# Reviewing assessment of student learning in interdisciplinary STEM education

Xiaoyi Gao[1], Peishan Li[1], Ji Shen[2*] and Huifang Sun[1]

## Abstract

STEM education has received increasing attention in recent years. However, developing valid and reliable assessment of interdisciplinary learning in STEM has been a challenge. This study is a comprehensive review of assessment of interdisciplinary STEM education during the last two decades. A total of 49 empirical research articles were collected from an initial library of 635 articles focusing on interdisciplinary STEM education. A two-dimensional framework was developed to examine the literature. The first dimension concerns the nature of disciplines being assessed and includes three categories: monodisciplinary, interdiscipline, and transdiscipline. The second dimension concerns the learning objectives and includes four aspects: knowledge, skill, practice, and affective domain. The results show that most assessments focused on assessments of monodisciplinary knowledge, monodisciplinary affective domains, and transdisciplinary affective domains. Although many programs aimed to improve students' interdisciplinary understanding or skills, their assessments did not align with their aims. Based on the review, future directions are proposed for developing assessments for interdisciplinary STEM educational programs.

**Keywords:** Interdisciplinary STEM education, Assessment

## Introduction

STEM (Science, Technology, Engineering, and Mathematics) has become a popular term in education worldwide. It is acknowledged that there is a need to have a workforce with adequate STEM knowledge and skills to meet the challenges in the future (e.g., Caprile et al., 2015; English, 2016; Marginson et al., 2013; National Science and Technology Council, 2013; The Royal Society Science Policy Centre, 2014).

While STEM education has drawn increased attention and research, it is also filled with debates and dilemmas (STEM Task Force Report, 2014). There is little consensus on what STEM education means and how it should be realized in practice (NAE & NRC, 2014). In a broad sense, it can refer to either the sum of the individual disciplines involved in STEM or an interdisciplinary approach to STEM education that emphasizes the connections across disciplines. The latter is what this paper focuses on. That is, in this paper, STEM education stands for *interdisciplinary STEM education*, which will be further elaborated in the "Theoretical framework" section.

STEM education is driven by today's complex policy and economic, social, and environmental problems that require solutions integrated and interdisciplinary in nature (e.g., Bryan, Moore, Johnson, & Roehrig, 2015; English, 2017; Sanders, 2009). Simply put, it is a means for linking students' learning across the STEM disciplines (NRC, 2009; STEM Task Force Report, 2014). However, given that the traditional discipline-based approach is still dominant in the educational system, how interdisciplinary STEM education should be assessed and evaluated has raised many concerns, such as inadequate research in evaluating the efficacy of integrated instruction and challenges in assessing students' interdisciplinary understanding (Griese et al., 2015; Herro et al., 2017; Shen, Liu & Sung, 2014; You et al., 2018).

This review study aims to address the timely issue on assessing student learning in STEM education to move the field forward. Through reviewing relevant literature

* Correspondence: ji.shen1221@gmail.com
[2]Department of Teaching and Learning, School of Education and Human Development, University of Miami, Coral Gables, FL, USA
Full list of author information is available at the end of the article

and using a two-dimensional framework, the study summarizes and examines approaches in assessing student learning in STEM education beyond the elementary or primary level. While acknowledging there is a body of literature focusing on interdisciplinary STEM education at the elementary level (e.g., English & King, 2015; Sinatra et al., 2017), we did not include this level in this study for two reasons: (1) theoretically, students at this level tend to have a very naïve concept about the (nature of) disciplines involved in STEM; (2) pragmatically, we wanted to narrow down the scope of this study. The following question guided this review study: What are typically included in assessments of student learning in STEM education at the secondary and the tertiary level?

## Theoretical framework

Assessment in this paper refers to deliberate effort to observe student learning through different means to evaluate where students are with respect to one or more specific learning objectives. We follow the reasoning elaborated in the "assessment triangle" (NRC, 2001; Pellegrino, 2014) that includes three important ends: cognition, observation, and interpretation. Specifically, we submit that the design of any assessment should begin with the cognition end—the target knowledge, skills, or any aspects to be assessed in a specific subject domain and how such knowledge and skills can be represented. Since we focus on interdisciplinary STEM education that involves multiple disciplines, the cognition end involves two interrelated aspects: (a) the nature of the multiple disciplines to be assessed and (b) the learning objectives in relationship with the multiple disciplines.

### Nature of disciplines

As STEM education involves multiple STEM disciplines, it is necessary to clarify the nature of disciplines in our study. In general, the term discipline refers to a branch of human knowledge (Choi & Pak, 2006; Klein, 1990) and associated tradition, culture, and community related to that discipline. More specifically, in this study, disciplines include the main fields considered in STEM education such as science, technology, engineering, and mathematics, as well as their sub-fields such as physics in science and geometry in mathematics.

People use different terms (e.g., integrated, interdisciplinary, transdisciplinary) to describe the connections and integration among multiple disciplines (NAE & NRC, 2014; Shen, Sung, & Zhang, 2015; STEM Task Force Report, 2014). An interdisciplinary approach juxtaposes and integrates (parts of) two or more disciplines, focusing on establishing explicit connections between relevant disciplines (Klein, 2004; Miller, 1981). In education, we view interdisciplinarity as learners building connections between different disciplines, such as integrating

knowledge and skills from two or more disciplines in order to solve complex problems or explain sophisticated phenomena. We also submit that an interdisciplinary approach is built upon its disciplinary roots (Zhang & Shen, 2015). Therefore, in an interdisciplinary approach, learning in each discipline is clearly discernable and needs to be supported (NAE & NRC, 2014).

A transdisciplinary approach refers to the unity of knowledge and skills beyond disciplinary framing (Nicolescu, 2002). As Miller (1981) put it elegantly, "transdisciplinary approaches are articulated conceptual frameworks which claim to transcend the narrow scope of disciplinary world views and metaphorically encompass the several parts of the material field which are handled separately by the individual specialized disciplines" (p.21). Compared with an interdisciplinary approach, a transdisciplinary approach focuses on the problem space or issue at hand without paying much attention to traces of the individual disciplines (Klein, 2008). Similar to a transdisciplinary perspective, an integrated approach in educational settings often refers to combining ideas and subject matters of different disciplines into a seamless whole (Lederman & Niess, 1997).

### Learning objectives

A variety of approaches to STEM integration have been proposed (e.g., Guzey et al., 2014; Moore et al., 2014; NAE & NRC, 2014; Sanders, 2012). Oftentimes, the primary reason why these approaches differ is that each program has its unique learning objective(s). According to Bloom's taxonomy (Anderson & Krathwohl, 2001; Bloom, 1956), learning objectives may be categorized into three types: the cognitive, the affective, and the psychomotor domains. The cognitive domain includes knowledge and processes and skills that are mentally performed (e.g., understanding factual knowledge and conceptual knowledge, practicing procedural knowledge and metacognitive knowledge): The affective domain concerns constructs related to feelings and emotions (e.g., interest, attitudes, motivation, and values): The psychomotor domain deals with processes and skills that are physically performed (e.g., body movements, physical abilities). As there is no STEM educational program focusing on the psychomotor domain from our pool of literature, we chose to focus on the cognitive and affective domains in this study.

Considering the cognitive domain, a basic learning objective of STEM education is to help students develop content knowledge for one or more specific disciplines within STEM. For instance, a common approach in STEM education is to use engineering design to help students develop math and/or science knowledge (Becker & Park, 2011; Guzey et al., 2014; Sanders, 2009, 2012). In this approach, one (or more) specific

discipline(s) within STEM (e.g., math and science) is prioritized and the other discipline(s) (e.g., engineering) serves as a vehicle to deliver that discipline.

Another common learning objective of STEM education within the cognitive domain is to help students develop skills that go beyond a single discipline. Consistent with this perspective, a range of programs focusing on STEM integration have used learning tasks that are situated in the context of a complicated situation and require students to apply knowledge from multiple disciplines (e.g., Hurley, 2001; Moore et al., 2014). In this way, the importance for each discipline is treated at the same level (helping understand the situation), and the amount of knowledge for each discipline depends on the nature of the learning problem or situation.

While educators have long pointed out the problem of separating knowledge and skills in teaching and learning, integrating knowledge and skills as a measurable outcome poses significant challenges. For instance, the US Next Generation Science Standards (NGSS; NRC, 2012) adopted the term "practices" to "emphasize that engaging in scientific investigation requires not only skill but also knowledge that is specific to each practice" (p.45). New approaches to assess NGSS in a more integrated manner have shown promises (NRC, 2014).

Many have argued that a significant outcome of an interdisciplinary STEM education approach is students' positive enhancement in the affective domain (e.g., Apedoe et al., 2008; Guzey et al., 2016). The affective domain in our study includes measures such as students' interest, engagement, attitude, and motivation for STEM contents and practices and career aspiration for STEM professions.

## Methods

In order to conduct a systematic review, we followed established procedure and guideline (e.g., Liberati et al., 2009; Namdar & Shen, 2015). We conducted a careful literature search and screening and then extracted key information from the studies included in our pool (see below). We also developed a coding framework according to our theoretical framework to analyze the studies. We reported both descriptive statistics and key information, as well as important insights, supplemented with examples for the reader to better understand these results.

### Literature collection

Since we targeted STEM education, "STEM" and "science, mathematics, technology, and engineering" were the initial terms to start with. Due to the different interpretations of STEM education, we included "integrated", "interdisciplinary", and "transdisciplinary" in the searched terms. In order to expand the initial pool, we used different combinations of these terms in the title, abstract, and keywords to search for literatures in ERIC, ProQuest, as well as major STEM educational journals such as International Journal of Science Education, Journal of Research in Science Teaching, Science Education, International Journal of Science and Mathematics Education, and International Journal of STEM Education. The search generated a number of 635 initial references. Fig. 1 shows the search result in each step.

The abstracts of these articles were reviewed to conduct a preliminary screening. The following criteria were then used to select relevant articles for our study.

- Only research articles that report interdisciplinary STEM educational programs for G6-college students empirically and explicitly were included. Again, as elementary or primary school students experienced much less time in solid disciplinary learning in STEM, studies targeting the elementary level were excluded in this review.
- The articles need to report quantitative or qualitative methods to assess student learning, including knowledge, skills, practice, or affective domains, as well as descriptions of curriculum and learning activities that allow us to better understand their assessments. We excluded literature focusing solely on developing specific program evaluation/assessment tool in STEM (e.g., Hernandez, et al., 2014; Herro, et al. 2017). Studies assessing teachers' development or pedagogical skills were also excluded.
- We only included articles published during the 2000–2019 period. Prior to 2000, STEM education paid more attention to improve mathematics and science outcomes separately and involved little integration of engineering and technology (Breiner et al., 2012; Bybee, 2010; Hoachlander & Yanofsky, 2011; Sanders, 2009).
- To narrow down our scope, we only selected articles published in peer-reviewed journal articles in English.

### Analysis

After selecting relevant articles, we read all the articles carefully and recorded all of the basic information (authors, journal name, year published, grade level, disciplines involved, types of assessment, main results, etc.) in a spreadsheet. In accordance with our theoretical framework, we then examined the assessment aspect of each article by establishing a coding framework that has two dimensions: (1) the nature of disciplines and (2) the learning objectives (Fig. 2). The first dimension includes three categories: monodisciplinary (i.e., the assessment targets individual disciplines), interdisciplinary (i.e., the

**Fig. 1** Literature screening flowchart
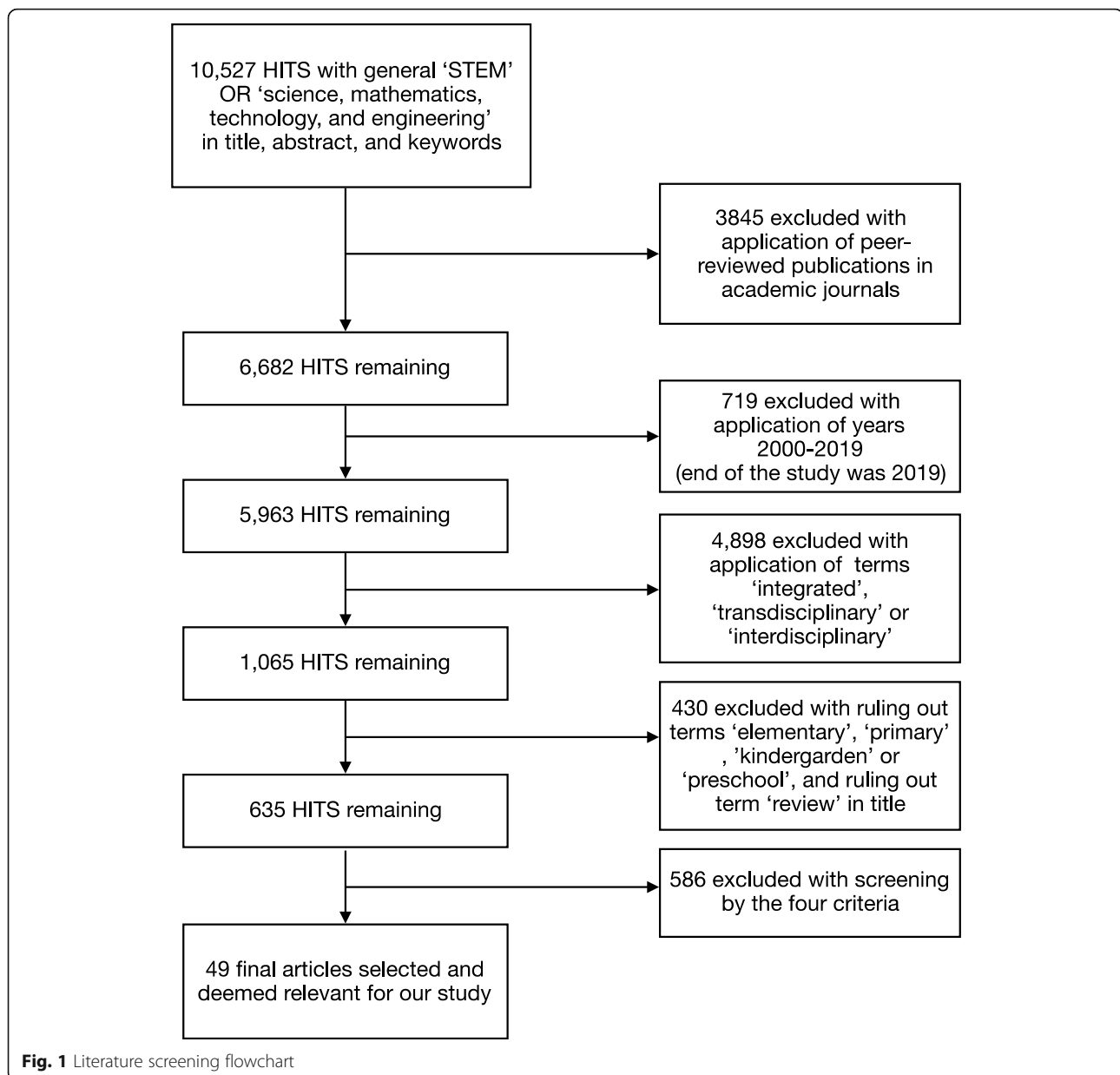
assessment targets connections between disciplines), and transdisciplinary (i.e., the assessment targets beyond disciplinary constraints).

The second dimension includes four aspects related to learning objectives: knowledge, skill, practice, and affective domain. Knowledge in our study refers to structured and systematized concepts organized in a single discipline (e.g., force in physics), and crosscutting concepts or general principles connecting multiple disciplines (e.g., concept of size and scale) (McCormick, 2004). Skills, on the other hand, focus on learner' ability to do something. When put in the context of disciplines, skills may include monodisciplinary skills (e.g., experimental skills in science), interdisciplinary skills (e.g.,

integrating knowledge from two or more disciplines), and transdisciplinary skills (e.g., creativity, critical thinking).

Note that since an article may report multiple types of assessment, each article can be coded into more than one category. Two graduate students were first trained to be familiar with the coding framework. Then, each person coded a set of the articles separately and checked the other person's coding. Issues emerging from the coding process were brought up for discussion and were all resolved with the research team. At the end, the two coders double checked all the coding together. Cohen's Kappa indicates that the data obtained from two raters had a high inter-rater

**Fig. 2** Coding framework

reliability ($F$ = 4.0, Sig = .000), which suggests that the raters coded the articles with a high level of consistency.

Before we report the results, the reader should keep in mind that the review has several limitations. As this review focused on student learning in STEM educational programs, the search was restricted to literature with an explicit focus on interdisciplinary STEM, not studies on the STEM programs consisting of individual disciplines of science, technology, engineering, or mathematics. Besides, the STEM education movement is still relatively young, only becoming prominent in international discourse in the last two decades (Blackley & Howell, 2015). Moreover, as we used G6-college as a criterion for the screening process and excluded the studies solely designed to develop assessments for STEM program

without the description of the program itself, we underestimated the work done in the field.

## Results

After a careful screening, a total of 49 articles were selected into our library from an initial pool of 635 references (see Supplementary Materials for the list of reviewed articles). In the following, we first describe general information extracted from the reviewed studies; we then report results related to our research question.

### General information about the studies

Figure 3 shows the distribution of the articles over the years. Not surprisingly, there is a trend of steep increase over the years, which indicates that research in STEM education is still on the rise. Figure 4 shows the
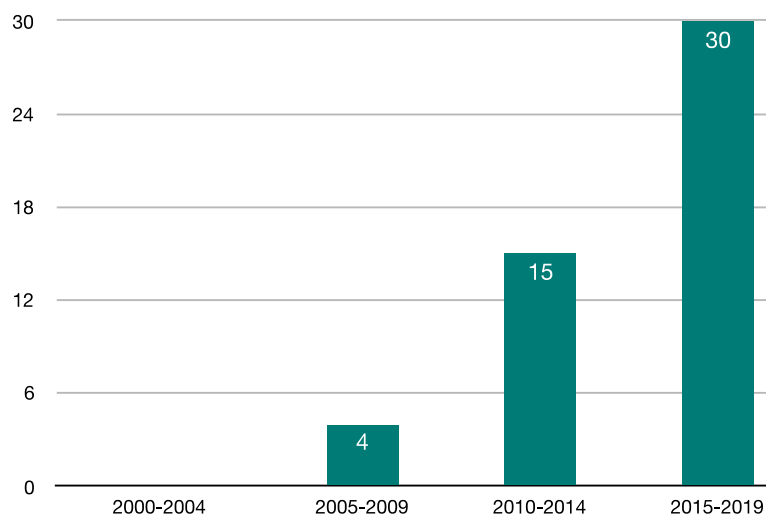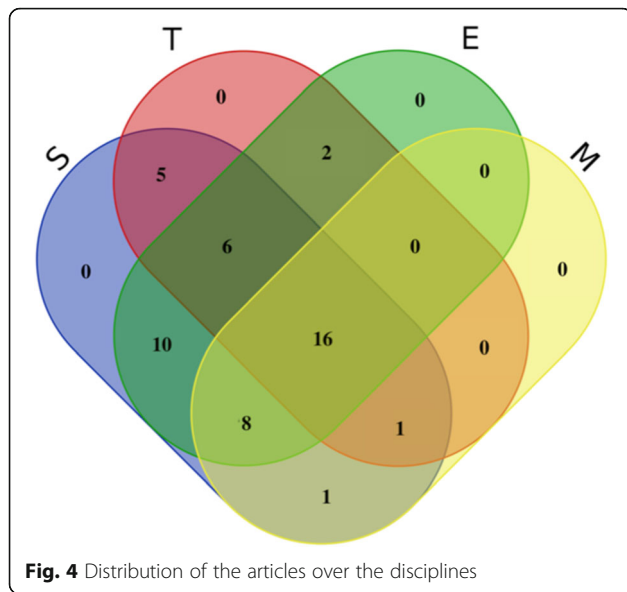


**Fig. 3** Distribution of the articles over the years

**Fig. 4** Distribution of the articles over the disciplines

distribution of the articles over the disciplines. Science was the dominant discipline ($n = 47$) and engineering was the second ($n = 42$). Most articles ($n = 40$) include both science and engineering in their studies. In terms of implementation context, there were more studies conducted in informal settings ($n = 28$) than in formal settings ($n = 21$). In terms of grades, there were 22 in middle grades (G6-8), 20 in high school level (G9-12), and 12 in tertiary level.

## Specific findings

We sought to answer the following research question: What are typically included in assessments of student learning in STEM education? In this section, we first summarize the assessment formats used in the studies and then report the results according to our coding framework (Fig. 2).

The assessment formats vary a great deal, ranging from traditional formats such as standardized tests to research-oriented observational methods (e.g., interviews, video analysis; see Table 1). The formats for monodisciplinary knowledge (MDK) and monodisciplinary affective (MDA) assessments are relatively universal across different studies whereas and those for interdisciplinary knowledge and practices (IDK, IDP) are more diverse. This is intuitive as assessing content knowledge in individual disciplines has a long tradition and is relatively easier. In contrast, assessing interdisciplinary knowledge, practices, and affective domains is much more challenging and calls for innovations.

In terms of analysis or interpretation of assessment results, there are more quantitative studies ($n = 38$) than qualitative studies ($n = 26$). For quantitative studies, a variety of inferential statistics were used (e.g., Students' $t$ test, regression study, ANOVA) besides descriptive and correlational analysis. Psychometric analyses were also performed in many studies to examine the validity and reliability of the instrument (e.g., applying Rasch modeling to examine the construct validity), but 10 studies did

**Table 1** Formats of assessments in STEM educational programs. (The numbers represent each of the 49 articles included in the study)

| Main formats | Specific formats | Articles |
|---|---|---|
| Questionnaire/test | Likert scale | 1, 3, 4, 6, 10, 11, 17, 18, 19, 23, 25, 27, 28, 30, 31, 33, 35, 36, 37, 38, 41, 46, 47, 48, 49 |
| | Multiple-choice | 1, 2, 3, 5, 6, 9, 12, 16, 32, 38, 39, 46 |
| | Open-ended items/design problems | 5, 8, 10, 12, 13, 16, 20, 23, 24, 32, 44, 45, 47 |
| | True/false | 8 |
| | Semi-differential scales | 4 |
| Interviews | Individual (structured) | 4, 18, 35 |
| | Individual (semi-structured) | 16, 22, 23, 30, 31, 33, 36, 44 |
| | Individual (structure unknown) | 7, 13, 29 |
| | Focus group | 22, 26, 27, 43 |
| Observation/field notes | | 1, 16, 29, 31, 33, 35, 39, 43, 44 |
| Learning/working sheet/activity booklets | | 15, 17, 27, 31, 33, 34, 35, 40 |
| Design product analyzing | | 15, 16, 34, 39, 40, 42, 45, 47 |
| Video analyzing | | 6, 8, 34 |
| Students' written work analyzing | | 10, 14, 19, 39, 40 |
| Application/graduate rates | | 12, 20 |
| Reflexive practice | | 26, 43 |
| Digital pile sort | | 32 |

not include any discussion on validity or reliability of their assessment methods.

Table 2 shows the distribution of the articles on the two dimensions: the nature of the disciplines and the learning objectives. In the following, we report results of selected cells in Table 2 with illuminating examples.

### Monodisciplinary knowledge

A total of 19 articles (38.8%) included assessments that focused on students' knowledge in individual disciplines or their sub-disciplines within STEM. All of them used pre-post design to examine students' content learning gains. For example, Apedoe et al. (2008) described an 8-week, design-based unit that integrated science and engineering for high school students to improve their understanding of chemistry concepts (e.g., energy changes in reactions). Students completed a three-phase design of a heating/cooling system prototype: conceiving the prototype, developing the subsystems, and presenting the design. The assessment of the unit (i.e., the pre/post-test) focused on chemistry knowledge and consisted of 24 items selected from the American Chemical Society's and the Chemical Concept Inventory test item pool. The results indicated that students' understanding of the key chemistry concepts (e.g., reactions, atomic interactions, energy) improved significantly.

Note that 13 out of the 19 studies actually included more than one discipline in their assessment. Guzey et al. (2017) presented an engineering design-based STEM program to improve elementary and middle school (G4-8) students' learning of science, engineering, and mathematics knowledge. The pre/post-test, developed by a team of classroom teachers, school curriculum specialists, and researchers, focused on assessing students' engineering, science, and mathematics knowledge. The test for middle grades had 45 multiple-choice items (15 for each discipline). The mathematics- and science-related items were chosen from publicly available large-scale assessments such as Trends in International Mathematics and Science Study (TIMSS) and National Assessment of Educational Progress (NAEP); the engineering items were developed by the authors. They found that student performances on the different science assessments were mixed when comparing

between the treatment students (whose teachers attended a corresponding professional development program) and the controlled students (whose teachers did not). Furthermore, no significant treatment effect was found for engineering and mathematics assessments. One interesting finding was that quality of curriculum units and type of engineering integration were associated with student learning gains.

Despite that all of the programs described in these 19 articles emphasized interdisciplinary connection in instruction, assessments developed or adopted by these researchers focused on measuring students' content gains within individual disciplines. Even when multiple disciplines were assessed simultaneously, they were done in a separate manner. Furthermore, the assessment items were typically adopted or adapted from existing standardized tests. Most of the instruments combined multiple-choice and open-ended items. Notably, the majority of these articles (9 out of 13) reported mixed results with respect to student learning gains in different disciplines.

### Interdisciplinary knowledge

In contrast, among the 49 articles selected, only 7 (14.3%) included assessments of students' IDK. The methods of measuring IDK are more diverse (e.g., presentation, written work, and tests that included multiple-choice and open-ended items). For instance, Riskowski, Todd, Wee, Dark, and Harbor (2009) developed a test assessing students' IDK in an interdisciplinary engineering module for eighth grade science class. The module aimed at promoting students' knowledge of an interdisciplinary topic, drinking water safety, by engaging them in designing, building, and testing a water purification device. To evaluate student learning in the module, the researchers developed a pre/post-test that consisted of five true/false questions (focusing on "factual knowledge of water quality" and taken directly from the textbook), five open-ended questions (focusing on "human impact of the availability of safe drinking water"), and one design question (asking students to use drawings, words, or phrases to "describe and explain what was needed to ensure safe drinking water and how their purification design addressed the water quality issues they

**Table 2** Content involved in assessments of STEM educational programs. (The numbers represent each of the 49 articles included in the study)

|  | Knowledge (K) | Skill (S) | Practice (P) | Affective (A) |
|---|---|---|---|---|
| Monodisciplinary (MD) | 1, 2, 3, 4, 5, 6, 7, 9, 12, 14, 16, 23, 27, 32, 37, 38, 39, 45, 46 | 16, 23, 43, 44 | 35, 43 | 1, 3, 4, 6, 7, 13, 18, 24, 29, 33, 35, 36, 38, 39, 44, 46 |
| Interdisciplinary (ID) | 8, 10,13, 27, 42, 43, 49 | 12, 48 | 15, 16, 17, 27, 34, 40, 42 | – |
| Transdisciplinary(TD) | – | 11,13, 14, 16, 22, 30, 33, 38, 39, 44, 47 | – | 12, 14, 19, 20, 21, 23, 24, 25, 26, 27, 28, 30, 31, 37, 38, 40, 41, 45 |

identified"). For the design question, the highest level of response needed to identify three different types of contaminants in the water and use evidence to justify the design. The instrument as a whole did target students' interdisciplinary understanding of drinking water safety and related issues. In analysis, students' total score was also analyzed separately as the science component score and the engineering component score. Unfortunately, it was not entirely clear how the researchers assigned the two component scores. Also, the authors described that they developed a rubric to evaluate different aspects of students' final presentations, but the rubric was not provided and this evaluation was not factored in the research analysis.

As a different example, Dierdorp et al. (2014) assessed students' written works and analyzed their reflections in an interdisciplinary unit. The unit focused on building meaningful connections among natural science, mathematics, and statistics through professional practices. Six design cycles were adopted for students' learning, and three types of professional practice were used as contexts: sports physiologist, dike height detection, and calibration of measuring instruments. Focusing on IDK development, the researchers analyzed their written works that were completed during the process of designing, as well as student questionnaires and interviews. Five possible connections (i.e., math-science, math-professional practices, statistics-science, professional practices-science, and professional practices-statistics) were explicitly assessed. Specifically, the written works were first coded in each discipline (e.g., "formulas using" represented math). The answers of questionnaires were coded into positive, negative, or inconclusive. The results indicated that integrating professional practices with science, mathematics, and statistics could help students build more connections among these disciplines.

Among the seven studies in the IDK cell, researchers all emphasized assessing students' interdisciplinary learning by identifying and evaluating the connections between different disciplines students built in the learning process. Compared with those studies from other cells, while more diverse formats of assessments were used, the analysis and/or coding process were much more cumbersome.

### Transdisciplinary skill
A total of 11 articles (22.4%) reported assessments of students' TDS. Ability tests, interviews, and informal observation were used in these studies. Five of the nine studies used pre/post-tests while the rest used interview or process assessments.

Chen and Lin (2019) reported a 2-year interdisciplinary STEM course, which consisted of maker-centered, project-based learning (M-STEM-PjBL). They aimed at enhancing rural middle school students' creativity, problem solving, science learning, and their attitudes toward science in Taiwan. The M-STEM-PjBL course had five phases: preparation, implementation, presentation, evaluation, and correction. The content of the first year consisted of designing GigoToys and writing S4A programs, and the second-year curriculum focused on creating green energy. The researchers assessed students' creativity and collected students' notes, written records, teachers' notes, and project products as empirical data. Their analysis found that the M-STEM-PjBL curriculum increased students' practical skills (such as doing hands-on activity) and creativity. However, although multiple types of qualitative data were enlisted, they did not provide sufficient information about the rubrics they used to evaluate these data.

Lou, Chou, Shih, and Chung (2017) presented a study in which they integrated STEM education with project-based learning (PBL) and investigated the effectiveness of this curriculum. The curriculum used $CaC_2$ steamship as its theme of design and had five stages: preparation, implementation, presentation, evaluation, and correction. Students' learning processes were evaluated by interviews, and their learning gain was examined by the "Creativity Tendency Scale", adapted from the Williams Creativity Assessment Packet revised by Lin and Wang (1994). Analysis of the Creativity Tendency Scale responses showed that students improved in all four aspects: adventurousness, imagination, curiosity, and challenge. However, although using a TDS test may be appropriate, the study did not provide much information regarding the test itself. Also, the study did not state how the different aspects of creativity were integrated into their curriculum activities.

The 11 articles in the TDS cell described assessments that focused on students' skills beyond specific disciplines. Ability tests (e.g., critical thinking) were used in four of these studies. Interviews were also used in five studies to capture students' perceptions of their own skill status and gains.

### Interdisciplinary practices
A total of 7 articles (14.3%) included some forms of assessment related to engineering design, such as designing a manufacture of a solar automatic trolley (Lou, Shih, Diez, & Tseng, 2011). Although design is a trademark of engineering practices, we categorized these studies in IDP because all of them focused on assessing students' design knowledge and skills simultaneously and, at the same time, attending to how students connect knowledge and skills from multiple disciplines.

English, King, and Smeed (2017) presented a study in which Australian sixth grade students engaged in designing an earthquake-proof structure using simple

materials. The students were first introduced to some basic science information (e.g., tectonics) and relevant engineering terms (cross bracing, tapered geometry, and base isolation). The students were then presented with the design challenge for which they worked in small teams to design and construct a two-story structure resistant to simulated "earthquakes." While completing the challenge, students needed to respond to an activity booklet that included several guiding questions related to their design (e.g., one such question was "What did you learn about your new building from the test, including any mathematics and science that you used?"). Using multiple sources of data (e.g., student design sketches and annotations, responses to questions in booklets, group work video transcripts), the researchers analyzed through open-coding how students engaged in the design process, as well as how they applied their STEM disciplinary knowledge in the process. The results shed light on how students coming from different regions approached the design task differently and how their STEM disciplinary knowledge developed in the design process. The research delineated the complex design processes through multiple angles and perspectives. However, the assessment method (open coding in data analysis) did not seem to be readily applicable in a practical classroom setting.

As one can see in English et al. (2017), it is very time consuming to assess practices as students engage in the processes. Other studies focused on assessing design products instead. For instance, Fan and Yu (2017) developed a STEM engineering module and compared it with a traditional technology module for high school students in Taiwan. The STEM engineering module was designed to intentionally integrate science (e.g., mechanics) and mathematics knowledge (e.g., geometry) and processes with that of technology and engineering. In both the STEM engineering module and the technology module, students spent the last 6 weeks of the 10-week curriculum to design a mechanical toy. Besides other types of assessment, students' final design products were evaluated to tap into students' design practices. More specifically, the products were scored on three aspects: mechanical design, working function, and materials used (each on a five-point scale). The product analysis showed that the students using the STEM engineering module outperformed those using the traditional technology unit in all three aspects. It was interesting to point out that in this study the quality of student products was not correlated with either students' conceptual knowledge or their higher order thinking skills.

Compared with IDK, assessments in IDP focused more on processes and presented more diverse formats. Out of the seven studies, five assessed both the design processes and the products, one assessed products only, and one assessed processes only. One observation is that, while having an iterative cycle is a defining feature of engineering design, all of the product-based assessments only focused on the final design after the iteration process.

## Monodisciplinary and transdisciplinary affective perspectives

The affective domain is the most popular assessment target in our pool of articles. A total of 16 articles (32.7%) included assessments of students' MDA; a total of 18 articles (36.7%) targeted TDA domains.

The themes of MDA assessments include awareness, attitudes, beliefs, motivation, and interest toward specific disciplines in STEM ($n = 11$) and perceptions of STEM-related careers ($n = 7$). For instance, Apedoe et al. (2008) showed that their design unit not only increased students MDK in chemistry (see the "Monodisciplinary knowledge" section for their program description) but also increased their interest in engineering careers. Students completed a five-point Likert scale survey to self-report to what extent they showed awareness and interest in engineering careers. The scale was divided into four dimensions: desire to attend more engineering classes, awareness of engineering as a career, desire to be an engineer, and desire to attend engineering informal activities. Results indicated that students' interest and awareness in engineering careers significantly increased after the unit. One intriguing point was that the researchers assessed students' content knowledge in science (chemistry in particular) while assessing their interest in engineering careers.

The themes of TDA assessment involve interests toward STEM, self-efficacy and willingness to major in STEM-related disciplines in college, or attitude towards STEM career without spelling out specific disciplines. For example, Musavi et al. (2018) reported the Stormwater Management and Research Team (SMART) program, which included a weeklong summer institute and an academic-year student research program for high school students. The program activities were designed to help students to increase interest and confidence, improve STEM knowledge and skills, and apply them to research and solve water quality problems in their local communities. Four surveys were used to assess students' attainment through the program. The post-institute and the midterm surveys were designed to assess the students' engagement in their engineering and science practices (e.g. "Working on global problems, such as stormwater, in my local community makes me want to pursue a career in STEM"). The specific items were not presented in the article. Results revealed that students' confidence to succeed in STEM was strengthened with

the combination of mentorship and a global problem in the local community.

One article covered both MDA and TDA. Nugent et al. (2015) presented a study examining a multitude of factors related to middle school youth's STEM learning and career orientation. The students were drawn from those who participated in robotics camps across the country as part of a larger STEM education project. Students attended 40 h (1 week) of hands-on activities that focused on using the LEGO Mindstorms NXT robotics platform to build and program robots. A number of variables including STEM interest (MDA), youth career orientation (TDA), and STEM self-efficacy (TDA) were assessed using Likert scales. The STEM interest questions focused on students' value and usefulness of STEM (e.g., "It is important for me to learn how to conduct a scientific investigation"). Career orientation was separated into two dimensions—interest in STEM-related jobs and willingness of taking secondary level science and mathematics lessons. The self-efficacy items focused on detecting their confidence in robotics tasks (e.g., "I am confident that I can record data accurately"). Results revealed a positive relationship between youth STEM self-efficacy and their career orientation, which were both influenced by STEM interest. While the study focused on testing a model of social, cognitive, and affective factors contributing to youth's STEM learning,

the corresponding assessments were constrained to only capture a cross-sectional snapshot of students' self-reported learning and affective factors.

Compared with the MDA where Likert scale was used in all the 16 studies, the formats of TDA assessment were much more diverse, including open-ended questions ($n = 3$), Likert scale ($n = 7$), structured or semi-structured interviews ($n = 5$), observation of participation ($n = 4$), and willing to major in STEM-related curriculum ($n = 3$).

## Discussion

Our review shows that most STEM assessments focused on MDK, MDA, and TDA, but much less work targeted the other aspects (Fig. 5). Although most of the STEM programs aimed at improving student's interdisciplinary understanding or skills, their assessments barely address this goal. In the following, we discuss several challenges in assessing STEM education as a way to steer the research community to directions that may need more attention.

### Challenges in assessing interdisciplinary learning

As one can see in Fig. 5, interdisciplinary assessments were least used when compared to monodisciplinary and transdisciplinary assessments. The challenge of assessing



**Fig. 5** Distribution of the articles over the categories
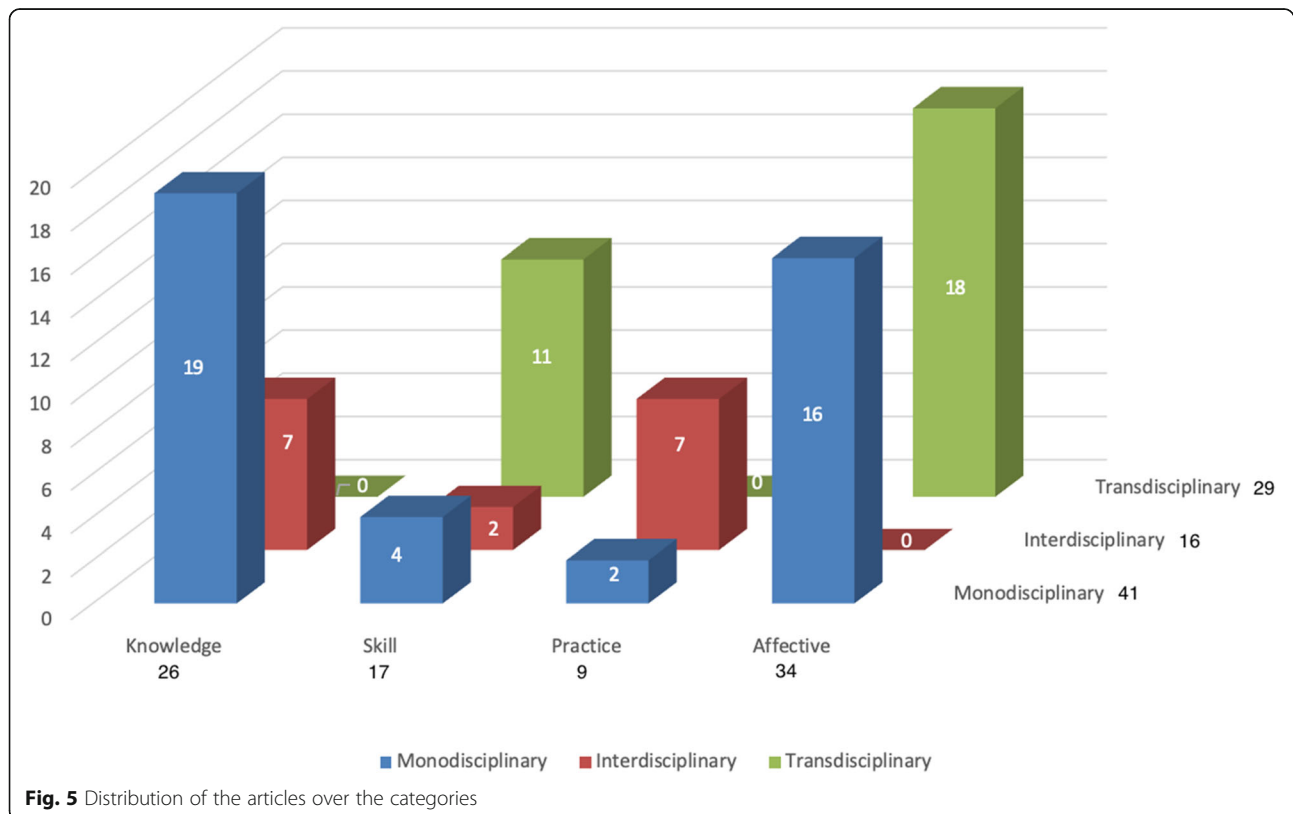
interdisciplinary learning has been well-documented (e.g., Herro et al., 2017; NAE & NRC, 2014).

The first issue is that interdisciplinarity in STEM education has been taken for granted. In reality, it is neither explicitly theorized nor well-articulated. STEM integration is not simply putting disciplines together as a conglomerate—it needs to be "intentional" and "specific" considering the connections across disciplines in the curriculum (Bryan et al., 2015; Vasquez et al., 2013). As Guzey et al. (2017) concluded in their study that "simply adding engineering into science instruction is not necessarily supportive of better student learning—teaching high-quality curriculum units that purposefully and meaningfully connect science concepts and the practices of those of engineering is essential to produce positive student outcomes" (p. 219).

The second issue is inconsistency among the curriculum, instruction, and assessment observed in these programs. As we can see in the literature, even though all the programs were categorized into STEM program, oftentimes the trace of interdisciplinary integration was implicit. All articles reviewed in this study emphasized the interdisciplinary nature of their educational programs, but very few of them went in details to describe the strategies they used to integrate and connect the disciplines.

Once the connections across disciplines are made explicit in curriculum and instruction, ideally these connections need to be assessed in order to capture students' interdisciplinary learning. As Moore et al. (2014) noted, just because these interdisciplinary connections might be emphasized in a curriculum, there is no guarantee that students will identify them or make the connections on their own.

Assessing the connections across disciplines in student understanding also helps understand if an interdisciplinary STEM education program does what the curriculum intends to. While most of the reviewed programs stated that the orientations of their curriculum were improving students' understanding of crosscutting concepts or interdisciplinary skills, few of them actually assessed them. As a result, the impacts of these programs were not properly evaluated (most likely underestimated), which hindered the effectiveness of the feedback drawn from these assessment results in order to further improve them.

## Challenges in assessing STEM learning processes/practices

Our review shows that although most of the articles ($n = 38$) developed learning activities based on engineering design, only a few ($n = 8$) assessed engineering design practices. This reflects the big picture of the challenges we all face in assessing learning processes similar to

other areas (e.g., Ashcroft & Palacio, 1996; Lederman et al., 2014).

Taking science as an example, the incorporation of practices in science education is a welcomed advancement internationally (e.g., McNeill et al., 2018; NGSS Lead States, 2013). How to assess these complex constructs framed under the three-dimensional science learning starts to take on some momentum. For instance, NRC (2014) suggested that assessment tasks had to be designed to provide evidence of students' ability to use the practices, to apply their understanding of the crosscutting concepts, and to draw on their understanding of specific disciplinary ideas, all in the context of addressing specific problems.

We would like to highlight two important aspects pertaining to assessing practices in STEM education. First of all, among the reported assessments of engineering practices, assessing final products was the most popular approach. However, assessing final products misses the critical process of iterative refinement of engineering design. This also applies to other STEM practices such as modeling (see, e.g., Namdar & Shen, 2015). The iterative nature of engineering design is particularly powerful for students since it prompts them to test and revise a solution to create the best possible outcome, encouraging idea improvement in generative learning (Crismond & Adams, 2012; Hamilton et al., 2008).

Second, there are many other learning practices key to STEM education that need to be identified and assessed. Given the diversity and heterogeneity of interdisciplinary STEM educational programs, we believe that a framework that incorporates practices (similar to that of NGSS) and a systematic set of assessment methods need to be developed. Besides engineering design, some key processes in STEM education for consideration may include problem solving (Prinsley & Baranyai, 2015), interdisciplinary reasoning and communication processes (Shen, Sung, & Zhang, 2015), and collaboration (Herro et al., 2017).

## Developing practical tools and guidelines for classroom use

Formatively, well-crafted assessments provide both the teacher and the students high-quality information about the effects of instruction and ways to improve. Developing practical tools and guidelines to assess students' learning in STEM education is still urged in classroom settings.

Among the literatures we reviewed, many reported assessments are research-oriented. Qualitative analyses such as video analyses and open coding are time and resource consuming. Therefore, these assessments typically function as research tools to advance our understanding rather than as practical tools in classroom

instruction settings. Many teachers do not have time or resources to accomplish these add-on assessments, and therefore, will not be able to use them to inform and assist student learning.

Moreover, using practical tools is more conducive to the development of longitudinal research. Similar to many other types of educational programs, interdisciplinary STEM education should aim for long-term impact (NAE & NRC, 2014). Among the 49 articles, only five longitudinal studies traced the progress of students' STEM learning and careers. Though short-term gains reflected important aspects of the effectiveness of an interdisciplinary curriculum, how students retain learning outcomes (especially those pertain to interdisciplinary learning) over time is largely unknown. Developing more practical tools could contribute to make long-term research more feasible.

Practitioners need clear guidelines and training to select and coordinate assessment formats for STEM education for different purposes. As can be seen in Table 1, a variety of formats were adopted in the reviewed works. Multiple formats/purposes for assessments should be balanced to meet student needs and provide basis for curriculum adjustments in specific settings. In practice, it is essential for teachers to align assessment and instruction for each STEM lesson (Solomon, 2003). In reality, how to choose appropriate assessment formats while keeping the assessment task feasible in different classroom settings requires further research.

## Conclusion

In our study, we reviewed assessments of interdisciplinary STEM educational programs from 2000 to 2019 and provided a two-dimensional framework of assessing students' learning in STEM education. The findings suggested that most assessments focused on assessments of monodisciplinary learning and transdisciplinary affective domains. Few assessments paid attention to interdisciplinary learning and practices.

Research in assessing interdisciplinary STEM learning has made important strides. Yet, there is still a long way to go. Based on the review, we would like to recommend the following suggestions as a way to help the community of researchers and practitioners who work in this area to better calibrate our work:

- The nature of the involved disciplines and mechanisms on how they are connected need to be made explicit in interdisciplinary STEM curriculum and instruction. More importantly, the connections across disciplines need to be operationalized and assessed to provide targeted feedback to students.
- STEM learning processes and practices are complex and manifold, especially when one considers the

many different features of disciplinary processes and practices. These core learning processes and practices need to be clearly delineated in learning objectives, and assessments need to be built around these objectives to capture the complex nature of interdisciplinary STEM learning.

- Developing practical assessment tools and guidelines for classroom use should be prioritized. While STEM education has penetrated many classrooms, most teachers have not received proper training on how to assess student learning in STEM. While our two-dimensional framework provides a theoretical starting point, building a network or repertoire of resources for practitioners would be a pragmatic step moving forward.

## Supplementary information

**Additional file 1.** Supplementary Materials: List of Reviewed Articles

**Author details**
[1]Faculty of Education, Beijing Normal University, Beijing, China. [2]Department of Teaching and Learning, School of Education and Human Development, University of Miami, Coral Gables, FL, USA.

## References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Apedoe, X. S., Reynolds, B., Ellefson, M. R., & Schunn, C. D. (2008). Bringing engineering design into high school science classrooms: The heating/cooling unit. *Journal of Science Education and Technology, 17*(5), 454–465.

Ashcroft, K., & Palacio, D. (1996). *Researching into assessment and evaluation in colleges and universities*. London, UK: Kogan Page.

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

Blackley, S., & Howell, J. (2015). A STEM narrative: 15 years in the making. *Australian Journal of Teacher Education, 40*(7).

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.

Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics, 112*(1), 3–11.

Bryan, L. A., Moore, T. J., Johnson, C. C., & Roehrig, G. H. (2015). Integrated STEM education. In C. C. Johnson, E. E. Peters-Burton, & T. J. Moore (Eds.), *STEM road map: A framework for integrated STEM education* (pp. 23–37). New York, NY: Routledge.

Bybee, R. (2010). Advancing STEM education: A 2020 vision. *Technology and Engineering Teacher, 70*(1), 30–35.

Caprile, M., Palmen, R., Sanz, & Dente, G. (2015). Encouraging STEM studies for the labour market (Directorate-General for Internal Policies: European Parliament). Brussels, Belgium: European Union. Retrieved on February 12th, 2020 from http://www.europarl.europa.eu/RegData/etudes/STUD/2015/542199/IPOL_STU%282015%29542199_EN.pdf.

Chen, C., & Lin, J. (2019). A practical action research study of the impact of maker-centered STEM-PjBL on a rural middle school in Taiwan. *International Journal of Science and Mathematics Education, 17*, 85–108.

Choi, B. C. K., & Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy. *Clinical and Investigative Medicine, 29*(6), 351–358.

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

Dierdorp, A., Bakker, A., Van Maanen, J. A., & Eijkelhof, H. (2014). Meaningful statistics in professional practices as a bridge between mathematics and science: an evaluation of a design research project. *International Journal of STEM Education, 1*(9), 1–15.

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

English, L. D. (2017). Advancing elementary and middle school STEM education. *International Journal of Science & Mathematics Education, 15*(1), 5–24.

English, L. D., King, D., & Smeed, J. (2017). Advancing integrated STEM learning through engineering design: Sixth-grade students' design and construction of earthquake resistant buildings. *The Journal of Educational Research, 110*(3), 255–271.

English, L. D., & King, D. T. (2015). STEM learning through engineering design: fourth-grade students' investigations in aerospace. *International Journal of STEM Education, 2*(14).

Fan, S. C., & Yu, K. C. (2017). How an integrative stem curriculum can benefit students in engineering design practices. *International Journal of Technology & Design Education, 27*(1), 1–23.

Griese, B., Lehmann, M., & Roesken-Winter, B. (2015). Refining questionnaire-based assessment of stem students' learning strategies. *International Journal of STEM Education, 2*(1), 12.

Guzey, S. S., Harwell, M., Moreno, M., Peralta, Y., & Moore, T. J. (2017). The impact of design-based stem integration curricula on student achievement in engineering, science, and mathematics. *Journal of Science Education & Technology, 26*(2), 207–222.

Guzey, S. S., Moore, T. J., Harwell, M., & Moreno, M. (2016). Stem integration in middle school life science: student learning and attitudes. *Journal of Science Education & Technology, 25*(4), 550–560.

Guzey, S. S., Tank, K., Wang, H., Roehrig, G., & Moore, T. (2014). A high-quality professional development for teachers of grades 3-6 for implementing engineering into classrooms. *School Science and Mathematics, 114*(3), 139–149.

Hamilton, E., Lesh, R., Lester, F., & Brilleslyper, M. (2008). Model-eliciting activities (MEAs) as a bridge between engineering education research and mathematics education research. *Advances in Engineering Education, 1*(2), 1–25.

Hernandez, P. R., Bodin, R., Elliott, J. W., Ibrahim, B., Rambo-Hernandez, K. E., Chen, T. W., et al. (2014). Connecting the stem dots: measuring the effect of an integrated engineering design intervention. *International Journal of Technology & Design Education, 24*(1), 107–120.

Herro, D., Quigley, C., Andrews, J., & Delacruz, G. (2017). Co-measure: developing an assessment for student collaboration in steam activities. *International Journal of STEM Education, 4*(1), 26.

Hoachlander, G., & Yanofsky, D. (2011). Making STEM real by infusing core academics with rigorous real-world work, linked learning pathways prepare students for both college and career. *Educational Leadership, 68*(3), 60–65.

Hurley, M. (2001). Reviewing integrated science and mathematics: the search for evidence and definitions from new perspectives. *School Science and Mathematics, 101*, 259–268.

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice*. Washington, DC: Association of American Colleges and Universities.

Klein, J. T. (2004). Interdisciplinarity and complexity: An evolving relationship. *E:CO Special Double Issue, 6*(1), 2–10.

Klein, J. T. (2008). Evaluation of interdisciplinary and transdisciplinary research: A literature review. *American Journal of Preventive Medicine, 35*(2S), 116–123.

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry-the views about scientific inquiry (VASI) questionnaire. *Journal of Research in Science Teaching, 51*(1), 65–83.

Lederman, N. G., & Niess, M. L. (1997). Integrated, interdisciplinary, or thematic instruction? Is this a question or is it questionable semantics? *School Science and Mathematics, 97*(2), 57–58.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of Clinical Epidemiology, 62*, 1–34. https://doi.org/10.1016/j.jclinepi.2009.06.006.

Lin, S. T., & Wang, M. R. (1994). *Creativity Assessment Packet*. Taipei, Taiwan: Psychological Publishing.

Lou, S., Chou, Y., Shih, R., & Chung, C. (2017). A study of creativity in $CaC_2$ steamship-derived stem project-based learning. *Eurasia Journal of Mathematics Science & Technology Education, 13*(6), 2387–2404.

Lou, S. J., Shih, R. C., Diez, C. R., & Tseng, K. H. (2011). The impact of problem-based learning strategies on stem knowledge integration and attitudes: an exploratory study among female Taiwanese senior high school students. *International Journal of Technology & Design Education, 21*(2), 195–215.

Marginson, S., Tytler, R., Freeman, B., & Roberts, K. (2013). *STEM: Country comparisons*. Melbourne, Australia: Australian Council of Learned Academies.

McCormick, R. (2004). Issues of learning and knowledge in technology education. *International Journal of Technology and Design Education, 14*(1), 21–44.

McNeill, K. L., Lowenhaupt, R. J., & Katsh-Singer, R. (2018). Instructional leadership in the era of the NGSS: principals' understandings of science practices. *Science Education, 102*(3), 452–473.

Miller, C. (1981). Varieties of interdisciplinary approaches in the social sciences: A 1981 overview. *Issues in Integrative Studies, 1*, 1–37.

Moore, T., Stohlmann, M., Wang, H., Tank, K., Glancy, A., & Roehrig, G. (2014). Implementation and integration of engineering in K-12 STEM education. In S. Purzer, J. Strobel, & M. Cardella (Eds.), *Engineering in pre-college settings: Synthesizing research, policy, and practices* (pp. 35–60). West Lafayette: Purdue University Press.

Musavi, M., Friess, W. A., James, C., & Isherwood, J. C. (2018). Changing the face of stem with stormwater research. *International Journal of STEM Education, 5*(1), 2.

Namdar, B., & Shen, J. (2015). Modeling oriented assessment in K-12 science education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education, 37*(7), 993–1023.

National Academy of Engineering (NAE) and National Research Council (NRC) (2014). *STEM integration in K-12 education: status, prospects, and an agenda for research. Committee on Integrated STEM education*. National Academy of Engineering and National Research Council of the National Academies. Washington, D.C.: The National Academies Press.

National Science and Technology Council. (2013). *A report from the committee on STEM education*. Washington, DC: National Science and Technology Council.

Nicolescu, B. (2002). *Monifesto of transdisciplinarity*. New York: State University of New York Press.

NGSS Lead States. (2013). *Next generation science standards: for states, by states*. Washington, D.C.: The National Academies Press.

NRC. (2001). *Knowing what students know: The science and design of education assessment*. Committee on the Foundations of Assessment.

NRC. (2009). *Engineering in K–12 education: Understanding the status and improving the prospects*. Washington, DC: National Academies Press.

NRC. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

NRC. (2014). *Developing assessments for the next generation science standards*. Washington, DC: The National Academies Press.

Nugent, G., Barker, B., Welch, G., Grandgenett, N., Wu, C. R., & Nelson, C. (2015). A model of factors contributing to stem learning and career orientation. *International Journal of Science Education, 37*(7), 1067–1088.

Pellegrino, J. W. (2014). A learning sciences perspective on the design and use of assessment in education. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 233–252). New York: Cambridge University Press.

Prinsley, R., & Baranyai, K. (2015). STEM skills in the workforce. Office of the Chief Scientist. *Occasional Paper Series*, 9.

Riskowski, J. L., Todd, C. D., Wee, B., Dark, M., & Harbor, J. (2009). Exploring the effectiveness of an interdisciplinary water resources engineering module in an eighth grade science course. *International Journal of Engineering Education, 25*(1), 181–195(15).

Sanders, M. (2009). STEM, STEM education, STEM mania. Technology Teacher, 68, 20–26. Retrieved from http://esdstem.pbworks.com/f/TTT + STEM+Article_1.pdf.

Sanders, M. E. (2012). Integrative STEM education as "best practice". In H. Middleton (Ed.), *Explorations of best practice in technology, design, & engineering education* (Vol. *2*, pp. 103–117). Queensland, Australia: Griffith Institute for Educational Research.

Shen, J., Liu, O., & Sung, S. (2014). Designing interdisciplinary assessments in sciences for college students: an example on osmosis. *International Journal of Science Education, 36*(11), 1773–1793.

Shen, J., Sung, S., & Zhang, D. (2015). Toward an analytic framework of interdisciplinary reasoning and communication (IRC) processes in science. *International Journal of Science Education, 37*(17), 2809–2835.

Sinatra, G. M., Mukhopadhyay, A., Allbright, T. N., Marsh, J. A., & Polikoff, M. S. (2017). Speedometry: a vehicle for promoting interest and engagement through integrated stem instruction. *The Journal of Educational Research, 110*(3), 308–316.

Solomon, G. (2003). Project-based learning: A primer. *Technology & Learning, 23*(6), 20–30.

Task Force Report, S. T. E. M. (2014). *Innovate: a blueprint for science, technology, engineering, and mathematics in California public education*. Dublin, California: Californians Dedicated to Education Foundation.

The Royal Society Science Policy Centre. (2014). *Vision for science and mathematics education*. London, United Kingdom: The Royal Society.

Vasquez, J., Schneider, C., & Comer, M. (2013). *STEM lesson essentials, grades 3–8: Integrating science, technology, engineering, and mathematics*. Portsmouth, NH: Heinemann.

You, H. S., Marshall, J. A., & Delgado, C. (2018). Assessing students' disciplinary and interdisciplinary understanding of carbon cycling. *Journal of Research in Science Teaching, 55*(3), 377–398.

Zhang, D., & Shen, J. (2015). Disciplinary foundations for solving interdisciplinary scientific problems. *International Journal of Science Education, 37*(15), 2555–2576.

## Publisher's Note