Integrating arts with STEM and leading with STEAM to increase science learning with equity for emerging bilingual learners in the United States

Bradley S. Hughes[1*], Michael W. Corrigan[2], Douglas Grove[1,3], Sage B. Andersen[1,4] and Joseph T. Wong[1]

Abstract

Background: To inform STEM education for benefiting emerging bilingual (EB) and English fluent (EF) students, the present study evaluated the order effects of integrating science and arts within a large-scale, ongoing effort investigating the efficacies of Next Generation Science Standards (NGSS)-aligned Science Technology, Engineering, and Math (STEM) methodologies to provide more equitable opportunities to students to learn science through Arts integration (STEAM). The experiment examines the curriculum integrating order of implementing combinations of STEM and STEAM approaches in fifth grade life and physical science instruction, comparing (STEM → STEAM) vs (STEAM → STEM).

Results: *T* tests and a three-way between-groups analysis of covariance examined the impact of instructional order, language fluency, and teachers' implementation fidelity. Findings indicate similar results in life and physical sciences, in which the STEAM first approach produced significantly higher science learning gains for both EF and EB students, revealing some higher learning gains for EF students, but with greater STEAM first order effect advantages for EB students overall. While EF students show higher learning gain scores in the high fidelity classrooms, the advantage of the STEAM first order is greater for EB students in all classroom fidelity levels and even within low to moderate implementation fidelity classrooms, as may commonly occur, such that the integration order of STEAM before STEM strategy is particularly advantageous to EB learners.

Conclusions: The integration pattern of leading with STEAM and following with STEM offers an important opportunity to learn for EB students, and increases equity in opportunities to learn among EB and EF learners of science. Both EB and EF students benefit similarly and significantly in high fidelity implementation classrooms. However, the gains for EF students are not significant in low fidelity implementation classrooms, while in such low fidelity implementation classrooms, the EB students still benefited significantly despite the poor implementation. These results suggest that a strong compensating STEAM first order effect advantage is possibly involved in the implementation system for the EB population of learners. Teaching science through the arts with STEAM lessons is an effective approach that can be significantly improved through introducing STEM units with the STEAM first order effect advantage.

*Correspondence: bhughes@uci.edu

[1] University of California Irvine, Irvine, CA, USA
Full list of author information is available at the end of the article

Introduction

Educating students in science is a U.S. national priority for both social and economic reasons (American Association for the Advancement of Science [AAAS], 1993; Bybee, 2014; Olson & Riordan, 2012). However, the US has trailed other developed nations in science-related skills (National Commission on Mathematics et al., 2000), with science performance trajectories remaining flat over the last two decades overall as measured by the PISA assessment (Organization for Economic Cooperation & Development [OECD], 2019). In the 2019 National Assessment of Education Progress (NAEP) science assessment report, researchers found that approximately only 30% of fourth-grade teachers identified that their students only participated once or twice in inquiry-based science activities during a given school year (NAEP, 2019). When compared to students, where teachers implemented a greater frequency of scientific inquiry-based activities (i.e., once or twice a day), on average, students with less access to science inquiry activities had significantly lower science test scores (NAEP, 2019). In both the US and abroad, learning science begins in the elementary grades, which is a critical timeframe for students as they are introduced to foundational and cross-cutting concepts necessary for later success in science achievement (National Research Council [NRC], 2012). Disparities in science achievement emerge early, and programs that address these issues need to target the early elementary and preschool years (Betancur et al., 2018). This is especially so for emerging bilingual (EB) students who are simultaneously learning science content and practices in addition to the primary language of instruction (Morgan et al., 2016).

Professionals responsible for facilitating the linguistic and academic competence of schools must address questions of what most effectively supports the learning processes for children whose first languages are not English (Jones, 2015). Babaci-Wilhite (2019) argues that the incorporation of arts into the inquiry-based approach, using local languages and cultural references will improve learning and human rights, and that it is a major challenge in STEM education to support teachers enacting inquiry-based instruction and integrating the arts into classroom curriculum. Seeking to help EB learners with challenges of learning science, the present study established a large-scale effort to investigate potential efficacies of a STEM and STEAM integrated curriculum, specifically integrating arts with inquiry-based methods, and examining patterns of integration to provide more equitable learning opportunities for students to learn.

With interest in identifying ways to increase opportunities to learn science, looking specifically at EB students in the US, concerning patterns in science achievement emerge in the elementary years. According to the California Department of Education, 2016 CAASPP CST (Hardoin et al., 2016), science results showed that among fifth grade EB students in California, only 16% are performing at or above "proficient" on the science assessment compared to 62% of their English fluent (EF) peers. In addition, findings from the National Assessment of Educational Progress (NAEP) reveal that EB students score lower at all grade levels and are more likely to score below basic (Kena et al., 2014). Considering that it is in these elementary school years when differences in science achievement begin to emerge both between student groups within the US and between the US and other developed countries (Provasnik et al., 2012), studies involving elementary STEM education methods may be particularly crucial in understanding and addressing the need for increased opportunities for all students to learn science, and particularly EB students.

A new wave of science reform efforts have attempted to improve science learning opportunities in recent years. This is evidenced in the creation and widespread adoption of the Next Generation Science Standards (NGSS), which represents a multidisciplinary approach to (S)cience learning, with elements of (T)echnology, (E)ngineering, and (M)ath heavily incorporated, in line with global efforts to increase STEM learning (Ritz & Fan, 2015). The NGSS call for a new vision of science education in which students make sense of scientific phenomena through authentic engagement in science practices (NGSS Lead States, 2013). However, this requires students to use language in increasingly complex ways (Lee, 2018), so the need for instructional support in STEM education may be greater for EB students than for the EF students (Goldenberg, 2013). For example, results of Afitska and Heaton's (2019) study showed that EB students performing below EF peers was largely a challenge of active language production and scientific vocabulary in assessments.

Proponents of the arts suggest that integrating arts with STEM efforts (STEAM) may play a supportive role in science learning (Catterall, 2009; Daugherty, 2013; Donovan & Pascale, 2012; Guyotte et al., 2014; Hardiman et al., 2014), and perhaps as a way to make science education

more equitable for EB students (Dewey, 1934; González-Howard & Suárez, 2021; Hadzigeorgiou, 2016; Lee et al., 2019). For the purposes of this study, defining the arts will include the areas of performing arts (i.e., dance, music, and theatre), the presenting arts (i.e., visual arts), and the producing arts (i.e., media arts), as described by the National Council for Core Arts Standards (2014).

Driven by the need for evidence-based techniques and curricula for science learning in the elementary grades, the present study established a large-scale National Science Foundation (NSF) funded research initiative investigating the efficacies of NGSS-aligned STEAM methodologies for providing equitable opportunities to students through arts integration. This report specifically details analysis of science and arts-integrated curriculum to inform the broader inclusive fields of STEM and STEAM education. Acknowledging that the terms science, arts, STEM and STEAM can have multiple meanings across educational communities, we define our interpretation for their usage within this particular study report pursuant to goals of communicating applicable breadth and simplicity for potential readers. By offering specific findings about integrating arts-based with inquiry-based science, we hope to contribute to the efforts to develop and identify increasingly effective methods for integration within these larger domains of the STEM and STEAM fields. In the interdisciplinary spirit of transdisciplinarity to reach beyond narrow silos of research for the pressing needs of student learning we refer to the more inclusive general terms of STEM and STEAM to avoid unnecessarily limiting the potential applications of this study to only science and arts and rather to include the audiences of technology, engineering, arts and math within the paradigm of STEM/STEAM educational change to more fully contribute evidence for enhancing effective integration approaches for STEM and STEAM. The curriculum in this study utilized two different approaches for science teaching and learning: (1) a NGSS-aligned science unit which integrates an inquiry-based approach to science instruction (henceforth referred to as STEM unit, STEM lessons, or simply STEM); and (2) an NGSS-aligned science–arts-integrated unit which integrates an arts-based approach to science instruction (henceforth referred to as STEAM unit, STEAM lessons, or simply STEAM). This study compared two patterns of lesson integration, the "order effects" of teaching STEM lessons before STEAM lessons (STEM → STEAM) or the reverse order of teaching STEAM lessons before STEM lessons (STEAM → STEM), and explored the extent to which teacher implementation fidelity of these instructional approaches mediates ordering effects.

Theoretical framework

This study and its design are grounded in a conceptualization of learning which synthesizes cognitive and social constructivist perspectives. Thus, we see learning as both occurring within each individual as they actively assimilate and accommodate new ideas into their existing schemas to construct new knowledge (Piaget, 1971), and also as mediated through social interactions (Vygotsky & Cole, 1978). Arends (2015) explains the personal construction of "meaning" by the learner through experience is further influenced by the interactions of prior knowledge and new learning events. The cognitivist individual constructivist approach of Jean Piaget (1936, 1945) promotes the process of adaptation, which involves the learner's assimilation of new information with their existing knowledge. This in turn enables learners to make the appropriate modifications to their existing intellectual framework to accommodate new information through the process of equilibration (Inhelder & Piaget, 1958), or state of cognitive balance, which is the cognitive mechanism that drives the learning process such that learners reconcile contradictions and inconsistencies in their knowledge structures (Piaget & Cook, 1952; Wadsworth, 1996). If assimilation is the cognitive process of fitting new information into existing cognitive schemas, then these schemas can represent "units of knowledge" or "mental representations of facts and ideas" (Piaget, 1963), and accommodation is the cognitive process of revising existing cognitive schemas, perceptions, and understanding so that new information can be incorporated (Piaget, 1963).

Drawing on Vygotsky's sociocultural view of constructivism, we acknowledge the aspect of learning that is socially constructed through interactions with peers, instructors, and more knowledgeable others to facilitate acquisition of cultural values, beliefs, norms, and problem-solving strategies (Dewey, 1938); (Posner et al., 1982; Vygotsky & Cole, 1978). In the context of science education, students learn using and applying their new and existing ideas as they work to make sense of scientific phenomena (Schwarz et al., 2017). Importantly, this is a deeply social process, whereby students engage collaboratively in the same science and engineering practices (SEP) employed by scientists in the pursuit of new scientific knowledge (NGSS Lead States, 2013; NRC, 2012). This process of constructing new knowledge for a particular purpose has been referred to as knowledge-in-use (Harris et al., 2016), a process which, at once, requires students to internalize their own newly constructed knowledge while actively negotiating this knowledge construction through social interaction with their peers. These perspectives draw on decades of research on bilingual language learning (e.g., Larsen–Freeman,

2007; Zuengler & Miller, 2006), as well as more recent research exploring how emerging multilingual students learn science (e.g., Lee et al., 2019). Specifically, bilingual education research supports the view of language as both a personal construct, with each person having their own personal idiolect to draw upon as they communicate (Otheguy et al., 2015), and as a set of dynamic practices (García & Wei, 2014; Wei, 2018) learned through social interaction (Larsen–Freeman, 2007; Zuengler & Miller, 2006). Thus, similar to new knowledge development, new language development occurs through using language for a particular purpose, a view referred to as language-in-use (Lee et al., 2013, 2019). Taken together, since teaching in group learning situations largely consist of sharing and negotiation of knowledge that are socially constructed through language or dialogue, these social interactions may leverage language development while affording learners to actively construct new knowledge through a sociocultural constructivist lens (Vygotsky & Cole, 1978).

Given the importance of social interaction for both language and science knowledge learning, it is important to consider the affective side of learning, particularly for those students who are learning science in a new language. Wang (2020) notes that teachers should employ a variety of teaching methods attending to students' emotional changes in class, while increasingly, the affective filter hypothesis is applied to guide English teaching to help students learn English more efficiently. Affective filter is a term usefully explained by Krashen (1985), who defines it as a psychological barrier, caused by distracting emotional factors, such as anxiety in competitive learning environments, preventing language learners from fully absorbing comprehensible knowledge, since it must first go through the affective filter before it can be absorbed. For example, it is theorized that cognitive load is increased for students who contend with the processing of new information in content areas as well as language, resulting in language-learning anxiety and raised affective filter (Pappamihiel, 2002). Schinke-Llano and Vicars (1993) found that most students in their study reported lower affective filter in student centered activities that involve negotiated interactions while also reporting being more uncomfortable during teacher focused activities. However, EB students may be more self-conscious about their own skills with language use, amplifying their affective filter, and contributing to being less likely to feel comfortable engaging in use of language that supports both their language and science knowledge learning.

Overall, these perspectives have multiple implications for how emerging bilingual students can best be supported in learning science. In the following paragraphs, we will explore contemporary approaches to science education that support all students' science learning, and

have the potential to specifically support EB students. Then, we explore the potential of arts-integrated science as a way to expand these opportunities for EB students.

Contemporary approaches to science education for EB students

Contemporary reforms in science education have focused on inquiry-based science learning as a way to authentically engage students in the practice of science (NGSS Lead States, 2013; NRC, 2012). Meaningful connections between prior knowledge and new knowledge may be actively generated through learning experiences, such as through sensemaking, problem-solving, and intentional student-centered activities (Brooks & Brooks, 2020; Hugener et al., 2009). For example, students in an inquiry science lesson may actively engage in learning activities, such as designing science experiments, engaging with real-world situations, exploring problems, explaining their understandings, elaborating their existing and new knowledge, and evaluating their own reflections during their learning experience (Dewey, 1938); (Bybee et al., 2006, 2015; Duffy & Raymer, 2010; Oliver, 2000).

EB students come into the classroom with rich experiences and interactions with the natural world, which can be used as resources for sensemaking in science, if teachers are able to tap into the affordances of these rich experiences. While inquiry-based instruction in which students are positioned as sense-makers provides an empowering context for building new science knowledge from students' previous experiences and ideas about scientific phenomena (Schwarz et al., 2017), teachers in inquiry-based learning contexts often treat their EB students' linguistic resources as their only access for disciplinary ways of communicating (taking a logocentric view), ignoring gestures, movement, or drawings, which limits the modalities for sense-making these students have available to them (Kusters et al., 2017). These issues are compounded by the fact that in science, teachers often privilege certain forms of language (Bang et al., 2012), particularly language that is considered to be "academic," standard English (Flores, 2020). This can prevent them from seeing the value in students' everyday ways of communicating or viewing these resources as inconsistent with academic science (Warren et al., 2001). These spoken or unspoken views about what counts as productive science language, or what academic science language should look and sound like may lead to more anxiety and higher affective filter for students who are learning science while simultaneously developing their proficiency in the English language.

The potential benefits of arts-integration in science for EB students

Arts integration in science can offer the support for learning across domains of both externally focused social processing and internally focused cognitive processes. A recent literature review conducted by Wahyuningsih et al. (2020), identified STEAM learning as a popular pedagogical methodology with evidence that supports early childhood education specifically to improve students' learning behaviors, such as creativity, problem-solving, scientific inquiry, critical thinking, and cognitive development. Considering the social processing of working in groups, students learning science might use an arts-integrated approach. For example, they may form small groups to utilize storytelling and bodily movements to role-play different stages of the plant life cycle, leveraging language and gestures to enact the scientific concepts mediated by their instructors (See Additional file 1: Table S3 and Video S2). Holzman (2016) further extends Vygotsky's research of play as a source for linking creative imitation with performance, asserting that creative imitation or modeling in a social environment of relating to oneself and others as a performer, dancer, speaker, or learner may provide meaningful STEAM learning for cognitive development (Martinez, 2017). Incorporation of arts-based instructional methods may provide additional means through which students can engage in science learning that draw upon students' ideas and creativity as a foundation for new learning (Hadzigeorgiou, 2016) in ways which support EB students' knowledge and language development (Lee et al., 2019), but do not privilege language over other ways of making meaning. Specifically, art offers additional opportunities for students to communicate their ideas with their peers in ways that leverage all of their different linguistic and semiotic resources for meaning-making (Kusters et al., 2017). In addition, using arts strategies engages students in the creative process, helping them take greater risks and decreasing anxiety (Morgan & Stengel-Mohri, 2014) and lowering students' affective filter and supporting new language acquisition (Krashen, 1983). Arts-integration in science inherently invites students to engage in science in ways that transcend typical boundaries between what is considered "academic" or "appropriate" in a science classroom and instead invites more creative ways to express ideas. Wong et al. (2022) found that teachers who went through a 10-week professional development training on how to integrate arts with NGSS science education recount a major theme identified was how the arts and dance really helped their students engage with vocabulary and retain the science information. Thus, integration with art opens up classroom spaces in which teachers can leverage and value the various meaning-making resources students enter the classroom with, inclusive of the various languages they already know or are learning, in addition to semiotic modes for communicating ideas and constructing new knowledge. By opening up science to student creativity in these ways, art can increase students' engagement with science in ways that lower their affective filter and reduce anxiety around communicating their science ideas.

Examining the more internally focused cognitive processes of arts, we can see how arts may assist learners in making connections between concepts and ideas in ways that support the assimilation and accommodation of new ideas into existing schemas initially described by Piaget (1971). The idea that the brain and body work together suggests that their connections produce higher cognition than would result from one process alone (Shapiro, 2014). Theory on embodied cognition (EC) suggests that the physical actions we perform ourselves shape our mental experience, in addition to the actions being performed by others near us (Lakoff & Johnson, 1999; Niedenthal, 2007). The initial definition of EC by Varela et al. (1991), explains cognition as embodied action that involves sensorimotor capacities embedded in larger biological, psychological and cultural contexts. Extending this, Bube (2021) explored the educational potentials of embodied art reflection, underscoring the importance of refining perception and attention through arts within pedagogical contexts of EC and arts in school settings. For example, in this study (See Additional file 1: Table S6 and Video S2), embodying scientific concepts may take the form of students' performing kinesthetic bodily movements within learning groups to use their bodies to create movements that convey heavy, medium, or light to represent gasses with different masses. Having students creatively generate and physically move toward these different concepts in an open classroom environment presents an alternative modality for students within sociable groups to visually and physically manipulate their bodies to generate opportunities for scientific sensemaking. Theorists of embodied cognition promote learning through multimodal representation employing bodily enactments of learning content (Paas & Sweller, 2012; Skulmowski et al., 2017). One such way to operationalize embodied cognition in an elementary educational context is through the interdisciplinary facilitation of an arts-integrated learning approach to teaching science (Agostini & Francesconi, 2021).

Arts may provide a rich and previously untapped classroom resource for both embodied cognition and social ways of processing science learning. Taking the perspective of science learning as a process in which students should be positioned as knowers and doers of science, who are capable of engaging in authentic science

practices (NGSS Lead States, 2013; NRC, 2012), we argue that there is a need to encourage and foster the creative side of science learning (Hadzigeorgiou, 2016), as a way to increase equitable learning opportunities for EB students. Traditional pedagogical approaches in science curricula have given attention to the necessary teaching tools for communicating and investigating scientific results, but have done less to provide support for the aesthetic tools, or students' imaginative engagement and transformative experiences required to conduct science (Hadzigeorgiou, 2016). Imaginative engagement in science education extends beyond the traditional teaching and learning context, fostering creativity, critical thinking, and problem-solving (Hadzigeorgiou et al., 2012). Like a scientist navigating from a list of observational data, to experimental design, and the interpretation of results—this complex process requires intuition, ingenuity, and imagination driving the scientific sensemaking (Kind & Kind, 2007). Eisner contends that "many of the most complex and subtle forms of thinking take place when students have an opportunity to work meaningfully on the creation of images... or to scrutinize them appreciatively" (Eisner, 2002, pp. xi–xii).

Both visual arts (i.e., painting, drawing, sculpturing) and performing arts (music, dance, drama) techniques offer pedagogical affordances that extend beyond traditional approaches toward learning science. Integrating the arts in science may provide students with a way to internalize concepts, process information, visualize and develop the ability to think metaphorically, and such "metaphor creates a space in human cognition, where individuals are free to rehearse new ideas of expression and form" (Efland, 2004, p. 757). In addition, contending with the vocabulary intensive nature of science content, arts integration in elementary classrooms has been shown to significantly increase both engagement and oral language skills for EB students (Brouillette et al., 2014). This is supported by research across content areas, which shows that art-integration has a positive impact on EB students' academic achievement in English language arts (Peppler et al., 2014), and multimodal research supports that integrating arts techniques, such as movement, gesture and expression into elementary classrooms specifically boosts language comprehension and memory of EBs (Peregoy & Boyle, 2008); (Gersten & Geva, 2003; Hardison & Sonchaeng, 2005; Kress, 2009; Rieg & Paquette, 2009).

Order of integrating arts and inquiry

Little is known about how to integrate arts and inquiry methods, particularly whether to lead or follow-up with the arts, before or after the inquiry methods. However, there is some research indicating potential efficacies from leading with the arts. English language learners become more fluent writers when using image creation as a pre-writing strategy and their vocabulary improves when students first express ideas through art (Andrzejczak et al., 2005). As conjecture based upon foundations within the theoretical framework discussed above, we hypothesized that leading science instruction with arts-integrated (STEAM-related) learning before implementing inquiry (STEM-related) learning, STEAM → STEM, will yield greater efficacies to EB learners than a STEM → STEAM order of integrating. We speculated that by leading with arts, in a STEAM → STEM order we are: decreasing the affective filter early in the instructional cycle; increasing the inclusive multimodal generation of an abundance of unfiltered new ideas for EB students to assimilate, which may include relatively more new concepts for EB students through a less filtered acquisition phase; following up in the learning cycle when these students consequently engage in inquiry and they are able to accommodate the new concepts, filtering their relative wealth of ideas through the inquiry process of testing before accommodating only those concepts that pass through the heightened scrutiny of the inquiry process. To put it more plainly, the STEAM → STEM order may produce a more robust, creative generation of ideas in a context which invites more students to participate in that idea generation, which students can then apply and funnel down as they engage in inquiry, ultimately solidifying learning through the process of accommodation.

By comparison, we speculated that in the reverse sequence of beginning with inquiry, in a STEM → STEAM order we are: beginning the instructional cycle with a relatively higher affective filter; attempting to have students accommodate new ideas without first generating a wealth of new ideas to assimilate; proceeding to follow up the accommodation of inquiry with the generation of more concepts that are not subsequently filtered through the inquiry process and may lead to potential misconceptions and fewer new concepts effectively accommodated. In other words, the STEM → STEAM order may privilege certain ways of communicating and certain types of ideas which may narrow the field of available ideas students generate and then work from as they engage in inquiry, while later, as they engage in arts, students broaden their available ideas creatively, but then do not have the opportunity to apply those new ideas in ways that facilitate their individual accommodation.

The purpose of this study was to explore the following research questions:

RQ1) Does leading with STEAM lessons before STEM lessons increase student knowledge gains in life science and physical science?

RQ2) Does leading with STEAM lessons before STEM lessons increase student knowledge gains in life and physical science for emerging bilingual English learners?

RQ3) To what extent does implementation fidelity contribute to the impact of the order effect of STEAM before STEM teaching efforts pertaining to emerging bilingual English learners?

We hypothesized that implementing STEAM before STEM would increase the opportunities to learn science for both EF and EB students, and particularly so for the EB students as explained above, ultimately leading to higher learning gains. We further hypothesized that these impacts will be more pronounced in classrooms with higher fidelity of implementation of the STEAM and STEM methods.

Methods

Study and curricular context

This study emanates from a multi-year collaboration between a large research university, a county performing arts center, and multiple school districts in California. The study program involved the development of two comparison sets of elementary science curriculum, with one being inquiry-based and the other arts-integrated that both covered the same NGSS performance expectations but differed in their pedagogical approaches. The program's curriculum included elementary grades three–five earth, life and physical sciences, of which this study reports on fifth grade life and physical science classroom student and teacher data. Comprehensive professional development training was provided to all teachers involved by experts in the fields of inquiry
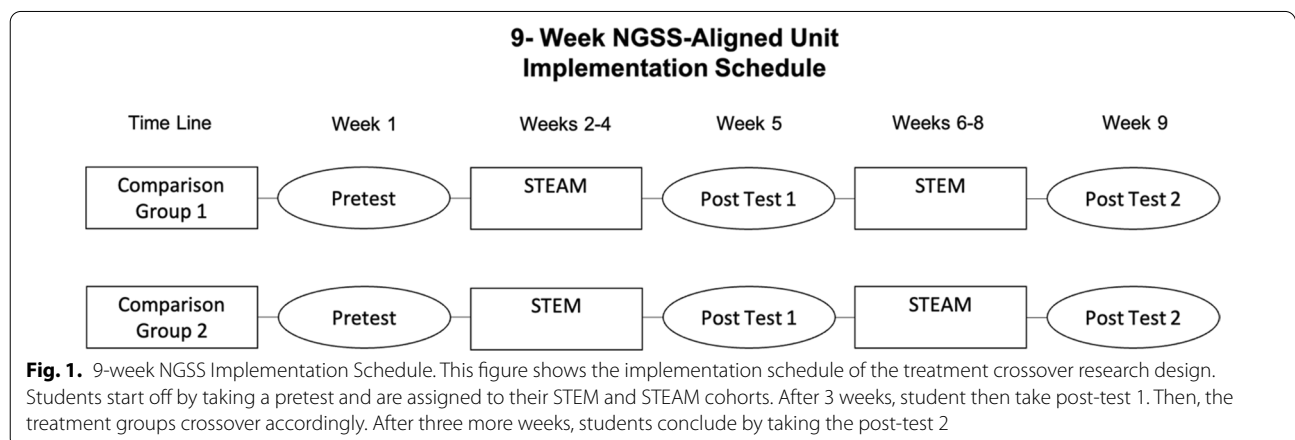
and arts NGSS curriculum. These units were designed as rigorous controlled comparison treatments to facilitate experimental measurement of randomized assignment of two alternative orders of integration of arts → inquiry or inquiry → arts, which we refer to as STEAM → STEM or STEM → STEAM, respectively.

NGSS-aligned instructional units

The NGSS-aligned life and physical science units each include lessons that are delivered through two approaches: (1) STEM and (2) STEAM lessons. The complete intervention spanned a 9-week timeline as illustrated below (see Fig. 1). The STEM lessons used guided inquiry as the main instructional framework, designed by science education experts and aligned to NGSS, include hands-on laboratory experimentation involving students developing questions, hypothesis testing, variable observations, measurements, analysis of results and drawing conclusions while addressing crosscutting concepts of patterns, cause and effect, scale, proportion, and quantity; systems and system models; energy and matter; structure and function; and stability and change. The STEAM lessons addressed the same NGSS performance expectations in addition to specific elementary level visual arts and dance standards, replacing the guided inquiries as the alternative instructional framework utilizing arts instead. Designed by arts education experts, the STEAM science lessons include an embedded focus on art elements, such as axial and locomotor movements, pathways, levels, and shapes in dance, as well as color, lines, shapes, and perspective in visual art.

STEM lessons

These fifth grade lessons addressed NGSS Science standard 5-LS2-1 in which students develop a model to describe the movement of matter among plants, animals, decomposers, and the environment. The Disciplinary

Fig. 1. 9-week NGSS Implementation Schedule. This figure shows the implementation schedule of the treatment crossover research design. Students start off by taking a pretest and are assigned to their STEM and STEAM cohorts. After 3 weeks, student then take post-test 1. Then, the treatment groups crossover accordingly. After three more weeks, students conclude by taking the post-test 2

Core Idea LS2.A covered the interdependent relationships in the ecosystems, focusing on organisms that are related in food webs in which animals eat plants and other animals eat the animals that eat plants. Through the NGSS-aligned STEM lessons students learn to understand and accept that air is matter, because it takes up space and has mass. Thus, students learn that gasses, being matter, have weight (mass), and all gasses do not have the same weight (mass). Each lesson involved one or more hands-on activities designed to engage students in scientific learning. These activities were typically an experiment in which students collected and analyzed data, recorded observations, and then discussed the results. For example, in one of the STEM lessons students played a predator/prey simulation game to learn about the variables related to the balance of producers and consumers in a food web and a biomass pyramid. They played several rounds and after each round modified the number of consumers so that there was a proper balance of predators and prey. Additional information about the STEM lessons can be found in the Additional file 1. NGSS standards covered in each lesson can be found in Additional file 1: Tables S1–S13.

STEAM lessons

In contrast, the NGSS-aligned STEAM unit lessons included both visual arts and the performing art of dance (VAPA) components in place of the STEM inquiry activities. Each lesson employed visual arts strategies to introduce a concept, then dance strategies to kinesthetically explore the concept further. The visual arts strategies employed throughout the STEAM lessons included: student-led pictorials generated by the students, observational drawings, active-NGSS unit art (using cut up pieces of paper to represent science vocabulary words or processes), and gesture drawings depicting student-generated conceptually symbolic gestures. Further, kinesthetic dance movements developed collaboratively by the students were directly and meaningfully associated with both arts and science vocabulary terms, as well as interpretation and analysis of visuals (Additional file 1: Video S2). While many of the dance movements are student generated, some specific teacher facilitated elements of dance used in the lessons included levels, pathways, patterns, shapes, axial and locomotor movements. These dance aspects of the STEAM lessons also addressed the same NGSS standards mentioned above in addition to addressing dance standards, such as demonstrating focus, physical control (e.g., proper alignment, balance), and demonstrating cooperation, collaboration, and empathy in working with partners and in groups (e.g., leading/following, mirroring, calling/responding, echoing, opposing), with the inquiry portions removed. In the

corresponding dance lesson students created a Biomass Pyramid Dance to demonstrate the relationship of the numbers of organisms at each level of the pyramid, representing how the energy is transferred up the pyramid and the consequence on an unbalanced pyramid (Additional file 1: Video S1). Additional information about the STEAM lessons can be found in the Additional file 1. NGSS, dance, and visual arts standards covered in each lesson can be found in Additional file 1: Tables S1–S6.

Professional development

A 5-day, week-long professional development (PD) institute took place in the summer prior to the implementation to prepare the teachers for instruction of the lessons. Teachers participated in equally immersive PD for both the three lesson STEAM unit and three lesson STEM unit to be implemented in the upcoming academic year. The STEM unit PD sessions were facilitated by teams of science curriculum developers and exemplary science teachers who were recognized for their expertise in content and in-service pedagogical training skill.

During the training, teachers were able to experience the different components of the lessons they would be teaching and they were provided with the full unit lesson plans.

Teachers participating in the program received 40 h of professional development in STEAM units and STEM units over a week-long Summer institute as well as 20 h of follow-up professional development to support implementation of the corresponding curriculum during the subsequent school year. Each year of the program introduced a new content area with year 1 focusing on earth, year 2 on life, and year 3 on physical science with 2 years of implementation on each. The current study focuses on the first year of physical science curriculum implementation and the second year of life science implementation.

In the effort to focus the tools of research toward increasingly pertinent areas of improving educational practice, it is important to identify teachers' efficacies to adopt and implement new professional development curricula learned when considering the dynamic complex ecologies of the everyday classroom context. With teachers in our study receiving over 40 h of intensive professional training, being well above the 15 h average that Estrella et al. (2018) typically found adequate to lead to positive treatment effects of PD programs, the extra training time included attending to classroom dynamics that are often difficult to feature meaningfully in a professional development training. For example, in our PD training teachers first experienced each lesson from the students' perspective before exploring pedagogical strategies from the teacher perspective, which research shows supports teachers in feeling more comfortable and

confident in their abilities to implement the curriculum in their own classrooms (Abd-El-Khalick, 2013; Gillies & Nichols, 2015; Yoon et al., 2012).

Experimental design

A longitudinal Pre–Post-Delayed Post-assessment design was employed to collect outcome measurements before, during, and after the learning interventions (Craig et al., 2012). We chose this within-subject treatment crossover measurement method in which all student participants in this study underwent both the STEM and STEAM learning conditions. Selecting this (Pre/Post/Post-test) design afforded the opportunity to explore how the order of augmenting science instruction integrated with the arts may increase science learning and equitability for emergent bilingual (EB) and English fluent (EF) students. Two randomly assigned groups of participants performed the same tasks in reverse order from one another, thereby allowing researchers the ability to monitor the effects of order in addition to changes of instructional gains over time (Crowder & Hand, 2017; Jones & Kenward, 1989). Two classroom level randomly assigned cohorts (STEM→STEAM or STEAM→STEM) eventually received both types of instructional units covering the same content, so they only differed in the implementation order, to assess whether leading with STEAM lessons provides greater benefit to EB students than to EF students compared to leading with STEM lessons (see Fig. 1).

Students were assessed at Time T1 (pre-intervention), Time T2 (post-intervention 1, after the first 3 lessons) and Time T3 (post-intervention 2, after completing 6 lessons) across 9 weeks (see Fig. 1). Treatment amounts were held constant between the STEM and STEAM units, and designed in recognition of the temporal limitations of teachers balanced with sensitivity to provide treatment levels that produce significantly measurable gains without over-teaching the concepts. Treatment frequency, quantity and duration were determined based on previous pilot interventions conducted by the authors. The treatment effects of implementing units with only three scaffolded lesson-sets preserves room for further learning gains that can be measured when combining two units of treatment. Hence, when the STEM→STEAM or STEAM→STEM unit implementations double the overall treatment magnitude the learning gains are still measurable with the total of six lessons within the combined units while measuring effects of their order of implementation. In other words, supposing each treatment had been five lessons instead of three lessons, then our previous trials suggested that the sensitivity of the measures might have been less capable of discerning the order effects, since learning saturation could already have been reached prior to any additive treatment.

Participants

Participants for this study were included from nine Southern California schools, typically contributing multiple classrooms for a total of 16 classrooms that were randomly assigned to receive the intervention in either of two patterns, specifically differing in the order of integrating the lessons with either the STEAM unit before the STEM or in the alternative order with the STEM unit first. The random distribution in this study is advantageous for its repeated measures design in which order effect is specifically analyzed while statistically reducing 'teacher' bias as each teacher randomly tests at all points across all treatments. This study design features a distribution pattern in which the life science treatment was administered to (N = 149) fifth graders in eight classrooms, while the physical science treatment was administered to (N = 152) fifth graders in the other eight classrooms. 83 of the 149 life science students and 86 of the 152 physical science students were designated as EB students. With half of these classrooms being randomly assigned to either implement the STEAM-first or STEM-first instructional method, based upon the classroom a student was assigned to, we categorized students as either STEAM before STEM or STEM before STEAM. This enabled us to precisely estimate the order effect or impact of one integrating approach vs the other.

Measures

We utilized two researcher-designed assessment tools, one based on the fifth grade life science NGSS and the other based on the fifth grade NGSS for physical science, with the key outcome variables of interest for this study being to determine gains in science knowledge. The first iterations of the grade level science knowledge tests were developed by the program's instructional design team of science experts. These versions for fifth grade were further refined after sending the science knowledge tests to another collection of external experts to provide feedback and suggestions. As a result, face and content validity was established for these two tools to be scored and utilized in the analysis, and a series of pilot tests offered further support for test–retest validation (construct validity) with alpha reliability (internal consistency coefficient) ranging from 0.68 to 0.77. The assessment questions administered can be found in the Additional file 1. Additional nominal and ordinal data collected on implementation levels and school level testing related to language fluency were also used to create grouping variables to test for interaction effects, that is, if there were any differences in impact for different groups of students.

As a part of accountability and evaluation of the large grant funded program, various aspects of teacher implementation were measured and serve to help to reflect upon the nature of teacher adoption of the new approaches to innovative methods of integration and account for fidelity in the analyses. Understanding the nature of educational reform includes the recognition that shifts in teacher practice occur over time and involve a variety of differentiated attitudes and approaches to adopting new strategies that may be beneficial to consider in larger scale efforts at change. Fidelity of implementation (FOI) characterizes the determination of how well a novel approach is implemented according to its original design (Lee et al., 2009; Mowbray et al., 2003). Measuring FOI, especially in the case of a new curriculum being tested, offers researchers, curriculum designers, and PD experts a glimpse into what went right, and arguably more importantly, what may have gone wrong or was inadequate (Dusenbury et al., 2003; Raudenbush, 2007).

Data scoring and analysis plan
Students were assessed three times, at pretest, between and after implementation of the STEM and STEAM units, depending on the order of implementation (either pretest → STEM → post-test   1 → STEAM → post-test 2, or pretest → STEAM → post-test 1 → STEM → post-test 2). The dependent variable was students' change scores, a measurement documenting the change in science knowledge obtained from pretest and then subsequent administration of two post-test assessments. These change scores consist of increases or decreases in life or physical science knowledge from pretest to post-test 1 (i.e., Post-test 1 minus Pretest = PrePost1 Change Score), and increases or decreases in life or physical science knowledge from pretest to post-test 2 (i.e., Post-test 2 minus Pretest = PrePost2 Change Score). We calculated total increases or decreases in life or physical science knowledge (i.e., PrePost1 Change Score plus PrePost2 Change Score = total increase in science knowledge) to be our dependent continuous variable for the analyses. To assign grouping variables for fidelity of implementation we utilized implementation logs kept on participating teachers during the study. By monitoring the activity and implementation levels of the teachers (e.g., assessing the dosage 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 to designate which teachers documented high (1) vs moderate to low (2) implementation.

Three independent-samples *t* tests were performed to determine significant differences between the change scores of the participating fifth grade students' gains in life and physical science knowledge. The first *t* test compared student change scores based on ordering of instructional approaches taught. Next, a *t* test was conducted to compare the change scores with students categorized as either English fluent (EF) or emerging bilingual (EB). Finally, change scores were compared by implementation fidelity (high vs moderate to low). The results of the *t* tests provide us with a first level view for documenting whether STEAM before STEM would lead to significantly higher knowledge gains.

To control for Type I error, and not rely only upon a series of disconnected *t* tests, while statistically controlling for pretest scores, a confirmatory three-way between-groups analysis of covariance (ANCOVA) was conducted to explore the impact of these three independent variables measuring the implementation fidelity, English language fluency and the instructional model utilized, pertaining to STEAM first or STEM first order of approaches. To assign grouping variables for fidelity of implementation, we utilized implementation logs that participating teacher documented during the study. By monitoring the activity and implementation levels of the teachers (i.e., assessing the dosage of well-delivered completed lessons provided as part of this study's intervention being provided to students) we were able to measure and designate which teachers documented high (1) vs moderate to low (2) implementation. On a scale of one to six, the range of scores all placed between three to six. Therefore, with no teachers showing lower scores (i.e., ones or twos), and to also have comparable sample sizes and large enough cell sizes in each category, a median split was conducted such that those scoring five to six were designated high, and those scoring three to four were designated moderate to low. In addition, to compare the effect of the curriculum and delivery methods pertaining to students designated as EB learners, the participants were divided into two groups based on the English language proficiency measured by the school district (EB vs EF), as well as being further assigned to two subgroups based upon whether the curriculum was delivered using STEAM first or STEM first. Pretest scores were administered to document the participants' initial knowledge of life and physical science were used as covariates to control for individual differences.

Results
Life science findings
A series of Independent-samples *t* tests were performed to study several grouping variables and a continuous variable assessing the change scores of the participating fifth grade students' gains in life science knowledge during the 9-week intervention. The first independent *t* test was conducted to compare a dependent variable

assessing student change scores (increase in life science knowledge from pretest to the final post-test) and an independent variable designating instructional method order, if students first received the life science curriculum using a STEAM/STEM order approach or STEM/STEAM order approach. There was a significant difference in change scores for STEAM/STEM classroom students ($m = 6.00$, SD $= 4.45$) and STEM/STEAM classroom students ($m = 2.93$, SD $= 3.91$; t (178) $= 5.50$, $p = 0.001$). The magnitude of the differences in the means (mean difference $= 3.40$, 95% CI 2.18–4.62) was large (eta squared $= 0.15$).

The next independent-samples t test was conducted to compare change scores in life science knowledge gains to students' English language fluency level. Based on English proficiency testing collected by the participating school districts, students were categorized as either EF or EB learners. There was a significant difference in change scores for EF students ($m = 5.36$, SD $= 4.18$) and EB students ($m = 3.25$, SD $= 4.35$; t (147) $= 2.99$, $p = 0.003$). The magnitude of the differences in the means (mean difference $= 2.11$, 95% CI 0.72–3.50) was a moderate effect (eta squared $= 0.06$).

An independent-samples t test was performed to explore and compare student change scores and an independent variable assessing implementation fidelity levels by grouping students into either of two categories: 1) high or 2) moderate to low. There was a statistically significant difference in change scores between the high implementation classroom students ($m = 4.60$, SD $= 4.61$) compared to the moderate to low implementation classroom students ($m = 3.10$, SD $= 3.57$; t (100) $= 2.68$, $p = 0.009$). The magnitude of the differences in the means (mean difference $= 1.79$, 95% CI 0.46–3.11) was small (eta squared $= 0.03$). The t tests documented that high implementation of such curriculum in classrooms were significantly more effective than the classrooms with moderate to low implementation levels. For more details on the descriptive statistics related to these outcomes see Table 1 below, which illustrates that this was the case for those in both the high and mid to low implementation categories as well as EF or EB learner categories.

A confirmatory three-way between-groups analysis of covariance (ANCOVA) was conducted to explore the impact of these three independent variables measuring implementation fidelity, English language fluency, and the instructional order method utilized pertaining to STEAM and STEM order approaches. The dependent variable was a measurement documenting the increase in life science knowledge obtained from pretest and then subsequent administration of two post-test assessments administered at 3-week intervals during the 9-week intervention.

Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variance, and reliable measurement of the covariate. After statistically adjusting for the pretest scores assessing preexisting baseline life

Table 1 Descriptive statistics via ANCOVA analysis of implementation levels, English fluency and instructional method related to total change scores for life science knowledge gained

Implementation fidelity	Language fluency	Instructional method	Mean	Std. deviation	N
High implementation fidelity	English fluent	STEAM/STEM	7.29	4.49	24
		STEM/STEAM	4.64	4.07	22
		Total	6.02	4.45	46
	Emergent bilingual	STEAM/STEM	6.11	4.97	18
		STEM/STEAM	2.50	3.84	44
		Total	3.55	4.47	62
	Total	STEAM/STEM	6.79	4.68	42
		STEM/STEAM	3.21	4.018	66
		Total	4.60	4.61	108
Low to moderate implementation fidelity	English fluent	STEAM/STEM	4.00	3.44	14
		STEM/STEAM	3.50	2.07	6
		Total	3.85	3.05	20
	Emergent bilingual	STEAM/STEM	5.00	3.465	5
		STEM/STEAM	1.56	3.815	16
		Total	2.38	3.945	21
	Total	STEAM/STEM	4.26	3.385	19
		STEM/STEAM	2.09	3.495	22
		Total	3.10	3.57	41

science knowledge, there was a significant interaction effect between level of implementation fidelity and a student's English fluency, F (1, 140) = 4.44, $p < 0.037$, with a small effect size (partial eta squared = 0.031). None of the other interaction effects between implementation fidelity, language fluency and instructional method were significant. There was a statistically significant main effect for instructional method order: F (1, 140) = 5.82, $p = 0.017$, with a small effect size (partial eta squared = 0.04). Neither language fluency nor implementation fidelity, however, produced a highly statistically significant main effect. The possibility exists that the small sample size (e.g., cell sizes) played a part in not producing more significant interaction and main effects.

Physical science findings

To further explore the above findings connected to life science instruction and the impact of implementation levels, language fluency and instructional methods, a series of Independent-samples t tests also were performed to study the change scores of the participating fifth grade students' gains in the physical science knowledge during the 9-week intervention. The first independent t test was conducted to compare student change scores (increase in physical science knowledge from pretest to the final post-test) and an independent variable designating if students received the physical science curriculum using a STEAM/STEM approach or STEM/STEAM approach. There was a statistically significant difference in scores for STEAM/STEM classroom students ($m = 6.08$, SD = 4.44) and STEM/STEAM classroom students ($m = 2.59$, SD = 4.60; t (182) = 5.68, $p = 0.001$). The magnitude of the differences in the means (mean difference = 3.62, 95% CI 2.35–4.90) was large (eta squared = 0.15).

The next independent-samples t test was conducted to compare change scores in physical science knowledge to a students' fluency level of the English language. Based on English proficiency testing collected by the participating school districts, students were categorized as either EF or EB learners. There was a significant difference in scores for EF students ($m = 4.82$, SD = 4.62) and EB students ($m = 3.31$, SD = 4.50; t (150) = 2.02, $p = 0.045$). The magnitude of the differences in the means (mean difference = 1.50, 95% CI 0.032–2.98) was small (eta squared = 0.026).

An independent-samples t test was performed to explore and compare student change scores and an independent variable assessing implementation fidelity levels by grouping students into either of two categories: 1) high or 2) moderate to low. There was a significant difference in scores for the high implementation classroom

students ($m = 4.29$, SD = 4.90) compared to the moderate to low implementation classroom students ($m = 3.10$, SD = 3.57; t (105) = 2.17, $p = 0.033$). The magnitude of the differences in the means (mean difference = 1.47, 95% CI 0.12–2.82) was small (eta squared = 0.025). For more details on the descriptive statistics related to these outcomes, please see Table 2.

To control for Type I error, a confirmatory three-way between-groups analysis of covariance (ANCOVA) was conducted to explore the impact of these three independent variables measuring the implementation fidelity, English language fluency and the instructional method model utilized pertaining to STEAM and STEM-based approaches. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variance, and reliable measurement of the covariate. After adjusting for the pretest scores assessing preexisting physical science knowledge, there was not statistically significant interaction effects between level of implementation fidelity, language fluency and instructional method. There was, however, a statistically significant main effect for the STEAM/STEM order of instructional method: F (1, 143) = 4.32, $p = 0.037$, with a small effect size (partial eta squared = 0.03). There also was a near statistically significant main effect for implementation fidelity: F (1, 143) = 3.45, $p = 0.065$, with a small effect size (partial eta squared = 0.024). Language fluency did not have a significant main effect.

Findings summary

In general, the results and descriptive statistics above suggest that a STEAM first approach can be beneficial to future life and physical science instruction efforts with both English fluent (EF) students and emerging bilingual learners (EB). The Independent-samples t tests revealed that: implementation order of STEAM→STEM was significantly better in life and physical sciences for both EF and EB students than the reverse order; EF students' learning gains were significantly higher than EB with STEAM first in life and physical sciences; and in life and physical sciences there were significantly higher change scores with higher implementation fidelity.

As Tables 1 and 2 illustrate, the STEAM first approach produced higher mean scores for the gains in knowledge in both high and low to moderate fidelity implementation settings. For both EF and EB students in both life and physical sciences, the effectiveness of having STEAM before STEM is greater in classrooms with high fidelity implementation as compared to classrooms with moderate to low fidelity implementation, to varying extents. To further consider these learning patterns, we'll express the differentially beneficial learning gains for implementing

Table 2 Means analysis of implementation levels, English fluency and instructional method related to total change scores for physical science knowledge gained

Implementation fidelity	Language fluency	Instructional method	Mean	Std. deviation	N
High implementation fidelity	English fluent	STEAM/STEM	7.33	4.48	21
		STEM/STEAM	3.48	5.07	25
		Total	5.24	5.13	46
	Emergent bilingual	STEAM/STEM	6.50	4.89	20
		STEM/STEAM	2.33	3.95	45
		Total	3.62	4.65	65
	Total	STEAM/STEM	6.93	4.64	41
		STEM/STEAM	2.74	4.38	70
		Total	4.29	4.90	111
Low to moderate implementation fidelity	English fluent	STEAM/STEM	4.00	3.44	14
		STEM/STEAM	3.50	2.07	6
		Total	3.85	3.05	20
	Emergent bilingual	STEAM/STEM	5.00	3.46	5
		STEM/STEAM	1.56	3.81	16
		Total	2.38	3.94	21
	Total	STEAM/STEM	4.26	3.38	19
		STEM/STEAM	2.09	3.49	22
		Total	3.10	3.57	41
Total	English fluent	STEAM/STEM	6.00	4.37	35
		STEM/STEAM	3.48	4.61	31
		Total	4.82	4.62	66
	Emergent bilingual	STEAM/STEM	6.20	4.62	25
		STEM/STEAM	2.13	3.90	61
		Total	3.31	4.50	86
	Total	STEAM/STEM	6.08	4.44	60
		STEM/STEAM	2.59	4.18	92
		Total	3.97	4.60	152

STEAM → STEM compared to STEM → STEAM as the 'STEAM first order effect advantage.' Notably, the STEAM first order effect advantages trend differently between EF and EB students at different fidelity levels:

(1) In high fidelity implementing life science classrooms the trend differences are relatively subtle (EF 2.65, EB 3.61), with EF students' mean change score for STEAM first was 7.29 compared to STEM first with 4.64 (EF Mean Difference = 7.29–4.64 = 2.65), while EB students showed a mean change score for STEAM first at 6.11 compared to STEM first at 2.5 (EB Mean Difference = 6.11–2.50 = 3.61).

(2) In moderate to low fidelity implementing life science classrooms EF students show a large drop in their STEAM first order effect advantage (0.50) with the mean change score for STEAM first was 4.00 compared to the STEM first change score of 3.50 (EF Mean Difference = 4.00–3.50 = 0.50),

while contrastingly, the EB students maintain their relatively high STEAM first order effect advantage (3.44) with the mean change score for STEAM first at 5.00 compared to the STEM first change score of 1.56 (EB Mean Difference = 5.00–1.56 = 3.44).

(3) Physical science classrooms reveal the same pattern of a large STEAM first order effect advantage for EF students (3.85) in high fidelity implementing classrooms, with the difference in mean change between STEAM → STEM and STEM → STEAM was (EF 7.33–3.48 = 3.85), compared to a low STEAM first order effect advantage (0.50) in moderate to low fidelity implementing classrooms (EF 4.00–3.50 = 0.50). While again, we see EB students maintaining their relatively high STEAM first order effect advantages in both high and lower fidelity implementations (4.17, 3.44), observing that in high implementing physical science classrooms the difference in mean change between STEAM → STEM and STEM → STEAM was similar (6.50–

2.33 = 4.17) compared to results (5.00–1.56 = 3.44) in moderate to low fidelity implementing classrooms.

Discussion

This study sought to investigate how to integrate STEAM and STEM methods, focusing on the order in which these approaches are combined and implemented within elementary life and physical science courses while considering the fidelity of implementation by the teachers and also the English language competencies of the students. Assessing whether test scores showed statistically significant gains, we performed a series of means analyses to document such differences. A rigorous examination of STEAM and STEM integration methods may be instrumental in identifying specific integration practices for improving science instruction through increasing opportunities to learn and raising student science achievement. The results of this study suggest that patterns of integration that lead with STEAM (arts-integrated science lessons) before STEM (inquiry-based science lessons) approaches do increase gains in life and physical science knowledge significantly for both English fluent (EF) and emerging bilingual (EB) students generally, providing more opportunities to learn science for all students. While it may not be surprising that overall learning gains were greater for EF students compared to EB students, it is particularly interesting that EB students consistently demonstrated greater STEAM first order effect advantages, illustrating the potential for the STEAM first order to offer more equitable opportunities to learn science.

RQ1) Does leading with STEAM lessons before STEM lessons increase student knowledge gains in life science and physical science?

Comparison of integration order of the STEAM and STEM lessons revealed large statistically significant differences for both life and physical sciences. The *t* tests and ANCOVA results suggest that a STEAM first approach is significantly more productive in increasing life science knowledge gains for EF and EB students, with higher overall gains for EF students, and with nearly identical patterns for physical science results. Seeing this clear pattern across the two very different scientific disciplines offers additional verification that integrated design is important to consider in elementary science instruction. We hypothesized such a result, speculating (see section Order of Integrating Arts and Inquiry for full rationale) that implementing with STEAM first decreases affective filter and increases inclusive multimodal generation of new ideas to assimilate during the early STEAM art phase, leading to a more abundant set of concepts to accommodate during the later STEM inquiry phase. We also speculated that the STEM first order would be less effective, since it begins with a relatively higher affective filter that squelches the generation of new ideas to assimilate in the early STEM inquiry phase and then follows up in the later STEAM phase with generating new ideas that are not subsequently filtered through an inquiry phase.

While a conclusive explanation for the consistent efficacy of integrating science instruction with arts (STEAM) before inquiry (STEM) will require additional study, it may also be partially explained by the involvement of gesturing, which is thought to facilitate the learner's representations of problems to be translated into perceptual and motor information, making it more readily available when solving problems (Goldin-Meadow & Alibali, 2013). Interestingly, it has been suggested that gesturing may provide scaffolding for learning in the future (Brooks & Goldin-Meadow, 2016). Therefore, perhaps leading with STEAM methods (STEAM → STEM), involving visual and performing arts (VAPA) gestures for scientific information, is efficacious due to the scaffolding of STEAM that is then leveraged during the next stage of the STEM method implementation, rather than vice versa when the scaffold of arts would come too late (STEM → STEAM) within the implementation integration sequence to build upon productively.

RQ2) Does leading with STEAM lessons before STEM lessons increase student knowledge gains in life science and physical science for emerging bilingual English learners?

Both life and physical science showed significant differences in scores between EF students and EB students, with a moderate effect for life science and a smaller effect for physical science. We observed higher impacts of the STEAM first order effect advantage for EB students than EF students, which is perhaps remarkable in that this shows potential for increasing equitable opportunities to learn among EB and EF through the STEAM → STEM instructional method. While our theoretical framework offers some potential explanation for the advantages to EB students for beginning STEAM methods, considering that sociolinguistic negotiations may offer a more open modality for students to initially conceptualize abstract science concepts by breaking down the barriers, such as difficult vocabulary and causal relationships through low pressure arts-based activities. Then, when arts-integrated instruction is followed by inquiry-based learning activities, this sequencing or ordering of instruction may allow students to draw on their prior knowledge from the STEAM lessons and offer opportunities for rehearsal and revision of ideas, accommodating to further elaborate,

adjust, or enhance students' capacity for science content mastery.

Other potential mechanisms may include the embodied cognition and representations of science concepts in different modalities. Rather than reading text, simply filling in worksheets, or diving straight into an experiment, students are using movement, gestures, and stories to verbally and physically represent scientific vocabularies, processes, and developing conceptualizations. STEAM teaching methods have been found to further support scientific language development by decreasing cognitive load and making abstract concepts more concrete and accessible through multimodal representation (Campbell et al., 2018; Wahyuningsih et al., 2020). Researchers in the field of embodied cognition argue that kinesthetic movement and multiple representational activities lead to reduced cognitive load, increasing executive functioning resources, such as student's working memory capacity to further develop schema formation and identify potential conceptual misalignments (Goldin-Meadow & Alibali, 2013); (Begolli & Richland, 2016; Richland & Hansen, 2013; Wilson, 2002). It has also been proposed (Wellsby & Pexman, 2014) that by examining sensorimotor information in student's conceptual and linguistic understanding and determining when their thinking becomes less reliant on such sensorimotor knowledge and more abstract and complex, developmental researchers could help advance theories of EC generally for the future, and it may be fruitful to measure how such effects might be impacted by the pattern and implementation of integrating arts with science. With arts methods potentially inviting students, particularly students who struggle with the language, to more fully engage in science, teachers may have a very useful tool for integrating curriculum by specifically designing lessons that attract students by leading with STEAM and following with STEM to give their students the best foundation for learning challenging concepts of science, whether they are language fluent or EB.

RQ3) To what extent does implementation fidelity contribute to the impact of the order effect of STEAM before STEM teaching efforts pertaining to emerging bilingual English learners?

Taking the level of implementation into account revealed an interesting pattern related to language proficiency of students and the 'STEAM first order effect advantage' (See Findings Summary for full explanation) for EB students in particular. The life science findings in Table 1 and the physical science findings in Table 2, along with the Findings Summary illustrates that while EF students benefit from the STEAM first order effect advantage in the high fidelity classrooms, they do not benefit

significantly in the low to moderate fidelity classroom. EB students in particular, do benefit more highly than EF students from the STEAM first order effect advantage in both high and low to moderate fidelity. Perhaps levels of implementation fidelity might be explained by teacher's comfort and confidence in implementing the STEAM approach. Findings from Wong et al. (2022) showed that teachers felt more confident and self-efficacious to implement STEAM pedagogies in the classroom after completing a 10-week online professional development that specifically focused on supporting teachers' content knowledge and STEAM teaching perceptions. Drawing on a similar teacher–learner observer professional development model (Corrigan et al., 2022; Wong et al., 2022), this study also supported teachers in 40 h of face-to-face professional development on how to implement both STEAM and STEM pedagogical strategies with a major emphasis on supporting diverse learners in the classroom. As such, we might attribute the results of high fidelity implementing teachers and their success on EB and EF students to the quality of PD support teachers received. This in turned helped EB students experience this advantage to their learning gains in every classroom fidelity environment, which presents a noteworthy opportunity to learn science with increased equity.

Conclusions

Even at moderate to low levels of teaching implementation fidelity, the STEAM first approach produced higher mean scores for the gains in knowledge, although these impacts were far greater for EB than for EF students. While both EB and EF students benefit similarly and significantly in high fidelity implementation classrooms, the gains for EF students are not significant in low fidelity implementation classrooms. Yet, in such low fidelity implementation classrooms, the EB students still benefited significantly despite the poor implementation. Implementation fidelity correlation with increased efficacy leads to further verification that the STEAM method employed was an effective treatment if precisely administered. That the lower fidelity implementation of STEAM before STEM was still significantly effective for EB students despite poor implementation, may suggest that a strong compensating STEAM first order effect advantage is possibly involved in the implementation system for this EB population of learners specifically, supporting the claim that a STEAM first order effect increases the EB students' opportunity to learn science. If equity building is of concern to curriculum designers, then teaching science through the arts with STEAM lessons is an effective approach (Corrigan et al., 2022), and introducing STEM units with STEAM may effectively improve the outcomes

further for teaching life and physical sciences with the STEAM first order effect advantage.

With limitations such as ensuring greater implementation fidelity, small-to-large effect sizes identified and cell size issue challenges due to random sampling methods, future research controlling for such methodological limitations further exploring these variables will be beneficial to determining more conclusive empirical results on inquiry-based and arts-based approaches collectively. In addition, while working with a highly transient population of students in Title 1 schools reflecting wide variability between students' academic performance, we also encountered issues with large standard deviation values. Statistical challenges were amplified due to large variability ranging from very low to very high change scores, and future research would benefit from replicating this study in more schools. Future work to utilize fidelity of implementation scores should continue to seek to further quantify teachers' instructional implementation levels via implementation logs, scored observation visits, and cognitive interviews to glean additional insights from teachers to better understand teachers' adaptability of the STEM/STEAM lessons to support their learners' needs. While supports and scaffolds were included within each of the lessons and teacher scripts, further explanations on how teachers might have modified and iterated on their own lesson plans to enhance the learning experience to fit their diverse student body will certainly inform the scalability, efficacy, and design of the STEM/STEAM lesson plans developed.

Importantly, while a pre/post1/post2 measurement procedure was utilized in this study for assessing the effects of implementation order, a noteworthy limitation is that a follow-up delayed post-test was not deployed as a function for measuring retained learning over time without any additional treatment manipulations (Ramirez & Jones, 2016). The scope of this study assessed learning gains over the 9-week implementation period with the first post-test assessed STEAM vs STEM and the second post-test assessed STEAMfirst vs STEMfirst approach. Indeed, to properly claim and assess long term learning effects in these NGSS-aligned life and physical science course curriculums, future methodological iterations for data collections would incorporate a more traditional delayed post-test, in addition to treatment crossover postest2 examinations, to measure retained learning knowledge of each group over time (Haynie, 1997).

For the present, these results provide implications for how arts-integrated strategies might be best deployed in classrooms, schools, and districts to teach science (See Additional file 1: Tables S1 and S7). Many participating teachers reported having previously felt the pressure to integrate without knowing how to do it well. This studies' STEAM integration endeavor to improve science learning for both EB and EF students offers a rare glimpse into the design efficacies and specific intricacies of integration with rigorous evidence for leading with STEAM approaches prior to STEM. With implications for practice and sustainability in mind, this program offers an expansive fully online training for all its materials, which can demonstrably increase the effective fidelity of implementing for teachers who intentionally choose to adhere to the lesson plans and approaches provided. Access to teacher materials can be found in the Online Teacher Resources and Implementation Materials within the Additional file 1. In these early stages of research investigating STEAM → STEM approaches, we have conjectured that the order of implementation of STEAM first may leverage elements of such mechanisms as embodied cognition and lowered affective filter as ways of improving the learning of science, potentially through efficiently orchestrating processes of more optimal assimilation and accommodation for EF and EB student. While our theoretical framework of a particular blend of social–cognitive constructivism may have helped us form a working model for STEAM first maximizing assimilation and STEM last maximizing accommodation, as a conjecture for our hypotheses of the order effect advantages supported by the study results, we acknowledge that this initial study is far from understanding the causality of the many intricacies potentially involved in STEAM learning. However, we do submit that this study demonstrates an aspect of the promise that arts and science integration holds for increasing opportunities to learn science, with the STEAM first order effect advantage being an example of a more equitable approach for science curriculum.

While it could be an interpretation that strictly following the specific methods of this particular program is important for the efficacy of EF student learning, we encourage an additional interpretation and implication. Specifically, we note that it may be more important to emphasize our experimental finding that implementation fidelity appears to be less important for EB students' efficacies. Considering that EB students maintain high STEAM first order effect advantages, even in low to moderate implementation classrooms, opens the door to responsible implementation experimentation. As we have demonstrated in this study, there appears to be ample affordances for teachers to experiment with integration approaches of their own construct and purposely veering in fidelity for their adaptations as expert practitioners in their own right. We urge future experimentation from informal to rigorous and quantitative to qualitative designs to quickly garner innovations that can help promote learner and teacher success in facing the mounting

challenges of learning science. Together, educational researchers, curriculum designers, and practitioners may delve more productively into innovative integrative ways of knowing and learning science that are cognitively informed, empirically based, and practice-originated, confidently knowing that teachers are tapping into a promising area of pedagogical practice and research in applying the STEAM→STEM order effect with inclusive excellence for serving a diverse body of students more equitable opportunities to learn.

Notes: 1 While traditionally in the US emerging bilingual (EB) students, who are in the process of learning the English language in addition to one or more other languages, have typically been referred to with a deficit-oriented label of English language learner (ELL) or English learner (EL), a growing number of science educators 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; Suarez, 2020; Ünsal et al., 2018; Wilmes & Siry, 2020). In this report we refer to such students as emerging bilingual (EB) learners, as we acknowledge the challenge and accomplishment of learning more than one language and recognize a 'bilingualism asset' perspective. In addition, we recognize that EB students are highly heterogeneous across factors, such as SES, race, ethnicity, and the other languages they use (González-Howard & Suárez, 2021).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40594-022-00375-7.

> **Additional file 1.** Methodology and additional tables.

Availability of data and materials
The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Author details
[1]University of California Irvine, Irvine, CA, USA. [2]Marshall University, USA, Huntington, West Virginia, USA. [3]Concordia International School, Hanoi, Vietnam. [4]University of Texas at Austin, Austin, Texas, USA.

References

Abd-El-Khalick, F. (2013). Teaching with and about nature of science, and science teacher knowledge domains. *Science & Education, 22*(9), 2087–2107. https://doi.org/10.1007/s11191-012-9520-2

Afitska, O., & Heaton, T. J. (2019). Mitigating the effect of language in the assessment of science: A study of English-language learners in primary classrooms in the United Kingdom. *Science Education, 103*(6), 1396–1422. https://doi.org/10.1002/sce.21545

Agostini, E., & Francesconi, D. (2021). Introduction to the special issue "embodied cognition and education." *Phenomenology and the Cognitive Sciences, 20*(3), 417–422. https://doi.org/10.1007/s11097-020-09714-x

American Association for the Advancement of Science. (1994). *Benchmarks for science literacy*. Oxford University Press.

Andrzejczak, N., Trainin, G., & Poldberg, M. (2005). From Image to Text: Using Images in the Writing Process. *International Journal of Education & the Arts*, 6(12), 1–17. https://www.learntechlib.org/p/101199/.

Arends, R. I. (2015). *Learning to teach* (10th ed.). McGraw-Hill.

Babaci-Wilhite, Z. (2019). *Educational tools to teach STEAM subjects integrating linguistic rights, collaboration, and critical thinking*. In: *Promoting language and STEAM as human rights in education*. Springer. https://doi.org/10.1007/978-981-13-2880-0_1

Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desettling expectations in science education. *Human Development, 55*(5–6), 302–318. https://doi.org/10.1159/000345322

Begolli, K. N., & Richland, L. E. (2016). Teaching mathematics by comparison: Analog visibility as a double-edged sword. *Journal of Educational Psychology, 108*(2), 194. https://doi.org/10.1037/edu0000056

Betancur, L., Votruba-Drzal, E., & Schunn, C. (2018). Socioeconomic gaps in science achievement. *International Journal of STEM Education, 5*(1), 1–25. https://doi.org/10.1186/s40594-018-0132-5

Brooks, J. G., & Brooks, M. G. (2020). *In search of understanding: The case for constructivist classrooms* (2nd ed.). Pearson College Div.

Brooks, N., & Goldin-Meadow, S. (2016). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science, 40*(7), 1831–1849. https://doi.org/10.1111/cogs.12292

Brouillette, L., Childress-Evans, K., Hinga, B., & Farkas, G. (2014). Increasing engagement and oral language skills of ELLs through the arts in the primary grades. *Journal for Learning through the Arts*. https://doi.org/10.21977/D910119300

Bube, A. (2021). Educational potentials of embodied art reflection. *Phenomenology and the Cognitive Sciences, 20*(3), 423–441. https://doi.org/10.1007/s11097-020-09685-z.

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

Bybee, R., Taylor, J. A., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., & Landes, N. (2015). *The BSCS 5E instructional model*. NSTA press.

Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Powell, J. C., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins and effectiveness. *BSCS, 5*, 88–98.

Campbell, C., Speldewinde, C., Howitt, C., & MacDonald, A. (2018). STEM practice in the early years. *Creative Education, 9*(01), 11. https://doi.org/10.4236/ce.2018.91002

Catterall, J. S. (2009). *Doing well and doing good by doing art*. Imagination Group.

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

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S„ Ogilvie, D., Petticrew, M., Reeves, B., Sutton, M., & Thompson, S. (2012) Using natural experiments to evaluate population health interventions: New Medical Research Council guidance. Journal of Epidemiology Community Health, 66(12):1182–1186. https://doi.org/10.1136/jech-2011-200375.

Crowder, M. J., & Hand, D. J. (2017). *Analysis of repeated measures*. Routledge.

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/3IojpnU.

Dewey, J. (1934). *Art as experience* (p. 362). Minton, Balch & Co.

Donovan, L., & Pascale, L. (2012). *Integrating the arts across the content areas*. Teacher Created Materials.

Duffy, T.M., & Raymer, P.L. (2010). A Practical Guide and a Constructivist Rationale for Inquiry Based Learning. *Educational Technology archive, 50*, 3–15. https://www.jstor.org/stable/44429836

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research, 18*(2), 237–256. https://doi.org/10.1093/her/18.2.237

Efland, A. D. (2004). Art education as imaginative cognition 1. In E. W. Eisner & M. D. Day (Eds.), *Handbook of research and policy in art education* (p. 757). UK: Routledge.

Eisner, E. W. (2002). *The arts and the creation of mind*. Yale University Press.

Estrella, G., Au, J., Jaeggi, S. M., & Collins, P. (2018). Is inquiry science instruction effective for English language learners? A meta-analytic review. *AERA Open*. https://doi.org/10.1177/2332858418767402

Flores, N. (2020). From academic language to language architecture: Challenging raciolinguistic ideologies in research and practice. *Theory into Practice, 59*(1), 22–31. https://doi.org/10.1080/00405841.2019.1665411

García, O., & Wei, L. (2014). Translanguaging and education. In O. García & L. Wei (Eds.), *Translanguaging: Language, bilingualism and education* (pp. 63–77). UK: Palgrave Macmillan.

Gersten, R., & Geva, E. (2003). Teaching reading to early language learners. *Educational Leadership, 60*(7), 44–49.

Gillies, R. M., & Nichols, K. (2015). How to support primary teachers' implementation of inquiry: Teachers' reflections on teaching cooperative inquiry-based science. *Research in Science Education, 45*(2), 171–191. https://doi.org/10.1007/s11165-014-9418-x

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.

Goldin-Meadow, S., & Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annual Review of Psychology, 64*, 257. https://doi.org/10.1146/annurev-psych-113011-143802

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: *Imaginative Science Education* (pp. 185–215). Springer, Cham. https://doi.org/10.1007/978-3-319-29526-8

Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. *Creative Education, 3*(05), 603. https://doi.org/10.4236/ce.2012.35089

Hardiman, M., Rinne, L., & Yarmolinskaya, J. (2014). The Effects of Arts Integration on Long-Term Retention of Academic Content. *Mind, Brain, and Education, 8*(3), 144–148. https://doi.org/10.1111/mbe.12053

Hardison, D. M., & Sonchaeng, C. (2005). Theatre voice training and technology in teaching oral skills: Integrating the components of a speech event. *System, 33*(4), 593–608. https://doi.org/10.1016/j.system.2005.02.001

Hardoin, M. M., Wise, L., Thacker, A., Dickinson, E., Wiley, C., Becker, D. S., & Schultz, S. (2016). *California Assessment of Student Performance and Progress (CAASPP)*. Independent Evaluation Report Volume.

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K. W. (2016). *Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications*. SRI International.

Haynie, W. J., III. (1997). Effects of anticipation of tests on delayed retention learning. *Journal of Technology Education, 9*(1), 20–30. https://doi.org/10.21061/jte.v9i1.a.2

Holzman, L. (2016). *Vygotsky at work and play* (2nd ed.). UK: Routledge. 10.4324/9781315676081.

Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K., & Klieme, E. (2009). Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction, 19*(1), 66–78. https://doi.org/10.1016/j.learninstruc.2008.02.001

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures, vol 22. *Routledge & Kegan Paul*. https://doi.org/10.1037/10034-000

Jones, A. (2015). Bilingual Education as a Professional Responsibility for Public Schools and Universities. In D. Mitchell & R. Ream (Eds.), *Professional Responsibility Advances in Medical Education.* (Vol. 4). Cham: Springer. 10.1007/978-3-319-02603-9_15.

Jones, B., & Kenward, M. G. (1989). *Design and analysis of cross-over trials*. Chapman and Hall/CRC. 10.4324/9780203009277.

Kena, G., Aud, S., Johnson, F., Wang, X., Zhang, J., Rathbun, A., ... & Kristapovich, P. (2014). The Condition of Education 2014. NCES 2014–083. *National Center for Education Statistics.*

Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. *Studies in Science Education, 43*(1), 1–37. https://doi.org/10.1080/03057260708560225

Krashen, S. D. (1983). The din in the head, input, and the language acquisition device. *Foreign Language Annals, 16*(1), 41–44. https://doi.org/10.1111/j.1944-9720.1983.tb01422.x

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. UK: Addison-Wesley Longman Limited.

Kress, G. (2009). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kusters, A., Spotti, M., Swanwick, R., & Tapio, E. (2017). Beyond languages, beyond modalities: Transforming the study of semiotic repertoires. *International Journal of Multilingualism, 14*(3), 219–232. https://doi.org/10.1080/14790718.2017.1321651

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought* (Vol. 640). Basic books.

Larsen-Freeman, D. (2007). Reflecting on the cognitive–social debate in second language acquisition. *The Modern Language Journal, 91*(s1), 773–787. https://doi.org/10.1111/j.1540-4781.2007.00668.x

Lee, O. (2018). English Language Proficiency Standards Aligned With Content Standards. *Educational Researcher, 47*(5), 317–327. https://doi.org/10.3102/0013189X18763775

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

Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching, 46*(7), 836–859. https://doi.org/10.1002/tea.20335

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for common core state standards for English Language Arts and Mathematics. *Educational Researcher, 42*(4), 223–233. https://doi.org/10.3102/0013189X13480524

Martinez, J. E. (2017). *The search for method in STEAM education* (pp. 111–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-55822-6

Morgan, C., & Stengel-Mohr, J. (2014). Language at the core: Developing meaningful connections between the arts and classroom instruction. *NYS TESOL Journal, 1*(1), 90–99.

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*(1), 18–35. https://doi.org/10.3102/0013189X16633182

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *The American Journal of Evaluation, 24*(3), 315–340.

National Commission on Mathematics, Science Teaching for the 21st Century (US), & United States. Dept. of Education. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. Diane Publishing Company.

National Center for Education Statistics. (2019). *College navigator*. U.S. Dept. of Education, Institute of Education Sciences, *National Center for Education Statistics*.Institute of Education Sciences, NationalAssessment of Educational Progress (NAEP), various years, 2009–2019 Science Assessments.

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*. The National Academies Press.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

OECD. (2019). PISA 2018 results (Volume I): What students know and can do. *OECD*. https://doi.org/10.1787/5f07c754-en

Oliver, K. M. (2000). Methods for Developing Constructivist Learning on the Web. *Educational Technology*, 40(6), 5–18. http://www.jstor.org/stable/44428633

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. In: *Executive Office of the President*. Executive Office of the President. Retrieved from https://eric.ed.gov/?id=ED541511

Otheguy, R., García, O., & Reid, W. (2015). Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review, 6*(3), 281–307. https://doi.org/10.1515/applirev-2015-0014

Paas, & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2

Pappamihiel, N. E. (2002). English as a second language students and English language anxiety: Issues in the mainstream classroom. *Research in the Teaching of English*, 327–355. Retrieved from https://www.jstor.org/stable/40171530

Peppler, K. A., Powell, C. W., Thompson, N., & Catterall, J. (2014). Positive impact of arts integration on student academic achievement in english language arts. *The Educational Forum, 78*(4), 364–377. https://doi.org/10.1080/00131725.2014.941124

Piaget, J. (1952) When thinking begins. In: *The Origins of Intelligence in Children*, International Universities Press, 25–36. https://doi.org/10.1037/11494-000

Piaget, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes,* (pp. xii, 384). U. Chicago Press

Piaget, J. (1936). *Origins of intelligence in the child*. Routledge & Kegan Paul.

Piaget, J. (1963). The attainment of invariants and reversible operations in the development of thinking. *Social Research, 30*(3), 283–299.

Posner, S. K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*(2), 211–227. https://doi.org/10.1002/sce.3730660207

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*.

Ramirez, H., & Jones, D. (2016). Effects of reading strategies and the writing process with written recasts on second language achievement. *Administrative Issues Journal: Connecting Education, Practice, and Research, 3*(1), 516. https://doi.org/10.5929/2013.3.1.7

Raudenbush, S. W. (2007). Designing field trials of educational innovations. *Scale up in Education: Issues in Practice, 2*, 1–15.

Richland, L. E., & Hansen, J. (2013). Reducing cognitive load in learning by analogy. *International Journal of Psychological Studies, 5*(4), 69. https://doi.org/10.5539/ijps.v5n4p69

Rieg, S. A., & Paquette, K. R. (2009). Using drama and movement to enhance English language learners' literacy development. *Journal of Instructional Psychology, 36*(2), 148–155.

Ritz, J. M., & Fan, S.-C. (2015). STEM and technology education: International state-of-the-art. *International Journal of Technology and Design Education, 25*(4), 429–451. https://doi.org/10.1007/s10798-014-9290-z

Schinke-Llano, L., & Vicars, R. (1993). The affective filter and negotiated interaction: Do our language activities provide for both? *The Modern Language Journal, 77*(3), 325–329. https://doi.org/10.2307/329101

Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). Moving beyond "knowing about" science to making sense of the world. In: *Helping students make sense of the world using next generation science and engineering practices*, 3–21. NSTA Press.

Shapiro, L. (2014). *The Routledge handbook of embodied cognition*. Routledge.

Skulmowski, & Rey, G. D. (2017). Measuring cognitive load in embodied learning settings. *Frontiers in Psychology, 8*, 1191–1191. https://doi.org/10.3389/fpsyg.2017.01191

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

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.

Wadsworth, B. J. (1996). *Piaget's theory of cognitive and affective development: Foundations of constructivism*. Longman Publishing.

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.

Wang, L. (2020). Application of affective filter hypothesis in junior English vocabulary teaching. *Journal of Language Teaching and Research*. https://doi.org/10.17507/jltr.1106.16

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A. S., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching, 38*(5), 529–552. https://doi.org/10.1002/tea.1017

Wei, L. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amx039

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in psychology, 5*, 506. https://doi.org/10.3389/fpsyg.2014.00506.

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

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625–636. https://doi.org/10.3758/BF03196322

Wong, J., Bui, E., Fields, D., & Hughes, B. (2022). A learning experience design approach to online professional development for teaching science through the arts: Evaluation of teacher content knowledge, self-efficacy and STEAM perceptions. *Journal of Science Teacher Education*. https://doi.org/10.1080/1046560X.2022.2112552

Yoon, H. G., Joung, Y. J., & Kim, M. (2012). The challenges of science inquiry teaching for pre-service teachers in elementary classrooms: Difficulties on and under the scene. *Research in Science Education, 42*(3), 589–608. https://doi.org/10.3390/systems4040038

Zuengler, J., & Miller, E. R. (2006). Cognitive and sociocultural perspectives: Two parallel SLA worlds? *TESOL Quarterly, 40*(1), 35–58. https://doi.org/10.2307/40264510

Publisher's Note