rises for the usage of scientific language.

The STEM versus STEAM Debate

Efforts for STEM and STEAM are global initiatives. Schools and universities throughout the United States and other countries, including China, Australia, the United Kingdom, France, and Taiwan, are implementing STEM and STEAM - focused curricula (Gess, 2017; Kelley & Knowles, 2016). In many countries STEM and STEAM embody educational approaches that embed hope for improving the quality of life through financial prosperity and opportunity by preparing the citizenry to compete in the marketplace of ideas where engineering, science, technology, and mathematics drive economies (Gess, 2017).

To some researchers and practitioners, however, a debate still exists on the differences between STEM and STEAM. At first glance, most would consider a difference between the two simply being the integration of the arts, but the hypothesized difference continues to be somewhat of a polarizing topic (Barcelona, 2014). Some taking a STEM only approach argue that the integration of s is not necessary as long as we just focus on the teaching and assessment of good mathematics and science standards (May, 2015). According to others, integrating the s while attempting to improve the instruction of STEM dilutes the time needed to focus on the STEM concepts (Heilig et al., 2010).

Proponents of the arts, however, suggest that integrating arts with STEM efforts (i.e., STEAM) may play a supportive role in science learning (Catterall, 2009; Donovan & Pascale, 2012; Guyotte et al., 2014; Hardiman et al., 2009). Some STEAM proponents contend the arts are necessary to increase both the creative thought process and allow students to be able to access the concepts more easily in STEM (Ghanbari, 2015). The STEAM paradigm also emphasizes the importance of STEM, noting that the arts can transcend across different disciplines and enrich learning in more than just art (Hetland et al., 2015). Utilizing STEAM can be considered as an educational teaching and learning approach.

According to Hetland's et al. (2015), integrating the arts provides a transdisciplinary epistemology to delivering STEM disciplines. arts-integrated literacy instruction helps students increase reading comprehension and writing skills (Podlozny, 2000; Walker et al., 2011), with all students benefiting, especially low-SES and EL students (Ingram & Riedel, 2003). STEAM methods aid in scientific language development by decreasing cognitive load and making abstract concepts more concrete and accessible through multimodality and embodied representation (Campbell et al., 2016; Wahyuningsih et al., 2020).

Further augmenting science instruction with the arts can enhance students' inquiry skills, problem solving skills, and creative thinking (Segarra et al., 2018), and perhaps to make science education more equitable for EB students (Dewey, 2005; González-Howard & Suárez, 2021; Hadzigeorgiou, 2016; Lee et al., 2019). Through such efforts, the STEM to STEAM movement can offer new insights and new vocabulary in interdisciplinary thinking (Madden et al., 2013).

Additional Logic and Evidence for Putting the A in STEM

There is a growing body of evidence suggesting the supportive role that the arts could play within STEM education (Catterall, 2009; Daugherty, 2013; Donovan & Pascale, 2012; Guyotte et al., 2014; Hardiman et al., 2009). The learning of STEM disciplines can be better facilitated with arts integration and provide access for more students (Hwang & Taylor, 2016). Increased motivation, engagement, and achievement may result with the integration of art into a STEM curriculum (Becker & Park, 2011; Hadzigeorgiou, 2016; Liao, 2016; Schlaack & Steele, 2018).

Perhaps ever more important to science and engineering, art education has a measurable impact on students' ability to think creatively (Lichtenberg et al., 2008; Luftig, 2000; Moga et al., 2000). Creativity leads to students showing more adaptability and flexibility in their thinking (Karakelle, 2009; Mehu, et al., 2008). According to Wahyuningsih et al. (2020), STEAM methodologies are an effective pedagogical strategy with an array of evidence supporting early childhood education to improve creativity, problem-solving, scientific inquiry, critical thinking, and cognitive development.

Furthermore, theorists in cognition research indicate that the use of visual and performing arts (VAPA) strategies such as kinesthetic movement, gestures, and purposeful body expression can augment science sensemaking, comprehension, and retention (Castro-Alonso et al., 2019; Glenberg, 2011). For example, STEAM techniques incorporating the arts, such as choreography and locomotor dance movements which include levels, sliding, chasing, and climbing aligned with NGSS content such as the phases of the moon or the movement of the sun around the earth, compared to more directly teaching

from the textbook, are positively linked with elementary students' effective recall, retention, and identifying misunderstandings (Edens & Potter, 2003).

Rationale

The achievement of English learners in general continues to be low and this same challenge continues within STEM education (Bravo & Cervetti, 2014; Huerta & Jackson, 2010). Approaches that are based on the NGSS standards and utilize STEAM approaches, however, may have promise for shrinking the gap for students who are not native English speakers. With a vision toward increasing science literacy for all students, while taking advantage of the potential benefits of arts-integrated science education and simultaneously providing equitable access to the arts, this study explores how science instruction might be beneficially augmented by utilizing the arts as an equitable alternative or supplementation to STEM approaches for elementary science instruction. We posit that further integration of arts into the STEM and STEAM education movements may increase science learning for elementary students.

This study sought to determine any differences in student science knowledge and learning outcomes based upon receiving STEM or STEAM instruction and the order in which they are received and/or combined. In this study the term unit refers to a sequential set of pedagogically consistent lessons, which will be detailed later under the description of the intervention. To test the possible efficacy and order effects of STEM-first and STEAM-first unit approaches to science instruction, the following research questions (RQ) were explored:

RQ1: Are there statistically significant differences on baseline assessments of earth science knowledge (Pretest) between the NGSS STEAM-first unit and NGSS STEM-first unit cohorts, and emerging bilingual and English fluent cohorts?

RQ2: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences between NGSS STEAM-first and NGSS STEM-first cohorts (regarding results at final Post-Test 2)?

RQ3: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences between emerging bilingual and English fluent cohorts (regarding results at final Post-Test 2)?

RQ4: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences in overall science knowledge gains (i.e., change scores) obtained during the 9-week intervention?

Methodology

Research Design

Funded by the National Science Foundation, this study is part of a multi-year collaboration between a large research university, a county performing arts center, multiple school districts in California, and an external evaluator. To specifically explore research efforts into the fourth-grade earth science investigations of the project, this study utilizes a treatment crossover repeated measures (Pre/Post/Posttest) design to assess how the order of augmenting science instruction integrated with the arts may increase science learning and equitability for emerging bilingual (EB) and English fluent (EF) students. Through random assignment of the participating teachers and their specific classrooms to two groups of participants performing the same tasks in reverse order from one another, similar to previous educational research efforts (Crowder & Hand, 2017; Jones & Kenward, 1989), this study examines the effects of order in addition to changes of instructional gains over time. The two-classroom level randomly assigned cohorts implementing either a STEM-first (STEM before STEAM) or STEAM-first (STEAM before STEM)) approach eventually received both types of instructional units covering the same content. This method created a scenario, however, where the two-classroom level randomly assigned cohorts mainly differed in the implementation order and allowed us to assess whether leading with NGSS STEAM lessons provide greater benefit to students compared to leading with NGSS STEM lessons.

Sample and Data Collection

The study efforts began with the recruitment of schools in multiple districts in California. First level criteria for the schools to be included in the selection pool was based solely on whether they were designated as a Title I elementary school. Title I schools in the USA more often reflect student populations consisting of highly transient and lower income based families, as well as more diverse collections of minorities or differing ethnicities. Additionally, to avoid potential spillover effects, all fourth-grade teachers at each school had to agree to participate in the intervention for the school to be included. Nine elementary schools from six districts were selected to receive the earth science intervention, and the order of the instructional intervention (STEAM-first or STEM-first) was randomly assigned to these schools.

This effort led to the random selection and assignment of 18 classrooms across the nine Southern California elementary schools. The participants assessed in this study, who completed the 9-week intervention, included 355 fourth grade students ($n = 355$). As a result, all participating fourth grade teachers in each of the schools were randomly assigned to

teach either the NGSS STEM unit followed by the NGSS STEAM unit (i.e., STEM-first), or to teach the NGSS STEAM unit followed by the NGSS STEM unit (i.e., STEAM-first).

Randomly assigning all participating teachers at the school level, to either implement the STEAM-first or STEM-first approaches, was performed to theoretically enabled more precise estimation of the possible order effects of combining both approaches and the impact of one approach versus the other. The goal of having one cohort lead with STEAM first was to see if using dance, song, digital media, design, poetry, or performance (the Arts) could more effectively help elementary students pull the hard-to-understand concepts and words out of textbook pages, and more clearly illuminate such ideas via a more social and cultural focused language that all can better understand and bring science to life. We also wanted to test if such efforts might help students find a bypass around possible language barriers. As Figure 1 illustrates, this nine-week process began with a pretest, and after completion of the first unit' multiple lessons and a first post-test, the groups switched and were taught by the other method.

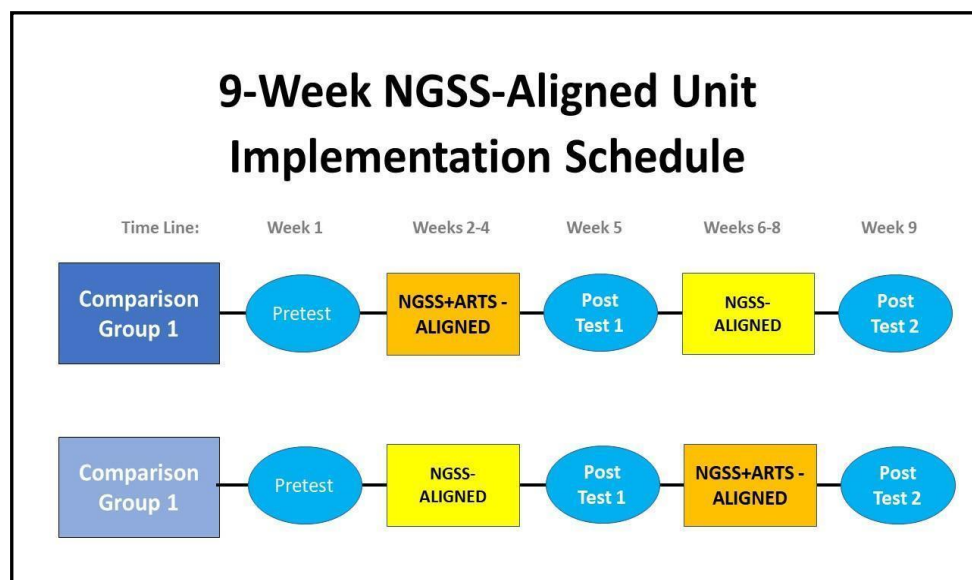


Figure 1. Experimental Classroom Implementation and Assessment Design

Instructional Intervention

As highlighted in Figure 1 above, the nine-week instructional unit consisted of three scaffolded NGSS earth science lessons being delivered twice. The three lessons were multi-pronged or multi-faceted, and with existing district mandated curricula requirements in place for teachers to simultaneously deliver, the lessons were completed weekly over a three-week span. Each of the NGSS STEM and NGSS STEAM lessons/units addressed identical scientific concepts based on grade-level NGSS performance expectations in earth science. The difference between these two approaches is the addition of the more highly developed arts augmentation the study put in place within STEAM unit lessons to replace the more traditional guided inquiry methods often found in STEM unit lessons.

Treatment frequency, quantity and duration were determined based on previous pilot interventions conducted by the authors, noting the temporal limitations of teachers balanced with the benefit of treatment levels that produce significantly measurable gains without over-teaching the concepts. Specifically, the treatment effects of the three scaffolded lesson sets, left room for further learning gains that could be tested in our combined order effect studies, in which the authors investigate the effects of combining two treatments, STEM and STEAM, effectively doubling overall treatment magnitude while measuring lesson unit order of implementation. Had each treatment been six lessons instead of three lessons, previous trials suggested that then the sensitivity of the measures could have been less capable of discerning the order effects since learning saturation could already have been reached prior to any additive treatment.

To be clear, the NGSS STEM lessons were aligned to the Next Generation Science Standards (NGSS) for earth science and used guided inquiry as the main instructional framework. The NGSS STEAM lessons addressed the same NGSS STEM performance expectations *in addition* to specific elementary level visual arts, dance and other art standards, while removing the guided inquiry components (facilitating equivalent scope and length of the STEM and STEAM lesson units). Examples of elements of art addressed in the NGSS STEAM science lessons include axial and locomotor movements, pathways, levels, and shapes in dance, as well as color, lines, shapes, and perspective in visual art. Please note that experts in the arts and art education were a part of the research team and on the staff who trained the educators.

The research team trained all participating teachers in the instructional procedures of the NSF project's NGSS lessons, balancing training between the STEM and STEAM units for gaining teacher implementation competencies and fidelity in both approaches. Accordingly, in providing instruction supportive of addressing EB's, the design of NGSS unit-based

curricula (for both the STEM and STEAM approaches) and teacher professional development (PD) were grounded in the literature on effective science teaching for ELs (Lee, 2005). These practices were, therefore, incorporated into both the NGSS STEM unit and the NGSS STEAM unit curricula to maintain the typical support methods for EB literacy constant between the two instructional intervention units, so that effects of arts integration are not confounded by such other methods. Research methods and assessment procedures were put in place to monitor and ensure the fidelity of the training and implementation.

Measures

The key outcome variable of interest for this study was to determine gains in fourth grade earth science knowledge. Considerable effort and time were dedicated to developing a research-based test of science knowledge that focused specifically on fourth grade NGSS earth science topics. Despite this measure not necessarily being a psychometric tool (e.g., mental measurement) and instead being more of a measure of knowledge (e.g., a science test), this measurement tool went through multiple design and development steps in accordance with procedures to create sound psychometric assessments, as well as pilot testing numerous iterations of the tool.

The first iterations of the fourth-grade level science knowledge tool were developed by the grant's instructional design team of science experts. The proposed test was further refined after sending the science knowledge questions to another collection of external experts to provide feedback and suggestions. As a result, face and content validity was established for this 17-item tool to be scored and utilized in the analysis, and a series of pilot tests offered further support for test-retest reliability/validation (e.g., initial construct validity) with alpha coefficients (scale reliability) consistently ranging from .70 to .73. This 17-item assessment of fourth-grade earth science knowledge, serving as the measure for the differing dependent variables (outcome variables) utilized in the following analyses, was then used repeatedly and administered at pretest, posttest1 and posttest 2.

Our independent variables (grouping variables) consisted of a measure of English fluency and instructional intervention approach. Based on English proficiency testing collected by the participating school districts, students were categorized as either emerging bilingual (EB) or English fluent (EF) learners. And in accordance with the random assignment sampling techniques above, the participating schools and their teachers were randomly assigned to either provide a STEAM-first or STEM- First approach. Thus, the instructional intervention designation was based on if a student was in a school or more specifically classroom assigned to deliver the STEAM-first or STEM-first approach. Additionally, as to be covered in more detail, the earth science assessment delivered at pretest, and the measurement of the instructional intervention's implementation fidelity were used as covariates.

Analyzing the Data

To start, preliminary checks were conducted to ensure the dataset had no violations of the assumptions of normality, outliers, linearity, homogeneity of variance, and multicollinearity. Across the numerous analyses performed, significance levels of Levene's tests were larger than .05, suggesting equal variances could be assumed. Beyond a few slightly significant higher mean scores identified at pretest between the cohorts being studied, which will be clarified shortly, issues related to linearity, outliers and multicollinearity were not identified.

The analysis was intended specifically to measure learning gains (i.e., a means analysis) related to earth science knowledge gains (shrinking the gap), language fluency and the possible use of a STEAM-first instructional approach. Therefore, to first assess where the randomly assigned student participants were performing at pretest related to earth science knowledge, we utilize independent-samples t-tests. And to assess the gains in science knowledge, while having SPSS utilize regression analysis in the background accounting for Type I and II error, we performed two separate analyses of covariance (ANCOVA).

Again, students were assessed three times, at pretest, between (posttest 1) and after implementation (posttest 2) of the STEM and STEAM units, depending on the order of implementation. The dependent variables for the differing analyses were students' pretest scores, final posttest scores as well as a change score assessing total knowledge gained from pretest to final posttest. This change score consists of increases or decreases in earth science knowledge from pretest to final posttest (i.e., Posttest 2 minus Pretest = Overall knowledge gained).

Pretest scores documenting the participants' initial knowledge of earth science also were used as covariates to statistically control for individual differences. Additionally, to assess fidelity of implementation of the intervention, we utilized implementation logs and observations collected on participating teachers during the study. Given the workload placed on teachers and knowing from past research efforts that not all teachers put forth the same effort when participating in research projects, measuring implementation fidelity was a crucial task. The goal was to account for the differences in the implementation of the intervention, and document that the intervention did take place with some level of fidelity. By monitoring the activity and implementation levels of the teachers (e.g., assessing the dosage and quality with which this study's intervention was being provided to students), we were able to measure, and then use a means-based standard deviation split as a covariate to designate which teachers documented high (1) versus moderate to low (2) implementation.

Due to the random assignment sampling procedures, the intervention's cohort groups were assigned at the teacher or classroom level (e.g., an independent grouping variable used to assign students to the STEM-first cohort or STEAM-first cohort). Given the instruction of the lessons was provided at the classroom level and not individually to each student, however, student level outcomes were used as the unit of analysis. Unlike a medical study where a child might be being assessed or tested for a specific illness or medication, as well as individual treatment procedure, we did not believe any type of cluster level analysis would be fitting. Given teachers teach to their class the curriculum, and then try helping each student individually after the delivery of the curriculum, we felt taking a wider look at how the instructional interventions impacted the classroom efforts and the collection of students in general made more sense.

Findings / Results

Again, this study was funded to explore if leading with science instruction augmented by the arts could benefit early readers and EB students and Title I school students in elementary grades. Based upon the type and order with which participants received the instructional intervention (NGSS STEAM-first unit vs. NGSS STEM-first unit), the goal was to augment science instruction with a more social-based education and arts-augmented approach to instruction and determine the possible impact on student science knowledge learning. The focus was on determining if the intervention possibly could increase science knowledge gains by leading with STEAM efforts and possibly shrink the gap in academic performance between the participants. Therefore, the following research questions (RQ) were explored utilizing the analyses mentioned above to test the impact of the instructional intervention provided to the specific cohorts of participants being assessed:

RQ1: Are there statistically significant differences on baseline assessments of earth science knowledge (Pretest) between the NGSS STEAM-first unit and NGSS STEM-first unit cohorts, and Emerging Bilingual and English Fluent cohorts?

RQ2: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences between NGSS STEAM-first and NGSS STEM-first cohorts (regarding results at final posttest 2)?

RQ3: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences between Emerging Bilingual and English Fluent cohorts (regarding results at final posttest2)?

RQ4: While statistically controlling for possible existing pretest mean differences and differing levels of implementation fidelity of the instructional intervention, are there statistically significant differences in overall science knowledge gains (i.e., change scores) obtained during the 9-week intervention?

RQ1 Findings

When random selection or assignment of participants to the treatment groups is part of a study, it is always informative and beneficial to assess the pretest levels of the participants. Therefore, specific to RQ1, the first independent-samples *t*-test was conducted to compare a dependent outcome variable assessing earth science knowledge scores at baseline assessment, and an independent grouping variable designating if students where to initially receive the earth science curriculum using a STEAM-first (STEAM before STEM) approach or STEM-first (STEM before STEAM) approach. There was a significant difference in scores for STEAM-first classroom students ($m = 5.61$, $SD = 2.12$) and STEM-first classroom students ($m = 6.92$, $SD = 2.61$; $t(355) = 5.24$, $p = .001$). The magnitude of the differences in the means (mean difference = 1.31, 95% *CI*: .82 to 1.81) was a moderate (eta squared = .07). Therefore, and noting that random assignment was utilized before pretests to place the classrooms in either a STEAM-first or STEM-first cohort, the STEM-first cohort began this study with a statistically significantly higher knowledge level of the earth science information being taught and tested.

An independent-samples *t*-test also was conducted to compare earth science knowledge pretest scores between students' English language fluency level (i.e., emerging bilingual (EB) and English fluent (EF) cohorts). There was not a significant difference in earth science knowledge pretest scores for EF students ($m = 6.35$, $SD = 2.57$) and EB students ($m = 5.91$, $SD = 2.22$; $t(355) = 1.71$, $p = .089$). Although the EF student participants had a higher mean score on the pretest, both EF and EB students basically started with similar non-significantly different pretest scores. Please note, however, the magnitude of the differences in the means (mean difference = .44, 95% *CI*: .07 to .94) was a small effect (eta squared = .008). Table 1 highlights the mean scores and descriptive statistics/

Table 1. Pretest Earth Science Knowledge Mean Scores

	Instructional Intervention	Language Fluency	Mean	SD	N
Pretest/ Baseline	STEAM first	EF	5.34	2.12	93
		EB	5.82	2.12	115
		Total	5.61	2.12	208
	STEM first	EF	7.26	2.61	103
		EB	6.15	2.46	46
		Total	6.92	2.61	149
	Total	EF	6.35	2.57	196
		EB	5.91	2.22	161
		Total	6.15	2.42	357

Although research documents that EB often perform lower than their EF peers (NCES, 2014), due to this sample of schools being Title-I schools (i.e., lower income based and more highly transient low SES student populated schools), the nonsignificant pretest mean scores are understandable. As research previously cited on socio-economic status, poverty and transient school populations also supports, existence of performance gaps for all Title I students is often prevalent regardless of ethnicity or their first language. But as Figure 2 below highlights, due to our research project's random assignment procedures completed before pretest, as noted above the STEM-first cohort began this study with significantly higher knowledge of the earth science information being taught and tested. Furthermore, the EF students assigned to the STEM-first cohort had the highest of all scores, and the STEM-first EB students had higher pretest scores than the STEAM-first EB students. Meanwhile, the EF students assigned to the STEAM-first cohort had the lowest of all scores. Such documentation of pretest score similarities and differences are important to consider and account for and will be discussed below.

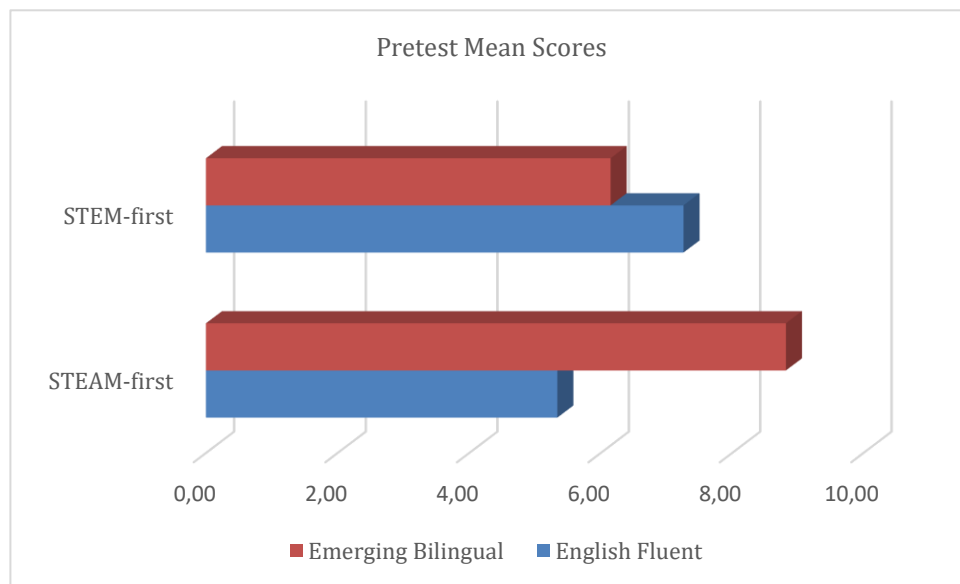


Figure 2. Pretest Earth Science Knowledge Mean Scores

RQ2 & RQ3 Findings

Given we know where the student participants started related to the pretest mean scores above, let's now consider what their final test scores documents. To explore research questions 2 and 3, we conducted a two-way between-groups analysis of covariance (ANCOVA) to next explore the outcomes of the final tests for the student participants. Regarding this ANCOVA, the measurement of the implementation fidelity of the instructional intervention was utilized as a covariate. As shared above, this measurement of implementation fidelity was obtained through regular observations by trained members of the research team and scoring of the participating teachers' efforts. Additionally, review and scoring of detailed activity logs on the quantity and quality of the intervention being provided by each teacher was included in the implementation fidelity index.

With the significant differences and yet also slight differences between the pretest scores of differing English fluency groups and those either receiving the STEAM-first or STEM-first intervention, we also utilized the student participants' pretest scores as a covariate to statistically control within the analysis. With existing knowledge of earth science and to what extent an intervention might have been provided with fidelity possibly having an impact on the outcomes of the

study, both variables could serve as suspects which might influence scores of the dependent variable (i.e., final posttest). Additionally, with IBM SPSS version 24 using regression procedures to remove variation in the dependent variable due to the covariates and performing the normal analysis of variance on the corrected or adjusted scores, one can increase the sensitivity of the F-test and reduce the probability of Type I and II error. Therefore, instead of performing a regression analysis which is used ideally for building prediction models once research has established associations, we used an ANCOVA to first explore if there is a difference between the group's knowledge gains impacted by the instructional intervention order effect. Our focus was on documenting evidence that the instructional approach could become a predictor in future research utilizing larger samples and regression analyses.

Therefore, to explore research questions two and three, a two-way between-groups analysis of covariance (ANCOVA) was conducted to explore the impact of the independent variables of instructional intervention approach (i.e., STEAM-first vs STEM-first) and the English fluency level of the participants (i.e., English fluent vs emerging bilingual), as measured by the 4th NGSS earth science final posttest. Again, preliminary checks were conducted to ensure that there was no violation of the assumptions or normality, linearity, homogeneity of variances, homogeneity of regressions slopes, and reliable measurement of dependent variables and covariates. While utilizing covariates to statistically adjust for pretest scores and the implementation fidelity of the intervention, there was a statistically significant interaction effect between instructional intervention and English fluency groups, $F(1, 351) = 9.00, p = .003$, with a small effect size (partial eta squared = .025).

Additionally, for the independent variable of instructional intervention there was a statistically main effect, $F(1, 351) = 15.25, p = .001$, with a small effect size (partial eta squared = .042); as well as for the independent variable of English fluency level, $F(1, 351) = 5.80, p = .017$, with a small effect size (partial eta squared = .016). Furthermore, for the pretest used as a covariate there was a statistically main effect, $F(1, 351) = 116.43, p = .001$, with a large effect size (partial eta squared = .25); as well as for the implementation fidelity covariate, $F(1, 351) = 15.43, p = .001$, with a small effect size (partial eta squared = .048). Both significant covariate outcomes suggest that the pretest score and level of implementation fidelity impacted their final test scores.

Table 2. Descriptive Stats for Final Earth Science Posttest Scores

	Instructional Intervention	Language Fluency	Mean	SD	N
Final Posttest	STEAM first	EF	9.40	3.46	93
		EB	9.93	3.29	115
		Total	9.69	3.37	208
	STEM first	EF	11.30	3.42	103
		EB	8.28	3.11	46
		Total	10.37	3.60	149
	Total	EF	10.40	3.56	196
		EB	9.46	3.31	161
		Total	9.97	3.48	357

As the ANCOVA results and means in Table 2 provide, there was a statistically significant difference between the final posttest mean scores related to the STEAM-first and STEM-first cohorts. The STEM-first cohort ($m = 10.37, SD = 3.60$) had slightly statistically significant higher final posttest scores than their STEAM-first peers ($m = 9.69, SD = 3.37$). As highlighted in response to RQ1, and the STEM-first EF students having significantly higher pretest scores at the start of the study, once again we see the highest score coming from this specific category of students. The STEM-first EF students ($m = 7.26, SD = 2.61$) had significantly higher pretest scores than their STEAM-first EF peers pretest scores ($m = 5.34, SD = 2.11$), and at final posttest the STEM-first EF students ($m = 11.30, SD = 3.42$) still have significantly higher scores than their STEAM-first EF peers ($m = 9.40, SD = 3.46$). This statistical significance at pretest and final posttest between the STEAM-first EF cohorts, and the STEM-first EF participants, with the STEM-first EF participants producing much higher scores than their STEAM-first EF peers and others, was most likely an impetus to the STEM-first producing higher scores.

Because, even though there was a statistically significant difference between the English fluency level groups, and the differences show that the EF student participants ($m = 10.30, SD = 3.56$) scored higher than the EB student participants ($m = 9.46, SD = 3.48$), it worth looking more closely at the differences in the final posttest mean scores of the EB student participants. While the STEM-first EF students ($m = 11.30, SD = 3.42$) had the definitively higher score, it was the STEM-first EB students ($m = 8.28, SD = 3.11$) who had the definitively lowest score. When we look more specifically at the EB students and consider that the STEAM-first EB students ($m = 9.93, SD = 3.29$) had the second highest score of the categories, it becomes apparent that there was also a statistically significant difference between the STEAM-first EB students and STEM-first EB students ($m = 8.28, SD = 3.11$).

The means of the final posttest provide initial evidence that the EB students potentially benefited more from receiving the STEAM-first approach. In fact, the STEM-first EB students ($m = 6.15, SD = 2.46$) had higher pretest scores than their

STEAM-first EB peers ($m = 5.82$, $SD = 2.12$). And as the analysis to follow exploring overall knowledge gained (change scores) highlights, though STEM-first cohort produced significantly higher final posttest scores than the STEAM-first cohort, and the EF cohort produced significantly higher final posttest scores than the EB cohort, it was the STEAM-first cohort who documented the highest overall knowledge gains during the 9-week study, and the EB students receiving the STEAM-first cohort far outshined the EB STEM-first cohort.

RQ4 Findings

For research question four, and to take a slightly different look at the knowledge growth, using a measure of total knowledge gained from pretest to final posttest, a two-way between-groups analysis of covariance (ANCOVA) was conducted to explore the impact of the independent variables of Instructional Intervention approach (i.e., STEAM-first vs STEM-first) and the English fluency level of the participants (i.e., English fluent vs Emerging Bilingual). Basically, instead of using the final posttest as the dependent variable as we did for research questions two and three, we subtracted the student participants' pretest scores from the final posttest score to compute a change score (e.g., final posttest – pretest = overall knowledge gains = change scores). The change score used in this analysis represents a number reflecting total earth science knowledge gained from pretest to final posttest. Again, however, scores on the fourth grade NGSS earth science pretest and assessments of the intervention implementation fidelity level provided to the students' classrooms were used as covariates to statistically control for individual differences.

Preliminary checks were conducted to ensure that there was no violation of the assumptions or normality, linearity, homogeneity of variances, homogeneity of regressions slopes, and reliable measurement of dependent variables and covariates. While statistically adjusting for the pretest scores and implementation fidelity of the intervention, there was a statistically significant interaction effect between instructional intervention and English fluency groups, $F(1, 351) = 9.02$, $p = .003$, with a small effect size (partial eta squared = .025). Additionally, for the independent variable of instructional intervention there was a statistically main effect, $F(1, 351) = 15.42$, $p = .001$, with a small effect size (partial eta squared = .042); as well as for the independent variable of English fluency level, $F(1, 351) = 5.80$, $p = .017$, with a small effect size (partial eta squared = .016). Furthermore, for the pretest used as a covariate there was a statistically main effect, $F(1, 351) = 20.06$, $p = .001$, with a nearly medium effect size (partial eta squared = .05); as well as for the implementation fidelity covariate, $F(1, 351) = 17.81$, $p = .001$, with a small effect size (partial eta squared = .048).

Table 3. Overall Earth Science Knowledge Gain Change Scores

	Instructional Intervention	Language Fluency	Mean	SD	N
Change Scores	STEAM first	EF	4.05	3.03	93
		EB	4.11	3.03	115
		Total	4.00	3.02	208
	STEM first	EF	4.04	2.89	103
		EB	2.13	2.91	46
		Total	3.45	3.02	149
	Total	EF	4.05	2.95	196
		EB	3.55	3.12	161
		Total	3.82	3.03	357

As Table 3 documents, differing from the final posttest outcomes, for this analysis on the overall change scores we find that it was the STEAM-first approach which produced the statistically significant higher score related to the instructional intervention variable. When examining the collection of mean scores in Table 3, however, one anomaly stands out, the mean score of the STEM-first EB student cohort ($m = 2.13$). While the other three categories of student knowledge gains reflected a mean score of 4.04 to 4.11, the EB students assigned to receive the STEM-first approach did half as well when it came to learning gains. Figures 3 and 4 further illustrate the greater advancements in science knowledge obtained by EB students in relation to STEAM-first versus STEM-first. Figure 3 and 4 clearly illuminates that those with lowest posttest scores and lowest overall knowledge gains (i.e., change scores), are the EB students who received a STEM-first approach.

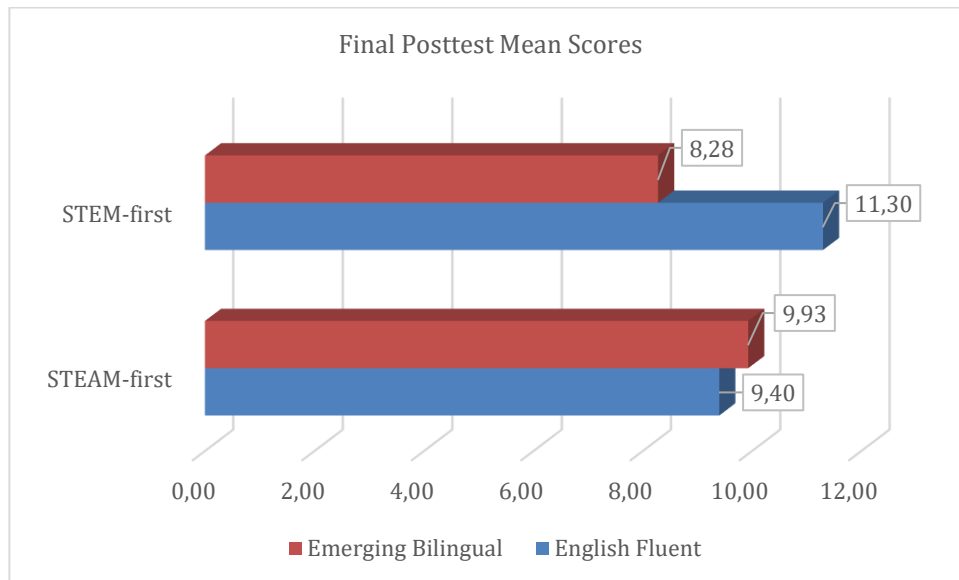


Figure 3. Final Posttest Mean Scores

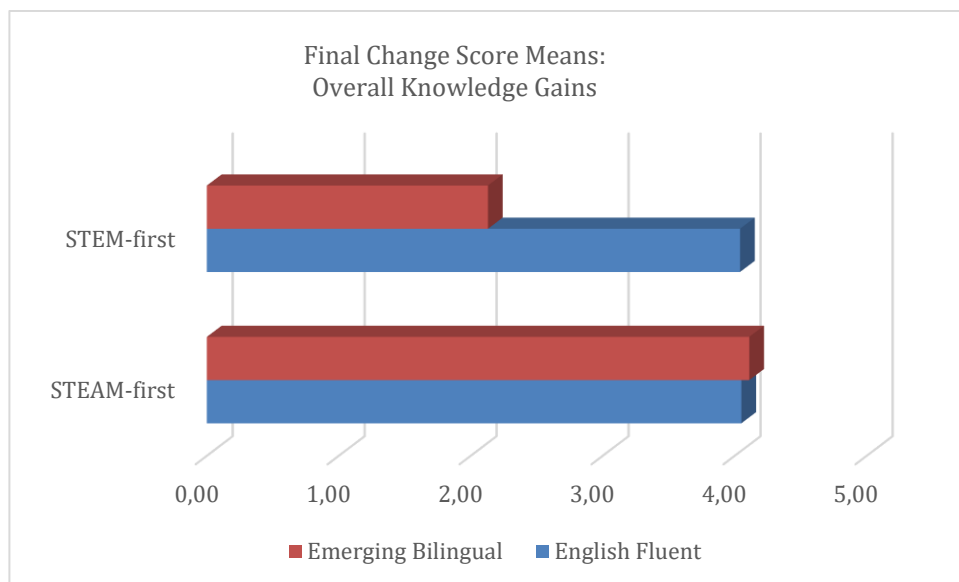


Figure 4. Final Change Score Means

Additional Detail on First Posttest

Some might wonder, however, why the first Posttest was not addressed within this study. The reason we left such elements out of the research questions was due to journal word count requirements and because the results were similar to the final posttest. As Table 4 below shares, when comparing the pretest and final posttest means provided above to the first posttest, the mean scores were higher than the pretest but lower than and yet somewhat like the final posttest. The means suggested that at halfway through the 9-week intervention the scores were rising, trending in the direction we identified with final posttest.

Table 4. Posttest 1 Earth Science Mean Scores

	Instructional Intervention	Language Fluency	Mean	SD	N
Posttest 1	STEAM first	EF	8.22	3.25	93
		EB	8.80	3.23	115
		Total	8.54	3.24	208
	STEM first	EF	10.15	3.35	103
		EB	7.59	3.12	46
		Total	9.36	3.48	149
	Total	EF	9.23	3.44	196
		EB	8.45	3.24	161
		Total	8.88	3.37	357

Discussion

Though the findings of this study are not definitive of the efficacy of leading with STEAM-first approaches, they do provide initial evidence which suggests further exploration could be beneficial. While results from research questions 2 and 3 showed significantly higher posttest scores for the STEM-first and English fluent cohorts, we must keep in mind the results from research question 1 which documented that the STEM-first and English fluent cohorts started the study with statistically significant higher pretest scores. Basically, these cohorts had a head start on the STEAM-first and EB student participants, similar to a 100-meter race and letting certain participants basically get to start 10 to 20 meters ahead of their competitors.

When it came to research question 4, however, we found it was the STEAM-first and EB student cohorts who produced statistically significant higher knowledge gains or change scores over the 9-week period. Figure 4, however, provides an illustration as to what we feel is one of the most important findings of this study. With outcomes showing how EB students assigned to the STEM-first experienced the lowest improvement in earth science overall knowledge gains measured via change scores, reflection on the possibilities why this took place are warranted. Is it possible the STEAM-first methods aided in decreasing cognitive load and making abstract concepts more concrete and accessible through multimodality and embodied representation as researchers have suggested (Campbell et al., 2016; Wahyuningsih et al., 2020)? As the results of this study documented, when it comes to fourth grade Title I and EB students learning earth science, a STEAM-first approach produced higher scores in overall knowledge gains than the STEM-first approach, and for the emerging bilingual students these higher scores and knowledge gains were significantly higher than their English fluent classmates.

An extensive literature review by our research team did not identify any studies similar to what we performed in relation to integrating the arts in hopes to improve science learning for elementary students. Similar to research in other subject matter areas exploring if the arts could help improve academic performance, however, we did find similar results. Like the Bravo and Cervetti (2014) and Huerta and Jackson (2010) research efforts, the results clearly show EB students continue to be challenged by STEM education. The results of this study, however, also suggest that leading with a STEAM-first approach is contributing to making the abstract concepts taught within STEM efforts more concrete and accessible for EB students, similar to Fillmore's findings (2007).

The results also suggest that STEAM-first approaches to science education might provide the support Goldenberg (2013) found to be of need by EL students. This study shows promise that integrating the arts does not dilute STEM efforts as some have hypothesized in the past (Heilig et al., 2010). As Figures 3 and 4 highlight, even the English fluent students performed better with a STEAM-first approach. And we hope this study provides encouragement to others to further explore how such efforts can increase reading comprehension (Podlozny, 2000; Walker et al., 2011), as well as enhance students' inquiry skills, problem solving skills, and creative thinking (Segarra et al., 2018). But most of all, this study personally encourages us to further explore how such efforts enhanced by the arts can make science education more equitable for EB students.

As the study limitations to follow share, we also were reminded that when you are studying a rather transient population of at-risk youth, attrition makes things difficult when it comes to tracking longitudinal progression. As a result, our sample could have been bigger and possibly contributed to more robust findings. Additionally, this longitudinal study funded by the National Science Foundation definitely taught us that unfortunately, sometimes, when performing research to truly test or explore if a novel intervention supported by past research can make the meaningful change, well, in pursuit to collect reliable and valid data via strict research methodology related to the study's focus you find results that some might label as finding nothing. Even if statistical significance is identified, with small effect sizes, to some it still means the study found nothing.

But we disagree. Because if time is taken to truly look at what the study completely documented, and go beyond what the stats survival manuals and professors suggest writing (e.g., test and document a statistically significant interaction effect was evident), one might discover that the study does identify valuable insights and promise to future research efforts.

For example, by simply using the pretest and posttest scores to create change scores representing overall knowledge growth, the results of research question 4 provided an interesting twist to research questions 2 and 3 findings. Such efforts to recode variables in slightly different formats clarified that though STEM-first and English fluent cohorts produced the statistically significant results for research questions 2 and 3, it was the STEAM-first and EB students who showed greater growth.

Conclusion

A most likely factor tied to producing the smaller effect sizes, however, was the size of the standard deviations documented in the tables. Of course, these scores were the results of the test that were taken by the students we selected. But by utilizing random assignment and sampling techniques, to better ensure the intervention would be delivered with fidelity and each school would be assigned to a specific STEAM-first or STEM-first cohort to implement consistently in each school, we ended up with samples of students who did not have the same earth science knowledge levels at pretest. Therefore, from the start we had some methodological challenges which we knew might lead to some analysis challenges.

Variability is a tricky thing or situation. You want some variability in the data you collect on participants, otherwise it might feel like you are assessing clones of the same individual. Also sometimes having little variability leads to not identifying differences. But too much variability can dramatically reduce your statistical power and statistical power is the probability that influences if you will detect a difference or effect.

Unfortunately, the sample randomly selected and utilized from our participating population of Title I schools, provided a large amount of variation between the scores. The reality is that not all Title I students are under performing, and there are some very smart children who attend Title I schools, and there are Title I families who see education as their child's most important tool for a better future. And some of these parents in the USA are some of the best parents you will find and put everything they have into their children's education. But statistics do not always appreciate the world of reality, and the larger the standard deviation being used to divide the difference of the population means, the smaller the effect size.

This is why those performing regression analysis are instructed to check for outliers and run an analysis that basically deletes such outliers. But given the participants in this study were all from Title I schools, basically this study was an analysis of outliers when compared to what the greater population of students reflect across the USA. And as the scores and standard deviations reflect in this study, there was great variability between the students who took the test and how they scored.

But beyond the methodological challenges, with the positive and promising findings as shared above, and given the cause is worthy (i.e., helping those struggling with learning materials and concepts taught in a language they have not yet fully grasped), we hope this study encourages others to replicate such efforts as well as expand and improve upon the research. Despite the small effect sizes, the statistically significant results provide evidence that the EB students benefited more from receiving the STEAM-first approach than the STEM-first approach, and that the STEAM-first approach led to statistically significant higher earth science knowledge gains than the STEM-first approach.

On a science test worth 17 points, however, the highest mean score established on the pretest was 7.26 (43%) and on the final posttest 11.30 (66%). These were not scores to celebrate an amazing breakthrough. Please note, however, that in the USA 74% of fourth graders in Title I schools do not read with proficiency (Annie E. Casey Foundation, 2020), and the science test utilized for this study required reading skills. Therefore, seeing improvement in this study on a sample of Title I student participants is promising, when considering the Title I population in general is not currently showing and has not shown any progress in reading proficiency year after year, decade after decade in the USA.

For this study, case in point, though not definitively supported, the goal was to eliminate the statistically significant differences in academic performance between students in the USA who spoke English fluently, and EB students. The goal was to "shrink the gap" between those who were born into the language all the textbooks were written and teachers spoke, and those who were still learning the language, emerging bilingual (EB) learners. But despite not identifying this specific goal, the study did show promise for incorporating the arts into science instruction, and how leading with the arts augmented STEAM science instruction before more traditional STEM efforts, might just help those challenged by the language or reading in general. As a result, we hope we have made an argument supporting the fact that sometimes finding nothing, is actually something very interesting.

Recommendations

As to how this study could be improved upon in the future, the biggest challenge from our experience comes from sampling and selection of the participants. To study a large enough sample of EB students and Title I students, future research will most likely have to focus on schools within large Title I districts. With such samples we are not sure one can avoid the limitations to follow or the variability issues we incurred. One could consider utilizing a more purposive or strategic approach to sampling to avoid having cohorts differ dramatically in pretest scores, but in order to run inferential and regression analyses you must use random sampling. Therefore, the recommendations, as discussed throughout the paper, is to be ready for methodological challenges, and do due diligence to limit such limitations.

Additionally, we have a recommendation for journals, professors and researchers. For decades in academia and the world of research, when it comes to manuscript publication, an unwritten rule has been shared with new faculty, researchers and graduate students. Basically, such *unspoken* guidance many have received suggests that to successfully publish a study, your research findings must document statistically significant outcomes. Although some recent efforts seek to move away from this strong reliance and focus on significance tests (Ioannidis, 2019), reality suggests *p*-values and effect sizes, inferential analysis (i.e., analyses we more safely can make inferences from) and identifying the probability and effect size of statistical outcomes is very important and informational.

But what about when testing a hypothesis or research question based on past successful studies, and efforts to replicate such studies on more diverse or specific samples end up not finding similar results to past efforts? Or what about trying to document how existing statistically significant differences between two samples in need of research and support have been reduced dramatically by utilizing a specific approach to treatment or instruction, but the statistical significance is not documented at the end or the effect sizes are small? In both cases, the possibility exists that such research outcomes, where statistically significant results were not documented or small effect sizes suggest the findings are not practical, could provide a beneficial contribution to those who read journal publications and seek to broaden a field's knowledge.

Science teachers share with students every year how studies must be replicated to be valid. Some of the greatest discoveries, inventions, and scientific breakthroughs ever achieved were the result of a series of failed studies. But if we are only publishing studies that find statistical significance with moderate to large effect sizes, we are creating a research world where no one ever hears about the failed studies. As a result, though a study might have been done showing significance, many never hear that the study was anomalous and that the intervention does not work.

Through learning from previous failed efforts and persistent replication of such studies seeking to improve methods, sampling, data collection, and analyses to achieve better statistical outcomes, is how scientists and researchers make progress. Yet, in today's world, many editors, and reviewers of journals, basically make it a publication prerequisite for external review that statistically significant findings were identified. Or as Amrhein et al. (2019) share, "Unfortunately, the false belief that crossing the threshold of statistical significance is enough to show that a result is 'real' has led scientists and journal editors to privilege such results, thereby distorting the literature" (p. 307).

Regardless of one's position, statistics are essential to determining the value, applicability, and generalizability of published outcomes. If we cannot receive a high level of assurance that the variables were measured with reliable and valid tools, the findings did not happen by chance and the outcomes were robust and meaningful, we might as well forget our worries of accounting for Type I and II error. If we are not going to require all to address and fulfill the many assumptions which accompany all inferential analyses (as far too many published studies continue to do), such as ignore the requirements of random sampling and established validity and reliability of assessment tools utilized (Cohen et al., 2013), we might as well just flip a coin and save a lot of effort. But given the most basic fundamentals of behaviorism encompass the pursuit to measure if certain stimuli lead to changes in response, and let's face it— sometimes stimulus don't lead to the expected response-- a heavy reliance if not requirement to identify statistical significance and large effect sizes when publishing a study actually can lead to excluding valuable insights to those in the field.

Moving beyond the norm of only publishing studies which utilize regression or only find massive significance and effect sizes is much needed in a society questioning the scientific method. Publishing failed studies or studies which find promising or conflicting results can help future researchers troubleshoot what might need changed, what's not working, and augmenting the intervention for better outcomes. Again, the point of this discussion is to highlight how sometimes finding *nothing* is actually something fairly interesting, and sometimes as this study on STEAM and STEM illustrates, represents a favorable outcome that though not robust, might be of value to individuals wanting to overcome the challenges EB students face.

Limitations

There were limitations due to the methodology of the study which as a result led to challenges impacting selection of analyses and the statistics produced. As with any large grant effort studying transient student populations, limitations due to recruitment, random assignment of teachers and attrition led to slight challenges with unequal samples or significant differences between baseline knowledge levels between the English fluent and EB students. Random assignment created some challenges to the comparative samples utilized for the analysis. Furthermore, student participants were not tested before the sampling and random assignments were completed. Although as the results suggest for research questions 2, 3, and 4 that higher mean scores were trending and effect sizes were small, the cell sizes of these samples most likely contributed to not being able to produce significance. And as discussed previously, the standard deviation and variability encountered led to effect size limitations.

Funding

This study was part of the Equitable Science Curriculum for integrating arts in Public Education (ESCAPE) funded by the National Science Foundation.

Authorship Contribution Statement

Corrigan: Concept and design, data acquisition, data analysis / interpretation, drafting manuscript, critical revision of manuscript, statistical analysis, securing funding. Grove: Concept and design, data acquisition, data analysis / interpretation, critical revision of manuscript, securing funding. Anderson: admin, technical support, material support. Wong: Critical revision of manuscript, statistical analysis, technical support, material support. Hughes: Concept and design, data acquisition, data analysis / interpretation, critical revision of manuscript, securing funding, admin, supervision.

References

- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>
- Annie E. Casey Foundation. (2020, January 15). *Reading scores: States not doing right by kids in Title I schools*. Kids Count Data Center. <https://bit.ly/3lkdIY8>
- Barcelona, K. (2014). 21st century curriculum change initiative: A focus on STEM education as an integrated approach to teaching and learning. *American Journal of Educational Research*, 2(10), 862-875. <https://doi.org/10.12691/education-2-10-4>
- Becker, K. H., & Park, K. (2011). Integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A meta-analysis. *Journal of STEM education: Innovations and research*, 12(5), 23-37.
- Bravo, M. A., & Cervetti, G. N. (2014). Attending to the language and literacy needs of English learners in science. *Equity & Excellence in Education*, 47(2), 230-245. <https://doi.org/10.1080/10665684.2014.900418>
- Brown, P. M., Corrigan, M. W., & Higgins-D'Alessandro, A. (Eds.). (2012). *Handbook of prosocial education*. Rowman & Littlefield.
- Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of science teacher education*, 25(2), 211-221. <https://doi.org/10.1007/s10972-014-9381-4>
- Campbell, T., Schwarz, C., & Windschitl, M. (2016). What we call misconceptions may be necessary stepping-stones toward making sense of the world. *The Science Teacher*, 83(3), 69-74. https://doi.org/10.2505/4/tst16_083_03_69
- Catterall, J. S. (2009). *Doing well and doing good by doing art: The long-term effects of sustained involvement in the visual and performing arts during high school: results of a 12-year national study*. Imagination Group/I-Group Books.
- Castro-Alonso, J. C., Paas, F., & Ginns, P. (2019). Embodied cognition, science education, and visuospatial processing. In J. C. Castro-Alonso (Ed.), *Visuospatial processing for education in health and natural sciences* (pp. 175–205). Springer, Cham. https://doi.org/10.1007/978-3-030-20969-8_7
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203774441>
- Corrigan, M. W., Higgins-D'Alessandro, A., & Brown P. M. (2013). The case for adding prosocial education to current education policy: Preparing students for the tests of life, not just a life of tests. *Kedi Journal of Educational Policy*, 10(3), 37-50. <https://Bit.ly/3nJKil8>
- Crowder, M. J., & Hand, D. J. (2017). *Analysis of repeated measures*. Routledge. <https://doi.org/10.1201/9781315137421>
- Daugherty, M. K. (2013). The prospect of an "A" in STEM education. *Journal of STEM Education: Innovations and Research*, 14(2), 10-15. <https://bit.ly/3lojpnU>
- Dewey, J. (2005). *Art as experience*. Penguin.
- Donovan, L., & Pascale, L. (2012). *Integrating the arts across the content areas*. Teacher Created Materials.
- Edens, K. M., & Potter, E. (2003). Using descriptive drawings as a conceptual change strategy in elementary science. *School science and mathematics*, 103(3), 135-144. <https://doi.org/10.1111/j.1949-8594.2003.tb18230.x>
- Fillmore, L. W. (2007). English Learners and mathematics learning: Language issues to consider. In A. H. Schoenfeld (Ed.), *Assessing Mathematical Proficiency* (pp. 333–344). Cambridge University Press. <https://doi.org/10.1017/CBO9780511755378.026>
- Gess, A. H. (2017). STEAM education: Separating fact from fiction. *Technology and Engineering Teacher*, 77(3), 39–41. <https://Bit.ly/3GNjp0t>
- Ghanbari, S. (2015). Learning across disciplines: A collective case study of two university programs that integrate the arts with STEM. *International Journal of Education & the arts*, 16(7), 1-22. <http://www.ijea.org/v16n7/>
- Glenberg, A. M. (2011). How reading comprehension is embodied and why that matters. *International Electronic Journal of Elementary Education*, 4(1), 5-18. bit.ly/3tHoZ0r

- Goldenberg, C. (2013). Unlocking the research on English learners: What we know--and don't yet know--about effective instruction. *American Educator*, 37(2), 4-11. <https://bit.ly/3rE0puM>
- González-Howard, M., & Suárez, E. (2021). Retiring the term English language learners: Moving toward linguistic justice through asset-oriented framing. *Journal of Research in Science Teaching*, 58(5), 749-752. <https://doi.org/10.1002/tea.21684>
- Guyotte, K. W., Sochacka, N. W., Costantino, T. E., Walther, J., & Kellam, N. N. (2014). STEAM as social practice: Cultivating creativity in transdisciplinary spaces. *Art Education*, 67(6), 12-19. <https://doi.org/10.1080/00043125.2014.11519293>
- Hadzigeorgiou, Y. (2016). 'Artistic' Science Education. In Y. Hadzigeorgiou (Ed.), *Imaginative Science Education: The Central Role of Imagination in Science Education* (pp. 185-215). Springer International Publishing. https://doi.org/10.1007/978-3-319-29526-8_7
- Hardiman, M., Magsamen, G. M., McKhann, G., & Eilber, J. (2009). *Neuroeducation: Learning, arts and the brain*. John Hopkins University. <https://bit.ly/3qNLQG1>
- Heilig, J. V., Cole, H., & Aguilar, A. (2010). From Dewey to No Child Left Behind: The evolution and devolution of public arts education. *arts Education Policy Review*, 111(4), 136-145. <https://doi.org/10.1080/10632913.2010.490776>
- Hetland, L., Winner, E., Veenema, S., & Sheridan, K. M. (2015). *Studio thinking 2: The real benefits of visual arts education*. Teachers College Press.
- Huerta, M., & Jackson, J. (2010). Connecting literacy and science to increase achievement for English language learners. *Early Childhood Education Journal*, 38(3), 205-211. <https://doi.org/10.1007/s10643-010-0402-4>
- Hwang, J., & Taylor, J. C. (2016). Stemming on STEM: A STEM education framework for students with disabilities. *Journal of Science Education for Students with Disabilities*, 19(1), 39-49. <https://doi.org/10.14448/jesed.09.0003>
- Ingram, D., & Riedel, E. (2003). *What does arts integration do for students?* [Report]. Center for Applied Research and Educational Improvement. <https://Bit.ly/3GM2tas>
- Ioannidis, J. P. (2019). The importance of predefined rules and prespecified statistical analyses: do not abandon significance. *Jama*, 321(21), 2067-2068. <https://doi.org/10.1001/jama.2019.4582>
- Jones, B., & Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman and Hall/CRC. <https://doi.org/10.4324/9780203009277>
- Karakelle, S. (2009). Enhancing fluent and flexible thinking through the creative drama process. *Thinking Skills and Creativity*, 4(2), 124-129. <https://doi.org/10.1016/j.tsc.2009.05.002>
- Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM education*, 3(1), 1-11. <https://doi.org/10.1186/s40594-016-0046-z>
- Lee, H. K. (2005). Rethinking arts marketing in a changing cultural policy context. *International Journal of Nonprofit and Voluntary Sector Marketing*, 10(3), 151-164. <https://doi.org/10.1002/nvsm.9>
- Lee, O., Llosa, L., Grapin, S., Haas, A., & Goggins, M. (2019). Science and language integration with English learners: A conceptual framework guiding instructional materials development. *Science Education*, 103(2), 317-337. <https://doi.org/10.1002/sce.21498>
- Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., Kastberg, D., & Jocelyn, L. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective. Highlights. NCES 2005-003*. US Department of Education. <https://Bit.ly/3fjJCB6>
- Liao, C. (2016). From interdisciplinary to transdisciplinary: An arts-integrated approach to STEAM education. *Art Education*, 69(6), 44-49. <https://doi.org/10.1080/00043125.2016.1224873>
- Lichtenberg, J., Woock, C., & Wright, M. (2008). Ready to innovate: Are educators and executives aligned on the creative readiness of the US workforce? *Americans for the Arts*. <https://bit.ly/3AhUqji>
- Luftig, R. L. (2000). An investigation of an arts infusion program on creative thinking, academic achievement, affective functioning, and arts appreciation of children at three grade levels. *Studies in Art Education*, 41(3), 208-227. <https://doi.org/10.1080/00393541.2000.11651678>
- Madden, M. E., Baxter, M., Beauchamp, H., Bouchard, K., Habermas, D., Huff, M., Ladd, B., Pearson, J., & Plague, G. (2013). Rethinking STEM Education: An Interdisciplinary STEAM Curriculum. *Procedia Computer Science*, 20, 541-546. <https://doi.org/10.1016/j.procs.2013.09.316>
- May, G. S. (2015, March 30). *STEM, not STEAM*. Inside Higher Ed. <https://Bit.ly/3fKcUiR>
- Mehu, M., Little, A. C., & Dunbar, R. I. (2008). Sex differences in the effect of smiling on social judgments: an evolutionary approach. *Journal of Social, Evolutionary, and Cultural Psychology*, 2(3), 103-121. <https://doi.org/10.1037/h0099351>

- Moga, E., Burger, K., Hetland, L., & Winner, E. (2000). Does studying the arts engender creative thinking? Evidence for near but not far transfer. *Journal of Aesthetic Education*, 34(3/4), 91-104. <https://doi.org/10.2307/3333639>
- National Center for Education Statistics. (2000). *The condition of education 2000* (NCES 2000-062). U.S. Department of Education. <https://nces.ed.gov/pubs2000/2000062.pdf>
- National Center for Education Statistics. (2007). *Highlights from PISA 2006* (NCES 2008-016). U.S. Department of Education. <https://nces.ed.gov/pubs2008/2008016.pdf>
- National Center for Education Statistics. (2014). *The condition of education 2014* (NCES 2014-083). U.S. Department of Education. <https://nces.ed.gov/pubs2014/2014083.pdf>
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. *NGSS Lead States. (2013). Next Generation Science Standards: For States, by States*. The National Academies Press.
- Olson, S., & Riordan, D. G. (2012). *Engage to excel: Producing one million additional college graduates with degrees in Science, Technology, Engineering, and Mathematics. Report to the President*. Executive Office of the President. <https://eric.ed.gov/?id=ED541511>
- Organization for Economic Co-operation and Development. (2007a). PISA 2006: Science Competencies for Tomorrow's World. <https://www.oecd.org/pisa/pisaproducts/39725224.pdf>
- Organization for Economic Co-operation and Development. (2007b). PISA 2006: Science Competencies for Tomorrow's World. OECD briefing note for the United States. <https://www.oecd.org/unitedstates/39722597.pdf>
- Organization for Economic Co-operation and Development. (2013). *PISA 2012 results: Ready to learn: Students' engagement, drive and self-beliefs (volume III)*. <https://Bit.ly/3AjVhQH>
- Organization for Economic Co-operation and Development. (2019). *PISA 2018 results (Volume I): What students know and can do*. <https://doi.org/10.1787/5f07c754-en>
- Organization for Economic Co-operation and Development. (2020). *Science performance (PISA) (indicator)*. <https://doi.org/10.1787/91952204-en>
- Ormrod, J. E. (2011). *Educational psychology: Developing learners* (7th ed.). Pearson.
- Podlozny, A. (2000). Strengthening verbal skills through the use of classroom drama: A clear link. *Journal of Aesthetic education*, 34(3/4), 239-275. <https://doi.org/10.2307/3333644>
- Poza, L. E. (2018). The language of ciencia: Translanguaging and learning in a bilingual science classroom. *International Journal of Bilingual Education and Bilingualism*, 21(1), 1-19. <https://doi.org/10.1080/13670050.2015.1125849>
- Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). Highlights from TIMSS 2011: Mathematics and science achievement of US fourth-and eighth-grade students in an international context. NCES 2013-009. *National Center for Education Statistics*. <https://Bit.ly/3rAa2eb>
- Schlaack, N., & Steele, J. S. (2018). The collaborative residency project: The influence of co-teaching on professional development in arts integration. *International Journal of Education & the Arts*, 19(11), 1-29. <http://www.ijea.org/v19n11/>
- Segarra, V. A., Natalizio, B., Falkenberg, C. V., Pulford, S., & Holmes, R. M. (2018). STEAM: Using the arts to train well-rounded and creative scientists. *Journal of Microbiology & Biology Education*, 19(1), 1-7. <https://doi.org/10.1128/jmbe.v19i1.1360>
- Suárez, E. (2020). "Estoy Explorando Science": Emergent bilingual students problematizing electrical phenomena through translanguaging. *Science Education*, 104(5), 791-826. <https://doi.org/10.1002/sce.21588>
- Ünsal, Z., Jakobson, B., Wickman, P. O., & Molander, B. O. (2018). Gesticulating science: Emergent bilingual students' use of gestures. *Journal of Research in Science Teaching*, 55(1), 121-144. <https://doi.org/10.1002/tea.21415>
- Wahyuningsih, S., Nurjanah, N. E., Rasmani, U. E. E., Hafidah, R., Pudyaningtyas, A. R., & Syamsuddin, M. M. (2020). STEAM learning in early childhood education: A literature review. *International Journal of Pedagogy and Teacher Education*, 4(1), 33-44. <https://doi.org/10.20961/ijpte.v4i1.39855>
- Walker, E., Tabone, C., & Weltsek, G. (2011). When achievement data meet drama and arts integration. *Language Arts*, 88(5), 365-372. <https://Bit.ly/3Ks0Sjg>
- Wilmes, S. E., & Siry, C. (2020). Science notebooks as interactional spaces in a multilingual classroom: Not just ideas on paper. *Journal of Research in Science Teaching*, 57(7), 999-1027. <https://doi.org/10.1002/tea.21615>