Review of Educational Research
December 2017, Vol. 87, No. 6, pp. 1042–1081
DOI: 10.3102/0034654317723009

A Bayesian Network Meta-Analysis to Synthesize the Influence of Contexts of Scaffolding Use on Cognitive Outcomes in STEM Education

Brian R. Belland, Andrew E. Walker, and Nam Ju Kim
Utah State University

Computer-based scaffolding provides temporary support that enables students to participate in and become more proficient at complex skills like problem solving, argumentation, and evaluation. While meta-analyses have addressed between-subject differences on cognitive outcomes resulting from scaffolding, none has addressed within-subject gains. This leaves much quantitative scaffolding literature not covered by existing meta-analyses. To address this gap, this study used Bayesian network meta-analysis to synthesize within-subjects (pre–post) differences resulting from scaffolding in 56 studies. We generated the posterior distribution using 20,000 Markov Chain Monte Carlo samples. Scaffolding has a consistently strong effect across student populations, STEM (science, technology, engineering, and mathematics) disciplines, and assessment levels, and a strong effect when used with most problem-centered instructional models (exception: inquiry-based learning and modeling visualization) and educational levels (exception: secondary education). Results also indicate some promising areas for future scaffolding research, including scaffolding among students with learning disabilities, for whom the effect size was particularly large ($\bar{g} = 3.13$).

Keywords: scaffold, Bayesian network meta-analysis, STEM, cognitive tutor, problem-centered instruction, intelligent tutoring systems

Having originated as a naturalistic description of how adults help toddlers learn solve problems (Wood, Bruner, & Ross, 1976), scaffolding has expanded to one that is used among diverse learners and in the context of many problem-centered instructional approaches (Hawkins & Pea, 1987; Hmelo-Silver, Duncan, & Chinn, 2007; Stone, 1998). Along with this expansion, many scaffolding approaches, forms, and empirical studies, have emerged. For example, scaffolding now encompasses one-to-one interactions with classroom teachers (van de Pol, Volman, & Beishuizen, 2010), interaction with similarly abled peers (Pifarre & Cobos, 2010), and computer-based tools (Devolder, Van Braak, & Tondeur, 2012; Reiser, 2004). Scaffolding is used among students of diverse educational levels and demographic

backgrounds (Cuevas, Fiore, & Oser, 2002; Hadwin, Wozney, & Pontin, 2005). Furthermore, scaffolding is often designed to affect knowledge and skills beyond problem-solving ability, including argumentation ability (Jeong & Joung, 2007) and deep content knowledge (Davis & Linn, 2000). Synthesizing work on this expanded conceptualization of scaffolding is important to help researchers and designers determine what works best in scaffolding among particular populations and contexts. Scaffolding synthesis work has been done, but all focus on between-subjects differences (Belland, Walker, Kim, & Lefler, 2017; Belland, Walker, Olsen, & Leary, 2015; Ma, Adesope, Nesbit, & Liu, 2014; Steenbergen-Hu & Cooper, 2013, 2014; Swanson & Deshler, 2003; Swanson & Lussier, 2001; VanLehn, 2011), leaving important questions of how much within-subject growth one might expect among average students unaddressed. In this article, we address this gap by using Bayesian network meta-analysis to synthesize pre–post growth among networks of student populations, STEM (science, technology, engineering, and mathematics) disciplines, educational levels, and assessment levels (Berger, 2013; Lumley, 2002; Mills, Thorlund, & Ioannidis, 2013).

Literature Review

Scaffolding Definition

Scaffolding can be defined as contingent support that structures and highlights the complexity inherent in problem solving, thereby supporting current performance and promoting skill gain (Reiser, 2004; Wood et al., 1976). Three key attributes characterize scaffolding: contingency, intersubjectivity, and transfer of responsibility (Wood et al., 1976). First, scaffolding is contingent on dynamic assessment, which indicates students' current abilities and where they need support. Scaffolding can be provided initially, and as dynamic assessment indicates that students are gaining skill or facing additional challenges, scaffolding can be faded or added, respectively (Collins, Brown, & Newman, 1989; Murray, 1999; Wood et al., 1976). Next, students need to recognize successful performance on the scaffolded task (Wood et al., 1976). Finally, scaffolding needs to engender independent task completion.

The concept of instructional scaffolding originated in describing one-to-one interactions with an ever-present tutor (Wood et al., 1976). Soon, researchers began to think about how the technique could be leveraged in other settings. One such way was one-to-one interactions from a classroom teacher who provided individualized help as students engaged with problems (van de Pol et al., 2010). Scaffolding is now used in the context of many instructional approaches, including project-based learning, problem-based learning, inquiry-based learning, and design-based learning (Belland, 2017). At the center of each is an ill-structured problem, defined as a problem that does not have just one correct solution, and which has multiple solution paths (Jonassen, 2011). To address such a problem, it is necessary to represent the problem qualitatively so as to recognize the critical factors and how they interact (Jonassen, 2003). Still, each problem-centered approach involves a different set of expectations, both in terms of process and product. For example, in design-based learning, students iterate designs that address the central problem (e.g., levee to prevent beach erosion of barrier islands;

Kolodner et al., 2003); meanwhile, in inquiry-based learning, students pose and address their own questions (Keys & Bryan, 2001).

As computing power increased, researchers began to think about how computer tools could provide scaffolding (Hawkins & Pea, 1987). Computer-based scaffolding is often designed to (a) help students with what to consider when addressing a problem (conceptual scaffolding), (b) bootstrap a strategy for addressing a problem (strategic scaffolding), (c) invite students to question their own understanding (metacognitive scaffolding), and (d) enhance interest, autonomy, self-efficacy, and other motivational variables (motivation scaffolding; Belland, Kim, & Hannafin, 2013; Hannafin, Land, & Oliver, 1999; Rienties et al., 2012). Specific strategies embedded in scaffolding include cognitive support such as highlighting critical problem features, modeling expert processes and demonstration, and motivational support such as recruitment, direction maintenance, and controlling frustration (van de Pol et al., 2010; Wood et al., 1976).

Existing Scaffolding Meta-Analyses

Some work has been done to synthesize existing empirical work, but most such synthesis work focuses on between-subjects differences—how students who used scaffolding performed when compared with the performance of students who did not use scaffolding. This is undeniably a crucial way to gauge the impact of an intervention, and it indicates that scaffolding is a highly effective intervention. Meta-analyses of between-group differences indicated that students using a variety of scaffolding types performed 0.53 (Belland et al., 2015) and 0.46 (Belland et al., 2017) *SD*s better than their control counterparts. Meta-analyses have also been performed among specific subtypes of scaffolding, such as that in intelligent tutoring systems (Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011), dynamic assessment (Swanson & Lussier, 2001), and scaffolding for students with learning disabilities (Swanson & Deshler, 2003), indicating that scaffolding can help experimental students perform substantially better than control students. While this work is important, it does not speak to the magnitude of cognitive growth that one might see in students who use scaffolding. There is a need for synthesis of pre–post cognitive growth resulting from scaffolding, and how that growth varies based on differences in the context in which scaffolding is used. The technique of network meta-analysis, which can address such growth, has emerged in medical research (Jansen et al., 2011; Lumley, 2002; Mills et al., 2013) and has potential in education research.

Contextual Issues Related to Scaffolding

It makes little sense to try to find a universal design for scaffolding that is most effective because scaffolding (a) employs a wide range of strategies that are grounded in different theories (Koedinger & Corbett, 2006; Puntambekar & Kolodner, 2005; Quintana et al., 2004; van de Pol et al., 2010) and (b) is used in the context of many different problem-centered instructional approaches and subject matters, and by learners diverse in grade level and demographics (Hmelo-Silver et al., 2007; Lin et al., 2012; Puntambekar & Hübscher, 2005; Stone, 1998).

Learners

The age level with which scaffolding is used has expanded from preschool to K–12, college, graduate, and adult. Scaffolding can be seen as potentially a good fit for such a wide range of different age groups in that all need to learn to address ill-structured problems (Jonassen, 2011). The need to be able to address ill-structured problems is reflected in the needs of employers (Carnevale & Desrochers, 2003) and is at the center of the Common Core (McLaughlin & Overturf, 2012; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and Next Generation Science Standards (Achieve, 2013; Krajcik, Codere, Dahsah, Bayer, & Mun, 2014). At the same time, it is likely that different combinations of scaffolding strategies need to be used across these different age groups. A comprehensive traditional meta-analysis indicated that effect sizes for computer-based scaffolding were higher among adult learners than among college, secondary, middle-level, or primary students (Belland et al., 2017). Still, it is natural to question whether the strength of pre–post gains of computer-based scaffolding varies based on education level.

The original education population among which the definition of instructional scaffolding was grounded was middle class and average-achieving (Wood et al., 1976). But with the expansion of the metaphor, scaffolding began to be used among students with a much wider range of demographic characteristics. Early efforts found success using scaffolding among lower achieving students (Dimino, Gersten, Carnine, & Blake, 1990; Palincsar & Brown, 1984) and students with learning disabilities (Englert, Raphael, Anthony, & Stevens, 1991; Stone, 1998). With the more widespread use of computer-based scaffolding, so too did computer-based scaffolding begin to be used among a wide range of learners, including students from traditional, low socioeconomic status (SES), and underrepresented backgrounds, as well as those who are lower achieving and higher achieving (Belland, 2017). Traditional meta-analysis efforts indicated that scaffolding leads to stronger between-subject effects among traditional students than among underperforming students (Belland et al., 2017). But it is also worthwhile to consider whether within-subject (pre–post) differences vary based on education population. This can be done through network meta-analysis.

Context of Use

The context in which scaffolding is used can vary widely, and this variation is associated with real differences in scaffolding strategy (Belland, 2017). Differences in context of use can be considered from two perspectives—the problem-centered instructional model with which scaffolding is used, and the subject matter in which the instruction is situated. Problem-centered instructional models with which scaffolding is used include project-based learning, problem-based learning, inquiry-based learning, design-based learning, case-based learning, and problem solving (Belland, 2017). These models all involve addressing an ill-structured problem, but the nature of the problem and what should be produced, as well as inherent structure for student learning, varies between the models. For example, problem-based learning is the most open-ended in that students are expected to produce and argue for a conceptual solution to the problem (Hmelo-Silver, 2004), while design-based learning and project-based learning constrain the solution type (e.g., video or designed product) students need to produce. The stages through which students

need to progress vary according to model as well. With such variation in process and product, it is natural to question if corresponding within-subject effect sizes vary. This can be addressed through network meta-analysis.

Problem-centered models and the nature of central problems tend to cluster differently by subject matter. For example, design-based learning (Chandrasekaran, Stojcevski, Littlefair, & Joordens, 2013; Silk, Schunn, & Cary, 2009) and problem-based learning (Galand, Frenay, & Raucent, 2012; Yadav, Subedi, Lundeberg, & Bunting, 2011) are often used in engineering education. Inquiry-based learning (Edelson, Gordin, & Pea, 1999; Marx et al., 2004) and project-based learning (Barron et al., 1998; Krajcik et al., 1998) tend to cluster in science education.

Assessment Level

Crucial to examining scaffolding outcomes is determining whether the magnitude of pre–post gains of scaffolding depend on assessment level, defined as the nature of learning outcome targeted by assessment. Assessment levels include concept (ability to state definitions of basic knowledge), principles (ability to describe or use relationships between facts), and application (ability to use concept- and principles-level knowledge to address a new problem; Sugrue, 1995). Traditional meta-analysis indicated that scaffolding's effect was greater when measured at the principles level than at the concept level (Belland et al., 2017).

Bayesian Network Meta-Analysis as a Potential Solution

Two techniques that can help researchers establish equivalence on response variables before the treatment is introduced are random selection and random assignment (Higgins et al., 2011). But little education research incorporates true random selection and assignment of participants, leading to high risk of bias in randomization (Higgins et al., 2011). Another method is to use students as their own controls through the use of a pretest that is equivalent to the posttest. Network meta-analysis allows one to synthesize pre–post differences across studies in order to make indirect comparisons between treatments that may not have been compared directly in any single study (Lumley, 2002; Mills et al., 2013). When taking a frequentist approach to network meta-analysis, all included studies need to contain a treatment and a control condition (Puhan et al., 2014). Thus, studies with multiple versions of a scaffolding treatment but no lecture control treatment cannot be included in a frequentist network meta-analysis. Taking a Bayesian approach to network meta-analysis allows researchers to include multiple treatment studies as long as each study has a treatment in common with another study (Bhatnagar, Lakshmi, & Jeyashree, 2014; Goring et al., 2016). Furthermore, taking a Bayesian approach sets up a decision-making framework that scaffolding researchers and funders can use to indicate which contexts hold the greatest promise for scaffolding (Jansen et al., 2011).

Research Questions

1. To what extent do learner characteristics moderate cognitive pre–post gains resulting from scaffolding?
    a. To what extent does education level among which scaffolding was used moderate cognitive pre–post gains?

   b. To what extent does education population among which scaffolding was used moderate cognitive pre–post gains?
2. To what extent does the context in which scaffolding is used moderate cognitive pre–post gains?
   a. To what extent does context of use of scaffolding moderate cognitive pre–post gains?
   b. To what extent does STEM discipline within which scaffolding was used moderate cognitive pre–post gains?
3. To what extent does assessment level moderate cognitive pre–post gains resulting from scaffolding?

Method

Design

For this synthesis effort, we followed a network meta-analysis approach from a Bayesian perspective. Network meta-analyses allow researchers to make direct and indirect comparisons of pre–post gains of different interventions that have a common comparator (Mills et al., 2013). Two principal advantages of network meta-analysis are its capacity to allow researchers to (a) make indirect comparisons among treatments that were never compared in a single study and (b) rank treatments according to effectiveness (Mills et al., 2013). However, the reliability of the indirect comparisons and rankings depends on the number of direct comparisons that are included in the network (Lumley, 2002; Mills et al., 2013). Furthermore, when the number of studies that represent a certain level of a moderator is low, the results for those moderator levels can be overweighted or biased. When (a) the number of direct comparisons among moderator levels is low and (b) there is no common comparator between moderator levels, one may opt to take a Bayesian approach to analysis. At a high level, in Bayesian approaches, rather than simply calculating the distribution of a collected sample without reference to what is already known (as one would do with a frequentist approach), one (a) determines possible prior distributions (considers what is already known about the distribution of the construct in the population of interest), (b) collects data from a sample, and (c) empirically approximates the posterior distribution (through, e.g., Markov Chain Monte Carlo [MCMC] sampling; see Figure 1; Carlin & Chib, 1995; Little, 2006; Lunn, Thomas, Best, & Spiegelhalter, 2000). For a relatively comprehensive and user-friendly introduction to Bayesian data analysis approaches, readers are directed to Gelman et al. (2013).

Following a Bayesian approach requires that one establish a prior distribution, defined as the distribution of the parameters in question according to prior research. All relevant prior meta-analyses about computer-based scaffolding focused on between-subject, rather than within-subject differences. Therefore, existing meta-analysis results are ill-equipped to form an informative prior distribution in this study. Furthermore, we wanted the current coding, rather than a prior distribution informed by between-subjects effects, to primarily drive the approximation of the posterior distribution (Jansen, Crawford, Bergman, & Stam, 2008). Therefore, this article employs a noninformative prior distribution model, which can be used when there is insufficient information about a treatment's effectiveness or there is no consensus about the effectiveness among scholars.

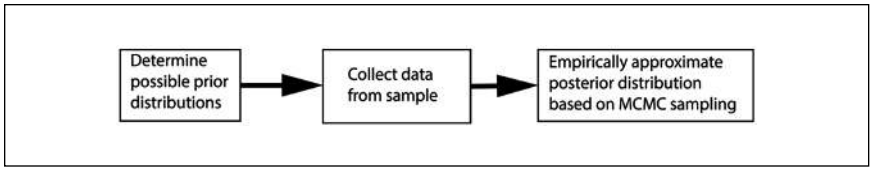

FIGURE 1. *Basic Bayesian approach.*

Among several possible prior noninformative distribution models, which have different assumptions about the variance between studies (e.g., maximum and minimum tau values), uniform prior distribution on tau (0, 5) was selected by deviance information criterion statistics (see Supplementary Table S1 in the online version of the journal), which evaluate and compare generated Bayesian models (Spiegelhalter, Best, Carlin, & van der Linde, 2002).

Next, one collects current data—in this study, this is our coding of articles collected through literature search. Then, one runs MCMC simulations informed by the prior distribution and the current data to empirically approximate the posterior distribution, defined as the distribution of true parameters. We did this using WinBUGs (Lunn et al., 2000; see Supplementary Table S1 in the online version of the journal for our WinBUGS code). Readers who are interested in learning more about how to perform the process of running calculations for a Bayesian network meta-analysis with the combination of STATA and WinBUGS are directed to the screencast available in Supplementary Video S2 in the online version of the journal. Readers interested in learning more about the foundations and application of coordinating Bayesian analysis between STATA and WinBUGS are directed to Thompson (2014). Many of the principles behind the commands and processes would be similar if combining WinBUGS with other statistical packages like R or SAS.

Literature Search

We used a three-pronged literature search to identify 7,589 potential studies, which were published between January 1, 1993, and December 31, 2015 (see Figure 2). The databases searched were ProQuest, Education Source, psycINFO, CiteSeer, ERIC, Digital Dissertations, PubMed, Academic Search Premier, IEEE, and Google Scholar, and search terms used were various combinations of the following terms: *scaffold**, *tutor**, *computer**, *intelligent tutoring system**, and *cognitive tutor**. Hand searches were conducted in journals that were recommended by experts or where we had found articles related to scaffolding in mathematics and engineering education: *Journal for Research in Mathematics Education*, *International Journal of Mathematical Education in Science and Technology*, *Journal of Professional Issues in Engineering Education and Practice*, and *Computer Applications in Engineering Education*. To gain additional coverage in the areas of special education and adult learning, we conducted hand searches in the following journals: *Journal of Special Education*, *Journal of Special Education Technology*, *BMC Medical Education*, and *Journal of Medical Education*. We ended up finding no potentially includable studies from *BMC Medical Education*

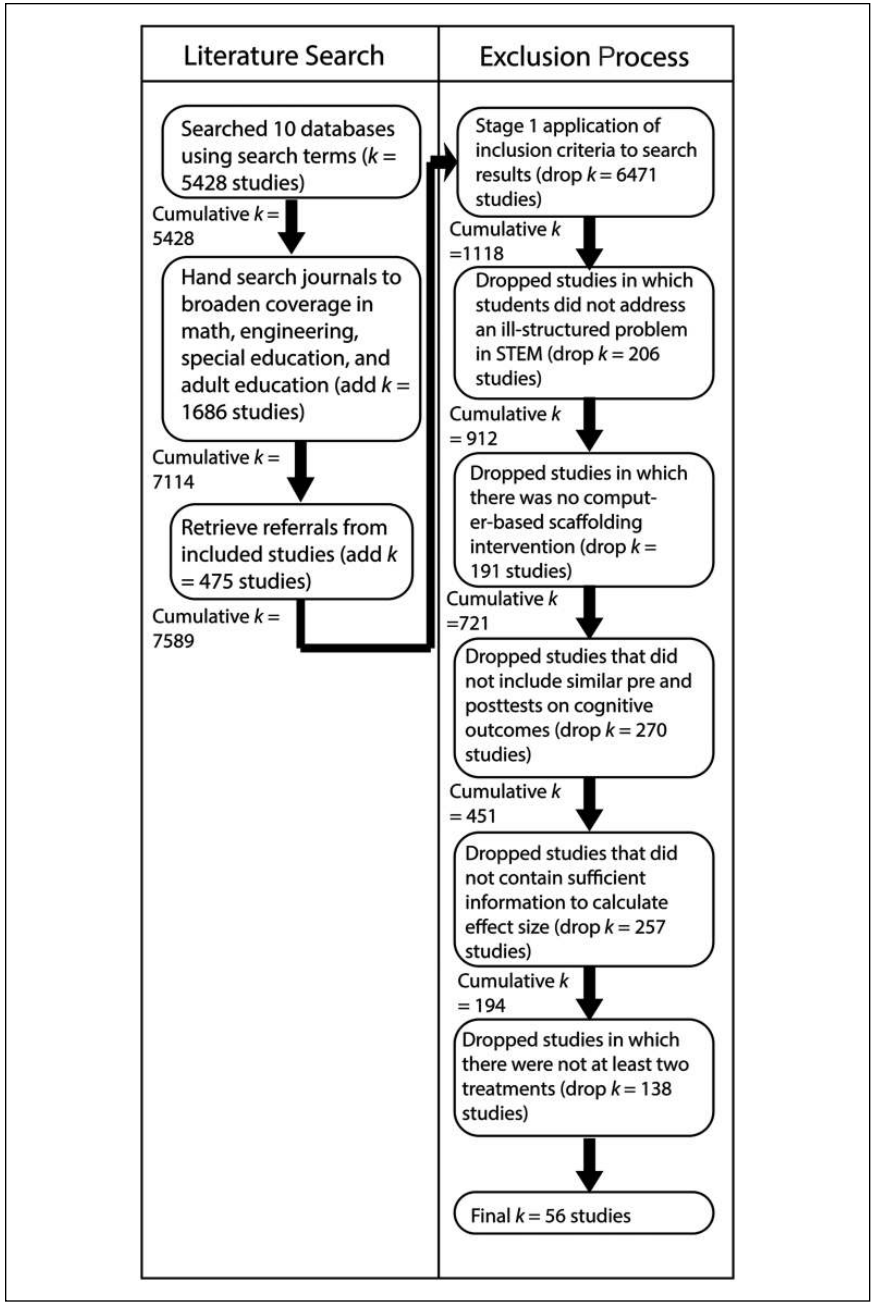

FIGURE 2. *Number of studies added at each stage of literature search and dropped at each stage of the exclusion process.*

or the *Journal of Medical Education*. Referrals were studies in the reference lists of included studies.

Application of Inclusion Criteria

Inclusion criteria were that (a) participants addressed an ill-structured problem in one of the STEM fields (science, technology, engineering, and mathematics); (b) participants used a computer-based scaffolding intervention; (c) participants took a similar pretest and posttest covering a cognitive variable; (d) sufficient statistics were reported to calculate effect size; and (e) there were at least two treatments. We defined ill-structured problems as those for which qualitative representation of the problem was necessary, and not all necessary information to do so were presented to students (Jonassen, 2011). All included studies had to have a treatment in common with at least one other study (Mills et al., 2013). Thus, if a study compared two scaffolding types that were not examined in any other study, then it would be excluded. When more than one study reported the same data, the one with the most information (e.g., dissertation) was retained.

Application of inclusion criteria proceeded in a two-stage manner. In Stage 1, the inclusion criteria were applied in a pre–pass manner to winnow the list of studies that resulted from the literature search (see Figure 2 for the number of studies dropped according to element of the exclusion process). Specifically, one researcher applied the inclusion criteria and only removed a study from consideration if it clearly did not meet the inclusion criteria. Stage 1 resulted in dropping $k = 6{,}471$ studies that resulted from the literature search ($k = 7{,}589$).

In Stage 2, alternating pairs of researchers read each article resulting from Stage 1 and applied inclusion criteria. Based on our inclusion criteria, 1,062 studies were excluded. The final number of included studies was $k = 56$. Stage 2 resulted in dropping a total of 1,062 of the studies remaining after Stage 1. The number of included outcomes varies slightly by moderator analysis, as detailed in the Results section (see Supplementary Table S1 in the online version of the journal for a list of included studies).

Coding Scheme

Articles were coded for the following characteristics—education population, education level, STEM discipline, and assessment level. Our coding process, along with examples from the coded studies, are shared in the following paragraphs.

Effect Size Calculations

All included studies had at least two treatments—usually one scaffolding treatment and a lecture control condition, but sometimes two different scaffolding treatments. For each treatment group, the sample size, pretest mean, pretest standard deviation, posttest mean, and posttest standard deviation were inputted into a free online tool (http://esfree.usu.edu/) to calculate effect size. All reported effect sizes used the Hedges's *g* calculation. Hedges's *g* was chosen because it (a) uses pooled standard deviation, which has the potential to be less biased than effect size estimates that use the control group standard deviation, and (b) is weighted according to sample size (Hedges, 1982).

Education Level (Primary/K–5, Middle Level/6–8, Secondary/9–12, College/ Vocational/Technical, Graduate/Professional, Adult)

Education level was coded as (a) primary when the majority of participants were enrolled in Grades K–5, (b) middle level when the majority of participants were enrolled in Grades 6 to 8, (c) secondary when the majority of participants were enrolled in Grades 9 to 12, (d) college/vocational/technical when the majority of participants were enrolled in a 4-year bachelor's program or 2-year associate's program, (e) graduate when the majority of participants were enrolled in a graduate degree program (e.g., master's or doctorate), or (f) adult when the majority of participants were over the age of 18 years but not enrolled in a college or graduate-level program.

Education Population (Traditional, High-Performing, Underperforming, English Language Learners (ELL), Underrepresented, and Persons With Learning Disabilities)

Education population refers to participant characteristics that may be associated with differences in educational outcomes in STEM (Heinrich, Knight, Collins, & Spriggs, 2016; Hernandez, Schultz, Estrada, Woodcock, & Chance, 2013; Molina, Borror, & Desir, 2016; Williams, Thomas, Ernst, & Kaui, 2015). Participants were coded as traditional when no argument was made that the majority of participants had a demographic characteristic or preexisting performance levels that makes them substantially different from students representing majority characteristics and typical performance for the country of study. For example, Chen, Kao, and Sheu (2003) noted having chosen their three participating schools because they were "located near 3 of the 10 best bird-watching sites in Taiwan" (p. 355). This was important because the scaffolding was aimed at helping students solve problems related to bird identification in the field. However, it does not have any bearing on education population characteristics, and thus participants were labeled as traditional. Sometimes, authors labeled participants as high-performing or low-performing based on preexisting measures of performance. For example, Liu (2004) reported pretest and posttest means separately for students who were identified as talented and gifted, those in the regular track, and those with learning disabilities or who were ELL. The author made the case that such groups represented high performers, traditional students, and underperformers, respectively. Sometimes, an argument was made that the entire school was high-performing or low-performing, and the education population was coded accordingly. For example, students in one study were coded as high-achieving because the study authors identified the participating school as having consistently ranked in the top 10 in its country according to an academic measure (Tan, Loong, & So, 2005). Education population was coded as ELL when the majority of participants spoke English as a second language but were instructed in English. For example, test scores were broken down according to participating school in Songer, Lee, and McDonald (2003). At one such school, only 38% of students spoke English as a primary language, and thus the corresponding scores were coded as ELL. Education population was coded as underrepresented when most participants are not typically represented in the target discipline. For example, participants in Bulu and Pedersen (2010) were 50% Hispanic and 35% African American, and the domain was science, where

individuals from these groups are underrepresented. Education population was coded as persons with learning disabilities when the majority of participants had a documented disability for which an individualized education program would be prepared and which would interfere with learning the target content. For example, of the nine elementary school students who used scaffolding in the context of mathematics instruction in Xin et al. (2017), three had learning disabilities, one had attention deficit hyperactivity disorder, and one had a mild intellectual disability.

Instructional Approach (Problem-Based Learning With Scaffolding, Project-Based Learning With Scaffolding, Inquiry-Based Learning With Scaffolding, Case-Based Learning With Scaffolding, Design-Based Learning With Scaffolding, Modeling/Visualization With Scaffolding, and Problem Solving With Scaffolding)

Problem-based learning with scaffolding was identified when (a) the problem was presented first, and was the driver of all subsequent learning; (b) teachers served as facilitators rather than information providers; and (c) computer-based scaffolding was provided (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004). For example, in Liu (2004), middle school students were presented with an ill-structured problem in which aliens are stranded, and needed to find a new home within our solar system. Student learning about characteristics of planets was driven by this problem, and teachers served as facilitators, rather than information providers. Project-based learning with scaffolding needed to involve learning focused toward the production of a real-world project/deliverable related to the central problem, and computer-based scaffolding needed to be provided (Helle, Tynjälä, & Olkinuora, 2006; Krajcik et al., 1998). For example, in Barak and Dori (2005), students addressed a sequence of chemistry problems and needed to construct a chemical model of the chemical that would address the problem. In inquiry-based learning with scaffolding, students needed to pose one or more question(s) related to the problem, devise and carry out a method to address the question(s), and be provided scaffolding (Crippen & Archambault, 2012; Edelson et al., 1999). For example, in Ardac and Sezen (2002), students used simulation software in which they could ask questions that they could then address by manipulating different variables related to a chemical reaction. In case-based learning with scaffolding, all necessary information is given to students often via lecture, then a case is provided, and students need to solve the case using the provided information and with the aid of scaffolding (Srinivasan, Wilkes, Stevenson, Nguyen, & Slavin, 2007; Thistlethwaite et al., 2012). For example, in Feyzi-Behnagh et al. (2014), participants needed to solve unique cases related to dermatology. Design-based learning with scaffolding was coded when students were invited to design and/or produce a product that would address an ill-structured problem (Kolodner et al., 2003; Silk et al., 2009). For example, Puntambekar, Stylianou, and Hübscher (2003) invited students to address authentic problems related to force by designing artifacts like roller coasters. Problem solving with scaffolding was identified when students needed to address an ill-structured problem, but the problem centered instructional model could not be classified as problem-based learning, project-based learning, inquiry-based learning, case-based learning, design-based learning, or modeling/visualization.

STEM Discipline (Science, Technology, Engineering, Mathematics)

We coded this category according to the problem students were addressing, rather than the discipline of the class in which participants were enrolled. We always coded according to a broad category (e.g., engineering), and a narrower category (e.g., electrical engineering). This decision was made for two reasons: (a) the subject matter of the class did not always align with the nature of the problem being addressed, and the nature of the problem being addressed was deemed to be more important to an examination of scaffolding; (b) participants were not always drawn from a formal class. As an example of the first point, participants in Magana (2014) were in an introductory educational computing course, but were addressing a problem related to scale (nanoscale, microscale, and macroscale). Because the goal was that students be able to order, classify, and sort shapes according to scale, the STEM discipline was coded as mathematics. As an example of the second point, participants in Chen, Kao, and Sheu (2005) engaged in a mobile butterfly watching activity. Within the study, participants needed to compare photographs they took with database photos of butterflies; for this reason, it was coded as science–ecology. There was a focus on engineering implications of electrical current in another study (de Jong, Härtel, Swaak, & van Joolingen, 1996). So while the participants were high school students enrolled in physics and engineering courses, the study was coded as electrical engineering.

Assessment Level (Concept, Principles, and Application)

Assessments were labeled on the basis of what students were asked to know and do with the target knowledge (Sugrue, 1995). Concept-level assessments measured whether participants knew basic knowledge. For example, a pretest and a posttest in one study asked declarative knowledge questions about scientific instruments, the solar system, and planet characteristics (Bulu & Pedersen, 2010). Principles-level assessment was coded when participants were asked to identify relationships/connections between facts, either in terms of directionality or scale. For example, an assessment invited students to read a scenario in which scientists were investigating a phenomenon, and students needed to indicate the hypotheses that was being tested (Tan et al., 2005). Application-level assessment was coded when participants needed to apply concept-level knowledge and principles-level knowledge to a new holistic/authentic problem. For example, high school students needed to use physics knowledge and principles to describe how a shuffle stone moves across a shuffleboard (Gijlers, 2005).

Coding Process

Alternating pairs of coders from a pool of four researchers with expertise in scaffolding, meta-analysis, or both, coded the studies. Two researchers independently coded each study, and then met to discuss coding discrepancies and come to consensus. We used Krippendorff's alpha to assess interrater reliability on initial coding because it can handle the range of scales (nominal, ordinal, and ratio) present in our coding data, and it adjusts for chance agreement (Krippendorff, 2004). Because Krippendorff's alpha adjusts for chance agreement, is appropriate to use with multiple scales, and can account for unused scale points, its values are typically lower than other popular indices of agreement such as percentage

agreement and Cohen's kappa, and thus should not be interpreted in light of such statistics. Two coders were drawn from a pool of 4, and 218 data points were used for the interrater reliability analysis. All alphas were greater than .667 (see Supplementary Table S1 in the online version of the journal), which represents the minimum standard for acceptable reliability (Krippendorff, 2004). The lowest Krippendorff's alpha values: .731 for assessment level, and .761 for context of use, were further analyzed using the *q* test bootstrapping method to examine the probability that the statistics were actually lower than .667 (Hayes & Krippendorff, 2007). *q* test results for assessment level coding shows that the chance of obtaining an alpha value below .67 was 3.13%; in other words, if the population of units were coded, reliability would likely be somewhere within the confidence interval for α_{true} of .67 to .79 (see Supplementary Table S1 in the online version of the journal). *q* test results for Context of Use show 90% probability that the alpha value was above .67 (see Supplementary Table S1 in the online version of the journal). While the probability to get an alpha value below .67 was 10%, the alpha value distribution followed a normal distribution, $p > .05$; thus, there was no concern about low reliability between coders.

Consensus codes were used in all analyses. An earlier version of the coding scheme was developed in two ways—through synthesis of the scaffolding literature and development of in vivo codes; this was then used for a pilot scaffolding meta-analysis project (Belland et al., 2015). We presented the coding scheme and our suggested additions to encompass a broader swath of literature to our advisory board. They then either confirmed that the coding categories and their associated levels were reasonable or suggested revisions. The revised coding scheme was then used in a comprehensive, traditional meta-analysis (Belland et al., 2017), and, with the exception of the calculation of ESs, the coding categories used in this article were the same.

Meta-Analytic Procedures/Statistical Analyses

The wide range of participants, context of use, study measures, and educational levels makes it unlikely that each outcome represents an approximation of a single true ES. This led us to use a random effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009). Analyses were conducted using the metan package of STATA 14 and WinBUGS 1.4.3. Specifically, WinBUGS 1.4.3 was used to run MCMC simulations using Gibbs sampling. We used 20,000 MCMC samples for each analysis. This study used the 2-level model $\theta_i = \beta_0 + \beta_{1x_{i1}} + \beta_{2x_{i2}} + \beta_{px_{ip}} + \delta_i + e_i$ including within and between study-level covariates for every moderator (Raudenbush, 2009). x_{ip} identifies study-level coding and β_p represents the regression coefficient. The random effect of studies, δ_i, has the following distribution: $\delta_i \sim N(0, \tau^2)$ and the sampling error, e_i, has a mean of zero and a sampling variance of σ. The seed for the random number generator was 1234 as the default setting and the starting value for beta and gamma parameter was zero. A total of 22,500 iterations for estimation of posterior distribution were generated by MCMC and 2,500 initial iterations were burned in to remove randomized initial values in every model for moderators in this study. Furthermore, we validated our models with graphical summaries (i.e., trace plot, autocorrelation, histogram and density plots). The pattern of trace was stable as the iteration number increased and the value of autocorrelation approached 0 as the lag increased.

Using a Bayesian approach helps address small study effects (Kay, Nelson, & Hekler, 2016; Mengersen, Drovandi, Robert, Pyne, & Gore, 2016). But another potential problem of publication bias is the file drawer problem, according to which studies with negative or no effects are often not published. To guard against this threat, we examined the underlying coding. We only found two positive outliers (z score >+3 *SD*), but no evidence of systematic bias. The conclusion of no systematic bias is further supported because (a) our previously published traditional meta-analysis indicated no publication bias in the literature on computer-based scaffolding in STEM education of the same time period (Belland et al., 2017), and (b) any publication bias inherent in the noninformed prior distribution assumptions and observed data will be corrected for with the posterior estimates (Kay et al., 2016; Mengersen et al., 2016) The two outliers were Galleto and Refugio (2012) and Kramarski and Zeichner (2001). Because there was no evidence of systematic bias, the mentioned studies were maintained in the list of included studies.

The presence of similar pretests and posttests within the same study can present a risk of testing bias. Within the overall Bayesian network meta-analysis of scaffolding in STEM education project, we also wrote an article covering scaffolding characteristics and risk of bias—a lens with which to code research quality that does not make assumptions when data are not present (Higgins et al., 2011). Results showed that there was no substantial risk of bias due to testing effect (Walker, Belland, Kim, & Piland, 2017).

MCMC simulations generate the posterior distribution, which represents the range of true ESs for each moderator. Using Bayesian probability, one can calculate the probability that each moderator level is the best (Jansen et al., 2011). We report this as "probability of the best." One can also calculate the probability that each moderator level is second best, third best, and so on. Averaging all such probability levels together for each moderator level allows one to arrive at a rank order for the levels of the moderator. We report this as "ranking."

The goal of Bayesian network meta-analysis is to model a network of evidence pertaining to scaffolding treatments and common treatments—sometimes lecture-based controls and sometimes other scaffolding treatments. Because not all scaffolding treatments will have been compared directly with control, it does not make sense to calculate a two-node network computing one effect size estimate for all scaffolding treatments versus control (Lumley, 2002).

Results

Research Question 1: To What Extent Do Learner Characteristics Moderate Cognitive Pre–Post Gains Resulting From Scaffolding?

Education Level

When interpreting the network plot (see Supplementary Figure S3 in the online version of the journal), one can see the number of unique outcomes for each level (e.g., middle level) of the target characteristic (e.g., education level). Each solid line between two circles represents the number of direct comparisons between the two levels of the target characteristic. For example, the solid line between middle level and control shows that there were eight direct comparisons of middle-level

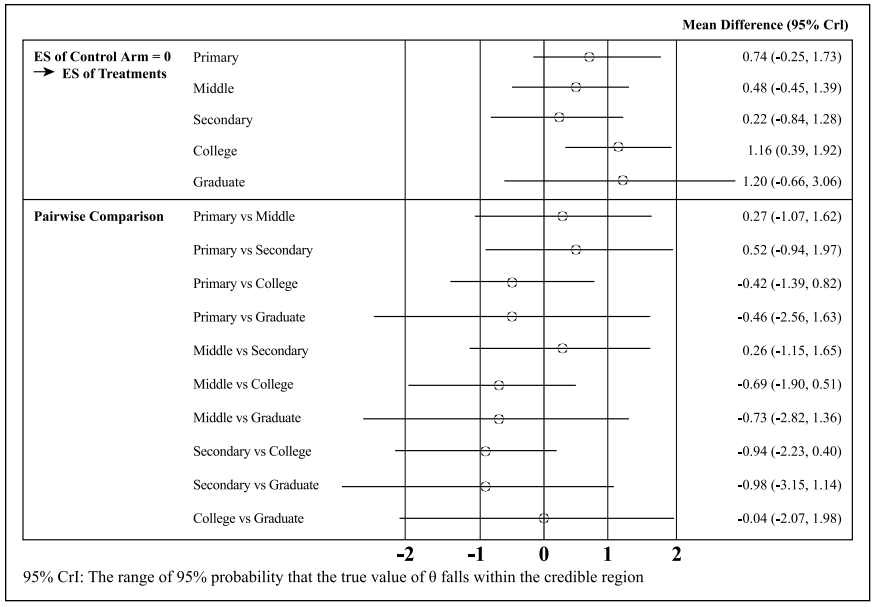

FIGURE 3. *Effect size (ES) estimates and 95% credible intervals (Crl) of scaffolding according to education level.*

students using scaffolding with students in a control condition. Of note, for education population, there are no studies that compared students at different educational levels, which is to be expected. Dotted lines indicate indirect comparison information that can be ascertained among treatment characteristics that were never directly compared in a single study. The number of outcomes were (a) greatest at the college/vocational/technical level ($k = 12$); (b) roughly equivalent among primary ($k = 7$), middle level ($k = 8$), and secondary ($k = 6$); and (c) lowest among graduate/professional ($k = 3$). Because this is a Bayesian network meta-analysis, the number of outcomes refers to actual coded outcomes; the degree of precision of effect size stimates depends on the number of coded outcomes. Also, not all included studies had a control condition. Thus, the number of control outcomes does not equal the number of included studies.

Pre–post effect size estimates are highest among college- and graduate-level learners, at $\bar{g} = 1.16$ and $\bar{g} = 1.2$, respectively (see Figure 3). The 95% credible intervals represent ranges of true pre–post effects of scaffolding in each respective category. There were some true effects that were below zero for all educational levels except college. This is a function of the number of coded outcomes on which the Bayesian simulations estimated the posterior distribution.

Using a Bayesian network meta-analysis approach allows estimation of the true effect size and enables rank ordering treatments and calculating the probability that each treatment is the best. Scaffolding led to the highest pre–post gains at the

TABLE 1

Ranking and probability of the best of scaffolding used at different education levels

Education level	Ranking	Probability of the best
College	1.98	35%
Graduate	2.32	47%
Primary	3.07	11%
Middle	3.8	4%
Secondary	4.5	2%

college and graduate levels (see Table 1), ranked first and second with a 35% and a 47% chance of being the best, respectively.

Education Population

The evidence is strongest for the comparison of traditional students using scaffolding versus control (25 outcomes; see Supplementary Figure S3 in the online version of the journal). There are some studies that contained multiple educational populations. For example, traditional students using scaffolding co-occurred with underrepresented students, high-performing students, underperforming students, and ELL in at least one study for each combination.

The pre–post gains are consistently positive and substantial across educational populations (see Figure 4). The number of outcomes for scaffolding used by traditional students was the greatest, leading the group to have the tightest credible interval. Note that $N = 36$ for control in the education level network, while $N = 35$ for control here and for other moderators. This is because one study contained outcomes associated with two different educational levels—middle level and secondary; for the education level analysis, such outcomes could not be combined, while for the other moderator analyses, the outcomes needed to be combined. Scaffolding for students with learning disabilities had the largest effect size estimate ($\bar{g} = 3.13$) by a large margin. This effect size estimate should be considered tentative, as the MCMC sampling was based on four outcomes from a single study. ELL also had a large pre–post effect size ($\bar{g} = 0.92$).

When examining ranking and probability of the best, one finds scaffolding to have a high probability of having the best ranking when used among students with learning disabilities (see Table 2). Indeed, the probability of the best is virtually nil for all other education populations.

Research Question 2: To What Extent Does the Context in Which Scaffolding Is Used Moderate Cognitive Pre–Post Gains?

Context of Use

With the exception of problem solving, the number of coded outcomes for each problem-centered instructional model was very small (see Supplementary Figure S3 in the online version of the journal). This resulted in a very large range of true effects as calculated through Bayesian simulations.

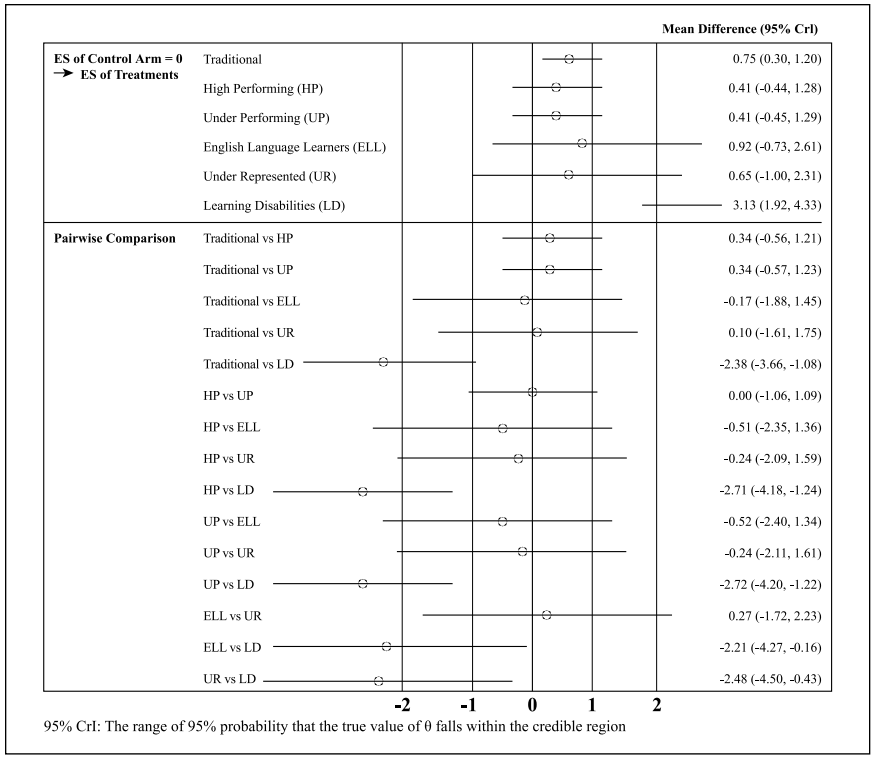

FIGURE 4. *Effect size (ES) estimates and 95% credible intervals (Crl) of scaffolding according to education population.*

TABLE 2

Ranking and probability of the best of scaffolding used among members of different education populations

Education population	Ranking	Probability of the best
Learning disabilities	1.03	96%
Traditional	3.47	0%
English language learners	3.51	2%
Underrepresented	4.16	1%
High-performing	4.75	0%
Underperforming	4.77	0%

The highest pre–post effect size was for project-based learning ($\bar{g}$ = 1.21; see Figure 5). Due to the low number of coded outcomes for the characteristics, the

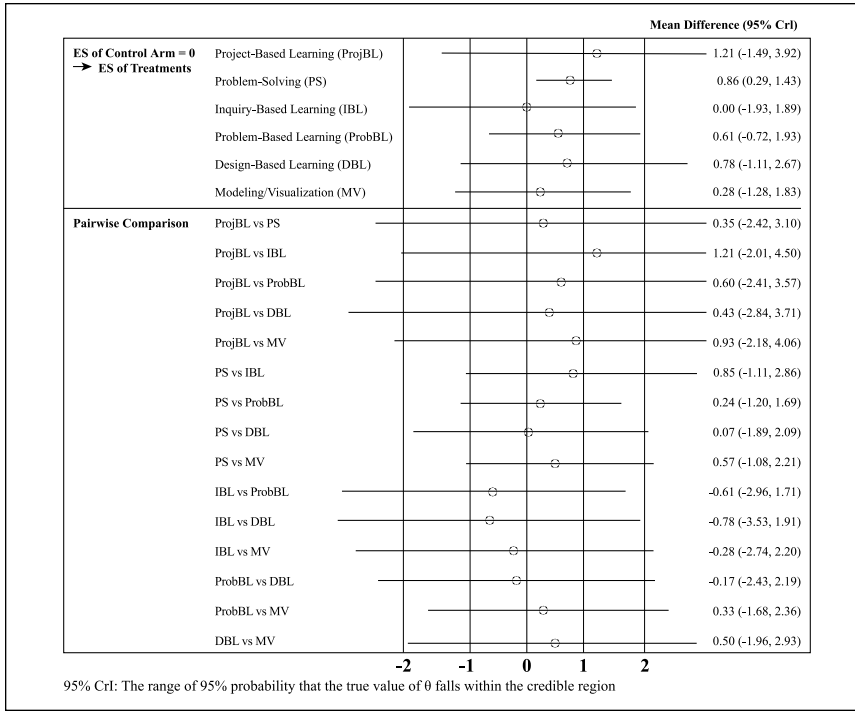

FIGURE 5. *Effect size (ES) estimates and 95% credible intervals (Crl) of scaffolding according to problem-centered instructional model with which scaffolding was used.*

range of true effects (credible interval) is wide. Thus, this ES needs to be interpreted cautiously. Most pre–post effect sizes were quite large, with the exception of inquiry-based learning ($\bar{g}$ = 0) and modeling/visualization ($\bar{g}$ = 0.28). This implies that scaffolding can lead to strong pre–post effect sizes across a wide range of problem-centered instructional approaches.

Proejct-based learning has the highest probability of the best (see Table 3). The ranking of problem solving is close behind that of project-based learning, but problem solving has a much lower likelihood of being the best.

STEM Discipline

Science and mathematics had the most coded outcomes, resulting in tighter credible intervals than in engineering and technology (see Supplementary Figure S3 in the online version of the journal).

Mathematics and technology had the highest pre–post effect sizes: $\bar{g}$ = 1.29 and $\bar{g}$ = 1.06, respectively (see Figure 6). Most studies coded as technology were from computer science instruction (n = 2), with the remaining outcome being from information technology. Mathematics and technology also had the highest and second highest probability of the best (see Table 4).

TABLE 3

Ranking and probability of the best of scaffolding used in the context of different problem-centered instructional models

Problem-centered instructional model	Ranking	Probability of the best
Project-based learning	2.81	44%
Problem solving	2.89	10%
Design-based learning	3.4	22%
Problem-based learning	3.7	11%
Modeling/visualization	4.55	7%
Inquiry-based learning	5.08	6%

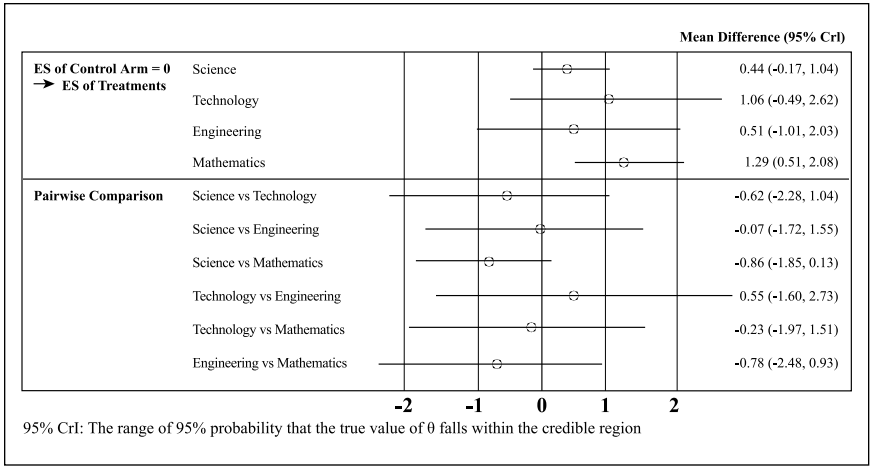

FIGURE 6. *Effect size (ES) estimates and 95% credible intervals (Crl) of scaffolding according to STEM discipline.*

Research Question 3: To What Extent Does Assessment Level Moderate Cognitive Pre–Post Gains Resulting From Scaffolding?

The network of evidence included a substantial number of direct comparisons among the assessment levels and between each assessment level and control, with the exception of between application and control (see Supplementary Figure S3 in the online version of the journal).

Scaffolding led to strong pre–post gains across assessment levels, with the lowest effect size estimate at the application level ($\bar{g} = 0.74$), and the highest at the concept level ($\bar{g} = 0.87$; see Figure 7). The credible intervals, which represent a range of true effect sizes, were relatively tight, pursuant to the large number of trials for each possible comparison, with the exception of application versus control. Accordingly, the credible interval for application was quite wide.

TABLE 4

Ranking and probability of the best of scaffolding used in the context of different STEM disciplines

STEM discipline	Ranking	Probability of the best
Mathematics	1.62	51%
Technology	2.23	35%
Engineering	3.23	12%
Science	3.33	1%

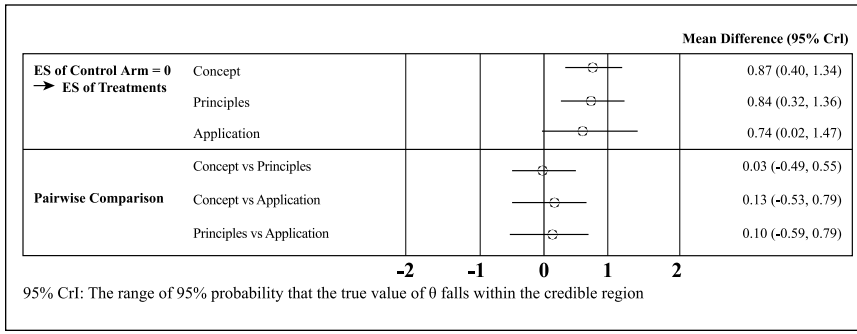

FIGURE 7. *Effect size (ES) estimates and 95% credible intervals (Crl) of scaffolding according to assessment level.*

TABLE 5

Ranking and probability of the best of scaffolding when measured at different assessment levels

Assessment level	Ranking	Probability of the best
Concept	1.79	41%
Application	1.93	34%
Principles	2.3	25%

The magnitude of difference among the assessment levels is minor. Comparing assessment levels through ranking similarly shows that there is little evidence to say that scaffolding is more effective at a particular assessment level than another (see Table 5).

Discussion

Implications for Instruction

It is often thought that when selecting an instructional approach, one needs to determine which level of educational outcome (e.g., concept level, problem solving) is most important, and select the approach that best aligns with the outcome

(Kuhn, 2007; Wiggins & McTighe, 2005). For example, many posit that using direct instruction is best at promoting strong conceptual knowledge (Kirschner, Sweller, & Clark, 2006). Still others argue that it is best to use problem-based learning to enhance problem-solving skill (Hmelo-Silver et al., 2007; Kuhn, 2007). In this way, teachers are often left in a quandary. Specifically, they often hear through professional learning and standards (e.g., Common Core and Next Generation Science Standards) that it is important to engage students in authentic problem solving (Drew, 2012; McLaughlin & Overturf, 2012). But they also know that if their students do not perform well on state standardized tests that emphasize declarative knowledge, their schools may be labeled low-performing, and other undesirable outcomes may ensue (Harman, Boden, Karpenski, & Muchowicz, 2016; Price, 2016). Thus, it is often difficult to convince K-12 teachers to integrate problem-centered learning (Keys & Bryan, 2001; Kim, Hannafin, & Bryan, 2007; Nariman & Chrispeels, 2015). Previous meta-analysis work implied that computer-based scaffolding leads to between-subjects differences that were statistically greater than zero and above $\bar{g} = 0.4$ across concept-, principles-, and application-level assessment (Belland et al., 2017). This is notable because such a diversity of strong effects is not found in problem-based learning by itself, which meta-analyses indicate leads to superior effects at the principles and application level, but equal or inferior effects at the concept level, compared with lecture (Gijbels, Dochy, Van den Bossche, & Segers, 2005; Walker & Leary, 2009). This article indicated that computer-based scaffolding leads to consistently strong pre–post effect sizes of at least 0.74 across concept-, principles-, and application-level assessment, with the strongest outcomes at the concept level. Thus, this article adds to the evidence (Belland et al., 2017) that scaffolding counteracts the purported weaknesses of problem-centered instructional models in helping learners achieve strong concept-level learning outcomes. In short, students using scaffolding when engaged in problem-centered instruction in STEM perform better than control students both in terms of between-group differences and within-group growth.

Turning to effect size estimates, it is important to remember that Bayesian network meta-analyses deal with a fundamentally different effect (within-subjects) than traditional meta-analyses (between-subjects). Thus, there is a need for extreme caution when comparing such. But there are metrics against which one can compare within-subjects effect sizes. One example is the average annual gain on standardized math scores, which ranges from $ES = 0.41$ to $ES = 1.14$ among elementary students, from $ES = 0.23$ to $ES = 0.26$ among middle school students, and from $ES = 0.06$ to $ES = 0.24$ among high school students (Hill, Bloom, Black, & Lipsey, 2008). Most scaffolding treatments coded in this study covered considerably less than one school year, with most closer to 1 or 2 weeks. However, the average within-subjects effects of computer-based scaffolding in this Bayesian network meta-analysis were at the high end of the range of annual gain scores in math achievement among high school students, at approximately the midpoint of the average annual gain scores among elementary students, and above the average annual gains among middle school students. Thus, in the span of 1 or 2 weeks, participants made cognitive gains akin to what students usually do in a whole academic year. Because no standardized annual exams apply across disciplines taken at the university or graduate level, a similar comparison of scaffolding's within-subject effects with average annual gains is not possible.

Variation in Scaffolding's Effect Among Learner Populations and Education Levels

Of note is that the within-subjects effect size was highest ($\bar{g}$ = 3.13) among elementary school students with learning disabilities. Caution is needed in interpretation, as the effect size was calculated from four effects from a single study. It is possible that such a large effect size is at least in part evidence of regression to the mean among the participants, who would likely have been performing at the low end of the scale before being exposed to the intervention. Indeed, six participants were excluded from the study "because their pre-assessment scores were above 60% correct" (Xin et al., 2017, p. 6). However, it is a very promising effect that warrants further research, as students with special needs constitute a group that is underrepresented in STEM (Israel, Maynard, & Williamson, 2013; National Center for Science and Engineering Statistics, National Science Foundation, 2013). Combatting underrepresentation in this group cannot likely be completely addressed solely through revision of instructional methods used in STEM education among the population, but boosting achievement in STEM among students with special needs on the order of over 3 *SD*s may lead such students to enroll in advanced STEM classes and degree programs (Kokkelenberg & Sinha, 2010; Riegle-Crumb & King, 2010), and have greater STEM self-efficacy (Britner & Pajares, 2006). It is also clear that waiting until later grades to address underrepresentation of students with special needs is unwise (Israel, Pearson, Tapia, Wherfel, & Reese, 2015). Rather, it is crucial to start early, which makes it especially promising that the included study (Xin et al., 2017) was set in third and fourth grades.

Turning to why scaffolding was so effective among students with special needs, one-to-one scaffolding has a long history in teaching students with learning disabilities (Palincsar, 1998; Stone, 1998). One way is through one-to-one support provided to mainstreamed students with special needs by teaching assistants (Radford, Bosanquet, Webster, & Blatchford, 2015). For example, teaching assistants may model and prompt the use of effective strategies (Radford et al., 2015). A key reason it has been advocated is its incorporation of dynamic assessment, which is considered of utmost importance in special education given the wide range of challenges and abilities that one can find among students with special needs (Tiekstra, Minnaert, & Hessels, 2016). While much computer-based scaffolding does not incorporate dynamic assessment, scaffolding embedded in intelligent tutoring systems does. The single coded study on scaffolding among special education students (Xin et al., 2017) was of an intelligent tutoring system in mathematics, which was used by elementary learners with a range of special needs, including learning disabilities, attention deficit hyperactivity disorder, and mild intellectual disabilities. In this way, vis-a-vis students with learning disabilities, the posterior distribution empirically approximated through MCMC sampling represents scaffolding embedded in intelligent tutoring systems. A meta-analysis showed that the effect size for dynamic assessment among students with special needs was highest (*ES* = 0.61) for students under the age of 10, while it was *ES* = 0.36, and *ES* = 0.38, among students aged 10 to 13 years and older than 13 years, respectively (Swanson & Lussier, 2001).

Problem-centered instruction has been implemented among students with varying cognitive and learning disabilities with success (Belland, Ertmer, & Simons, 2006; Belland, Glazewski, & Ertmer, 2009; Bottge, 2001; Bottge, Heinrichs, & Mehta, 2002). However, such efforts are not widespread, in part because direct instruction has long been considered to be highly efficacious among special education students (Datchuk, 2016; Gersten, 1985; White, 1988). Still, what is meant by direct instruction within special education differs from the model of an hour-long lecture. Rather, it refers to short, bite-sized instruction delivered in a rapid manner, with a goal of achieving mastery for all (Gersten, 1985; White, 1988). In this way, it is grounded in an idea of needing to maintain high expectations for students with special needs, which is also the rationale for using a scaffolding approach (Lutz, Guthrie, & Davis, 2006). A fundamental assumption in direct instruction is that it is best to minimize struggle/unsuccessful practice such that students learn as rapidly as possible. This is an assumption shared by developers of intelligent tutoring systems, many of which are based in the Adaptive Control of Thought–Rational (ACT-R) learning theory (Anderson, Matessa, & Lebiere, 1997). Thus, it is understandable that intelligent tutoring systems would be highly effective among students with special needs. At the same time, intelligent tutoring systems and direct instruction are not one and the same. Still, even for the staunchest of direct instruction advocates, a pre–post effect size of over 3 is hard to ignore. The magnitude of the effect can be determined with more clarity with the coding of more primary research, which would allow for a more accurate approximation of the population parameter through MCMC sampling and more robust indirect comparisons with scaffolding used among other education populations (Salanti, Higgins, Ades, & Ioannidis, 2008).

Interplay Between Highest Rankings Among College/Graduate-Level Learners and Elementary Learners With Special Needs

That the highest effect size estimates were among college- and graduate-level learners is similar to the finding of our traditional meta-analysis of computer-based scaffolding (Belland et al., 2017). It is no surprise that scaffolding is used in college and graduate-level populations, in that the promotion of skills like problem solving is critical at those levels (Jonassen, 2011). It is intriguing that an even greater effect can be found among third- and fourth-grade learners with special needs ($\bar{g} = 3.13$): It would be difficult to find learners who are further apart in cognitive abilities and development in the current data set. In short, computer-based scaffolding appears to be strongest in populations both furthest in age and cognitive development (college and graduate) from the target population (toddlers) of the original instructional scaffolding definition, and relatively close (third and fourth-grade students with learning disabilities; Wood et al., 1976). It is unlikely that one can find a complete explanation of why from the literature. And one cannot directly or indirectly compare elementary learners with special needs and college/graduate-level learners in the current meta-analysis because they were part of different networks of evidence. But one may think about this from the perspective of scaffolding's critical elements: dynamic assessment of student abilities, customization, and intersubjectivity (Belland, 2014). Three possible reasons that scaffolding fared so well among elementary students with learning

disabilities are as follows: (a) dynamic assessment is highly effective among the population (Swanson & Lussier, 2001); (b) the underlying design of the scaffolding the students used was informed by ACT-R, according to which it is best to minimize struggle, an assumption shared by direct instruction—a very successful strategy among the population; and (c) participants started out low on the pretest, so there was more room to grow. In college and graduate school, intersubjectivity may be more readily achieved than in K–12 settings because, in many cases, participants were majoring in the subject in which they were using scaffolding. For example, participants in Pfahl, Laitenberger, Ruhe, Dorsch, and Krivobokova (2004) were computer science graduate students solving problems related to software project management. Participants in Feyzi-Behnagh et al. (2014) were pathology/dermatology residents addressing dermatology problems. Such participants would be more likely to understand an appropriate solution in the class of problem being addressed (Mahardale & Lee, 2013; Mortimer & Wertsch, 2003) than would typical K-12 students engaging in a problem embedded in a discipline or profession. For example, among included studies, some problems addressed by middle school students related to thermoregulation (Roscoe, Segedy, Sulcer, Jeong, & Biswas, 2013) and finding homes for stranded aliens (Bulu & Pedersen, 2010), and some problems addressed by high school students related to chemical phase changes (Ardac & Sezen, 2002) and electric circuits (Korganci, Miron, Dafinei, & Antohe, 2014). The effect size of scaffolding used by elementary students ($\bar{g} = 0.74$) was closer to that of college ($\bar{g} = 1.16$) and graduate ($\bar{g} = 1.2$) than were the effect sizes of middle level ($\bar{g} = 0.48$) and high school ($\bar{g} = 0.22$). A possible reason is that out of seven coded studies at the elementary level, five were from mathematics; among STEM disciplines, mathematics had the largest effect size estimate ($\bar{g} = 1.29$).

Scaffolding and STEM Discipline

Effect size estimates were highest in mathematics and technology. This contrasts with our traditional meta-analysis of scaffolding, which found no significant difference in effect size estimate based on STEM discipline (Belland et al., 2017). That the effect size is highest in mathematics is not surprising, since much work on intelligent tutoring systems is done in mathematics (Steenbergen-Hu & Cooper, 2013; VanLehn, 2011) and it has long benefitted from more synthesis of research results and systematic refinement (Murray, 1999; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011) than other scaffolding types. Of note, many intelligent tutoring systems are grounded in ACT-R, according to which the goal of instruction is to present knowledge to students and give them practice applying such knowledge to problems, such that production rules for applying the knowledge are generated (Anderson et al., 1997). Such an approach fits well with the traditional approach to mathematics curricula in the United States, where even textbooks supposedly aligned with the Common Core focus largely on procedures and declarative knowledge (Polikoff, 2015). For example, in traditional algebra curricula, the focus is on helping students solve for variables, rather than framing variables as tools to characterize relationships (Nie, Cai, & Moyer, 2009). While traditional approaches to mathematics instruction are not the same as those of ACT-R-informed intelligent tutoring systems, production rules are similar to procedures, and so the foundations of the two approaches are in alignment.

Most studies in the technology category were from computer science. Scaffolding's strength in producing within-subjects gains is likely to be of great interest to those involved in the computer science for all initiative (K-12 Computer Science Framework Steering Committee, 2016; Obama, 2016). At the same time, it is not clear why pre–post effect sizes of scaffolding would be higher in computer science than in engineering and science. Further research is needed.

Implications for Meta-Analysis

Traditional meta-analysis has a long history in education research (Glass, 1976), where it has allowed researchers to take a step back from the body of research studies on a topic and read a (relatively) unbiased account of what the literature says. But there is bias in any literature review, meta-analysis or otherwise, which stems from factors like publication bias, how researchers frame the literature, inclusion criteria, choice of moderators, and unequal sample sizes. Following a Bayesian network meta-analysis approach does not mean that one avoids bias. Rather, it mitigates some biases but also introduces new biases. For example, its inclusion criteria allows one to synthesize results of early stage research for which the samples are too small to warrant a control group or the constructs are not narrowed down enough to allow for fine-tuned control of variables (Courgeau, 2012; Sutton & Abrams, 2001). Much scaffolding research is done in real-world settings, and does not benefit from a finely controlled study design. Such research is useful, but would be missed in a traditional meta-analysis. Most studies (70%) included in this Bayesian network meta-analysis were not covered in our traditional meta-analysis of computer-based scaffolding in STEM education (Belland et al., 2017).

Changing the prior distribution can lead to big changes in the posterior distribution, which is the source of much contention between frequentists and Bayesians (Efron, 2013; Little, 2006). Fit statistics (e.g., deviance information criterion) can provide evidence that the most suitable prior distribution was selected. But this does not sweep away the contention that arises from prior distributions. Still, using a noninformative prior distribution for which fit statistics are best may reduce bias in that the coding sample drives the approximation of the posterior distribution more than does the prior distribution (Jansen et al., 2008).

Bayesian network meta-analysis gained traction in pharmaceutical research in large part because it enhances decision-making by ranking all available treatments and determining the probability that each is the best (Jansen et al., 2011; Salanti, 2012). In this way, one could see with relative confidence which medication to treat condition *X* is the most effective. In a similar manner, educators and policy makers turn to meta-analyses to determine which instructional strategies are most effective and should be integrated into teaching, funded, or further researched. By ranking all available treatments and determining the probability that each treatment is the best, Bayesian network meta-analysis can help educators and policy makers determine which instructional strategy is most worthy of the classroom or funding. It is important to note that the accuracy of ranking and probability of the best depends on the number of studies representing each moderator level (Mills et al., 2013). It is likely that as Bayesian network meta-analysis is used more widely, stronger strategies for ascertaining accuracy of rankings and probabilities of the best will emerge (Casella & Moreno, 2006). At the same time, for Bayesian

network meta-analysis to maximally benefit educational researchers, there is a need for more multiple comparison studies. The reason for this is that educational researchers who do quantitative research have long held randomized controlled studies as the gold standard (Sullivan, 2011; U.S. Department of Education, Institute of Education Sciences, & National Center for Education Evaluation and Regional Assistance, 2006). Randomized controlled studies can establish the value of an intervention compared with lecture, but when it is clear that an intervention is better than control, as is the case with scaffolding (Belland et al., 2017, 2015; Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011), it makes sense to determine which versions work better under which circumstances. This can be done with multiple treatment studies and, by extension, Bayesian network meta-analyses. By including more multiple treatment studies, treatment networks would be more symmetrical and credible intervals of direct and indirect comparisons would be tighter (Salanti, Giovane, Chaimani, Caldwell, & Higgins, 2014). When a large proportion of studies included in a Bayesian network meta-analysis involve control conditions, this results in a radiating star network, in which the comparisons among various treatments individually with control are the most informative in that they reference the greatest amount of direct evidence, and comparisons among treatment types are least informative in that they reference the least amount of direct evidence (Salanti et al., 2008; Salanti et al., 2014).

Having more symmetrical treatment networks filled with more multiple treatment studies may also help address the issue of nesting of participants within classrooms, within schools, and within school districts, which often arises in education research. If the needed data are in the included research reports, one can use a hierarchical approach to meta-analysis, which accounts for nesting. This was done in a between-subjects meta-analysis of problem-based learning in medical education (Kalaian, Mullan, & Kasim, 1999), allowing the authors to find that students in medical schools with more experience with problem-based learning had higher medical content knowledge than students in medical schools that have less experience with problem-based learning. Hierarchical approaches to Bayesian network meta-analysis have been used in medical research (Stettler et al., 2007). While using a hierarchical Bayesian network meta-analysis could be advantageous in education research, few studies that we coded contained the needed classroom-level, school-level, and district-level data. Furthermore, one needs to have sufficient degrees of freedom across analyses to meaningfully detect intraclass correlation. To use hierarchical Bayesian network meta-analysis in education research, educational researchers need to include classroom-, school-, and district-level data on variables such as teacher experience, SES, and state standardized test scores. In this way, data would be available for future researchers to conduct hierarchical Bayesian network meta-analysis, which in turn may lead to new and more valid conclusions.

That more studies adopt within-subject designs is important not only for Bayesian network meta-analysis but also for social justice: compared with between-subject designs, within-subject designs may indicate better the extent to which members of marginalized populations (e.g., students from minority and low-SES backgrounds) benefit from scaffolding (McNeish & Dumas, 2017). Such students often score low on a single time point assessment, but this does not

illustrate their full capacity for learning (McNeish & Dumas, 2017). Scores taken at multiple time points can highlight areas of strength and growth of marginalized students, especially when one compares the trajectory and magnitude of growth among different student populations (McNeish & Dumas, 2017). Systematic synthesis of within-subject effects allows one to see which interventions hold promise for which populations, which in turn has the potential to enhance social justice. For example, despite the inclusion of 144 studies in our traditional meta-analysis (Belland et al., 2017), effect size estimates among education populations (high-performing, low-income, traditional, underperforming, and underrepresented) were indistinguishable statistically, except that the effect of scaffolding was greater among traditional students than among underperforming students. In contrast, the current study indicated that scaffolding has the greatest promise among special education and ELL students. Scaffolding also produced strong pre–post gains among underperforming and underrepresented students. Knowing that scaffolding is helpful across a wide range of educational populations is important, but it is equally important to understand the within-subjects growth that one might expect among members of different populations. Thus, we urge scaffolding researchers to adopt within-subject designs, especially when studying marginalized populations.

This article introduces an approach (Bayesian network meta-analysis) with which educational researchers can synthesize within-subject effects. When the goal is to model growth due to an intervention, it will accomplish synthesis goals more effectively than traditional meta-analysis. When the goal is to model a comparison between an intervention and control, traditional meta-analysis or Bayesian traditional meta-analysis would fit best.

Limitations and Suggestions for Future Research

A Bayesian approach to network meta-analysis was adopted because in traditional network meta-analysis, each included study needs to incorporate a control condition, which would have required us to exclude many multiple treatment studies. Through MCMC simulations informed by the prior distribution and current coding, we could strengthen predictions for the effect size for each context of scaffolding use. However, by taking this approach, the effect size estimates are empirical approximations of population parameters, which depend on use of the best possible prior distribution. Any change in prior distribution could produce different results. We verified the appropriateness of our prior distribution through deviance information criterion statistics. Furthermore, the strictness of the inclusion criteria and the fact that the majority of included studies were not included in our traditional meta-analysis (Belland et al., 2017) could mean that the nature of included scaffolding interventions was strikingly different. This may not be the case since the same operational definition of scaffolding was applied in both meta-analyses. In a follow-up to this study, we plan to (a) use the results of this article as an informative prior distribution and (b) code new studies not included in this meta-analysis. This may result in tighter credible intervals and more accurate effect size estimates.

No meta-analysis covers qualitative results, and all meta-analyses exclude some quantitative research. Conclusions of any meta-analysis are limited in these ways. For example, based on this study and our previous meta-analysis (Belland et al., 2017), college- and graduate-level education appear to be the most

promising contexts for scaffolding. It is possible that when synthesizing all empirical research (including quantitative research that did not meet the inclusion criteria and qualitative research) on the topic, one would reach a different conclusion.

Authors of the included studies chose the content covered in and when to administer the pre- and posttest. Choosing alternative test content or test administration time points could have led to different pre–post effect sizes. As such, the pre–post effect sizes reported in this article are an imperfect measure of the cognitive growth resulting from computer-based scaffolding.

The number of outcomes at some levels of coding categories were small. This could have led to large fluctuation of simulated effects, which in turn could have led to wide credible intervals. But when we checked the trace plot, there was no large fluctuation. Another possible reason is real inconsistency in computer-based scaffolding findings. Further research is needed.

Using the Sugrue (1995) framework for coding of assessment level may have led us to not fully capture the range of outcomes that are targeted by scaffolding, including conceptual change, particularly when helping students overcome misconceptions. Modifying the Sugrue (1995) framework may help more fully reflect outcome types targeted with scaffolding.

Finally, it is possible that our search terms did not uncover all relevant studies because some interventions may share essential characteristics with scaffolding but their name does not contain any of our search terms. We asked advisory board members (representing biology, chemistry, physics, engineering, technology, mathematics, cognitive science, learning sciences, and meta-analysis) for input on search terms. Authors of future Bayesian network meta-analyses of scaffolding research would be wise to carefully consider search terms.

Conclusion

Computer-based scaffolding is highly effective at improving cognitive learning from pre to posttest; this strength is largely consistent across measurement levels, education populations, and STEM disciplines. Scaffolding led to a pre–post gain of at least 1 SD among university-level students, graduate-level students and students with learning disabilities, and when used in the context of (a) project-based learning, (b) technology, and (c) mathematics. These are quite large effect sizes, which indicates that scaffolding's effect is strong in the contexts and warrants further exploration. Furthermore, effect size estimates were at least 0.74 across concept-, principles-, and application-level assessment. Scaffolding's consistent effect informs teachers that using problem-centered approaches does not preclude strong concept learning, which is often the focus of state standardized tests and, by consequence, teacher evaluation (Harman et al., 2016; Price, 2016). The within-subjects effect size at the concept level was $\bar{g} = 0.87$, a pre–post effect size with which any principal would be pleased, especially if it resulted from a 1- to 2-week unit.

The notably large effect size ($\bar{g} = 3.13$) among special education populations is rarely seen in education research. Further research is needed to see if the effect size estimate remains consistent with a larger number of coded studies, but doing

so is critical to enhancing social justice and the STEM workforce (Carnevale & Desrochers, 2003; Israel et al., 2013).

Also intriguing was that scaffolding's effect was strongest among college, graduate, and early elementary special education learners—the farthest and closest to the population for which scaffolding was originally proposed (Wood et al., 1976). Possible explanations include that dynamic assessment is a known strong intervention for special education students (Swanson & Deshler, 2003; Swanson & Lussier, 2001) and college and graduate students potentially exhibit greater intersubjectivity when they address problems related to their major.

Scaffolding showed its largest within-subject effects in contexts (i.e., college and graduate) far removed from its origins in early childhood education (Wood et al., 1976), which is consistent with our earlier traditional meta-analysis (Belland et al., 2017). Also, scaffolding has strong effects among special education students, ELLs, and students who are otherwise underrepresented, and when used with diverse problem-centered instructional models. This implies that scaffolding is a robust and versatile model.

This article also introduces Bayesian network meta-analysis to education research (Bhatnagar et al., 2014; Jansen et al., 2011; Salanti, 2012). But for Bayesian network meta-analysis to be of maximum utility in education research, there is a need for more multiple treatment studies to enhance researchers' ability to (a) strengthen comparisons (Salanti et al., 2014) and (b) use a hierarchical approach to Bayesian network meta-analysis so as to address nesting (Stettler et al., 2007). This may also help researchers get a better sense of the extent to which a treatment helped members of marginalized populations (McNeish & Dumas, 2017).

Note

This research was supported by the National Science Foundation under REESE Grant No. 1251782. Any opinions, findings, or conclusions are those of the authors and do not necessarily represent official positions of the National Science Foundation.

References

Achieve. (2013). *Next generation science standards*. Retrieved from http://www.nextgenscience.org/get-to-know

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*, 439–462. doi:10.1207/s15327051hci1204_5

Ardac, D., & Sezen, A. H. (2002). Effectiveness of computer-based chemistry instruction in enhancing the learning of content and variable control under guided versus unguided conditions. *Journal of Science Education and Technology*, *11*, 39–48. doi:10.1023/A:1013995314094

Barak, M., & Dori, Y. J. (2005). Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. *Science Education*, *89*(1), 117–139. doi:10.1002/sce.20027

Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., & Bransford, J. D. (1998). Doing with understanding: Lessons from research on problem- and project-based learning. *Journal of the Learning Sciences*, *7*, 271–311. doi:10.1080/10508406.1998.9672056

Barrows, H. S., & Tamblyn, R. M. (1980). *Problem-based learning: An approach to medical education*. New York, NY: Springer.

Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 505–518). New York, NY: Springer.

Belland, B. R. (2017). *Instructional scaffolding in STEM education: Strategies and efficacy evidence*. Cham, Switzerland: Springer. Retrieved from http://doi.org/10.1007/978-3-319-02565-0

Belland, B. R., Ertmer, P. A., & Simons, K. D. (2006). Perceptions of the value of problem-based learning among students with special needs and their teachers. *Interdisciplinary Journal of Problem-Based Learning*, *1*(2), 1–18. doi:10.7771/1541-5015.1024

Belland, B. R., Glazewski, K. D., & Ertmer, P. A. (2009). Inclusion and problem-based learning: Roles of students in a mixed-ability group. *Research on Middle Level Education*, *32*(9), 1–19.

Belland, B. R., Kim, C., & Hannafin, M. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist*, *48*, 243–270. doi: 10.1080/00461520.2013.838920

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, *87*, 309–344. doi:10.3102/00346 54316670999

Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Educational Technology & Society*, *18*, 183–197.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. New York, NY: Springer.

Bhatnagar, N., Lakshmi, P. V. M., & Jeyashree, K. (2014). Multiple treatment and indirect treatment comparisons: An overview of network meta-analysis. *Perspectives in Clinical Research*, *5*, 154–158. doi:10.4103/2229-3485.140550

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, England: Wiley.

Bottge, B. A. (2001). Building ramps and hovercrafts-and improving math skills. *Teaching Exceptional Children*, *34*, 16–23. doi:10.1177/004005990103400102

Bottge, B. A., Heinrichs, M., & Mehta, Z. D. (2002). Weighing the benefits of anchored math instruction for students with disabilities in general education classes. *Journal of Special Education*, *35*, 186–200. doi:10.1177/002246690203500401

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*, 485–499. doi:10.1002/tea.20131

Bulu, S. T., & Pedersen, S. (2010). Scaffolding middle school students' content knowledge and ill-structured problem solving in a problem-based hypermedia learning environment. *Educational Technology Research & Development*, *58*, 507–529. doi:10.1007/s11423-010-9150-9

Carlin, B. P., & Chib, S. (1995). Bayesian model choice via Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*, 473–484.

Carnevale, A. P., & Desrochers, D. M. (2003). Preparing students for the knowledge economy: What school counselors need to know. *Professional School Counseling*, *6*, 228–236.

Casella, G., & Moreno, E. (2006). Objective Bayesian variable selection. *Journal of the American Statistical Association*, *101*, 157–167. doi:10.1198/016214505000000646

Chandrasekaran, S., Stojcevski, A., Littlefair, G., & Joordens, M. (2013). Project-oriented design-based learning: Aligning students' views with industry needs. *International Journal of Engineering Education*, *29*, 1109–1118.

Chen, Y. S., Kao, T. C., & Sheu, J. P. (2003). A mobile learning system for scaffolding bird watching learning. *Journal of Computer Assisted Learning*, *19*, 347–359. doi:10.1046/j.0266-4909.2003.00036.x

Chen, Y.-S., Kao, T.-C., & Sheu, J.-P. (2005). Realizing outdoor independent learning with a butterfly-watching mobile learning system. *Journal of Educational Computing Research*, *33*, 395–417. doi:10.2190/0PAB-HRN9-PJ9K-DY0C

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.

Courgeau, D. (2012). *Probability and social science: Methodological relationships between the two approaches*. Dordrecht, Germany: Springer.

Crippen, K. J., & Archambault, L. (2012). Scaffolded inquiry-based instruction with technology: A signature pedagogy for STEM education. *Computers in the Schools*, *29*, 157–173. doi:10.1080/07380569.2012.658733

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, *30*, 433–464. doi:10.1023/A:1020516301541

Datchuk, S. M. (2016). A direct instruction and precision teaching intervention to improve the sentence construction of middle school students with writing difficulties. *Journal of Special Education*. Advance online publication. doi:10.1177/0022466916665588

Davis, E., & Linn, M. (2000). Scaffolding students' knowledge integration: Prompts for reflection in KIE. *International Journal of Science Education*, *22*, 819–837. doi:10.1080/095006900412293

de Jong, T., Härtel, H., Swaak, J., & van Joolingen, W. (1996). Support for simulation-based learning: The effect of assignments in learning about transmission lines. In A. D. de Illarraza Sánchez & I. F. de Castro (Eds.), *Computer aided learning and instruction in science and engineering* (pp. 9–26). Berlin, Germany: Springer.

Devolder, A., Van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: Systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning*, *28*, 557–573. doi:10.1111/j.1365-2729.2011.00476.x

Dimino, J., Gersten, R., Carnine, D., & Blake, G. (1990). Story grammar: An approach for promoting at-risk secondary students' comprehension of literature. *Elementary School Journal*, *91*, 19–32. doi:10.1086/461635

Drew, S. V. (2012). Open up the ceiling on the common core state standards: Preparing students for 21st-century literacy—Now. *Journal of Adolescent & Adult Literacy*, *56*, 321–330. doi:10.1002/JAAL.00145

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, *8*, 391–450. doi:10.1080/10508406.1999.9672075

Efron, B. (2013). Bayes' theorem in the 21st century. *Science*, *340*, 1177–1178. doi:10.1126/science.1236536

Englert, C. S., Raphael, T. E., Anthony, L. M. A. H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal*, *28*, 337–372. doi:10.3102/00028312028002337

Feyzi-Behnagh, R., Azevedo, R., Legowski, E., Reitmeyer, K., Tseytlin, E., & Crowley, R. S. (2014). Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instructional Science*, *42*, 159–181. doi:10.1007/s11251-013-9275-4

Galand, B., Frenay, M., & Raucent, B. (2012). Effectiveness of problem-based learning in engineering education: A comparative study on three levels of knowledge structure. *International Journal of Engineering Education*, *28*, 939–947.

Galleto, P. G., & Refugio, C. N. (2012). *Students' skills in mathematical computation using graphing calculator*. Paper presented at the proceedings of the 17th Asian Technology Conference in Mathematics, Thailand. Retrieved from http://atcm.mathandtech.org/ep2012/regular_papers/3472012_19901.pdf

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gersten, R. (1985). Direct instruction with special education students: A review of evaluation research. *Journal of Special Education*, *19*, 41–58. doi:10.1177/002246698501900104

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, *75*, 27–61. doi:10.3102/00346543075001027

Gijlers, H. (2005). *Confrontation and co-construction: Exploring and supporting collaborative scientific discovery learning with computer simulations* (PhD thesis). University of Twente, Enschede, Netherlands.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3–8. doi:10.3102/0013189X005010003

Goring, S. M., Gustafson, P., Liu, Y., Saab, S., Cline, S. K., & Platt, R. W. (2016). Disconnected by design: Analytic approach in treatment networks having no common comparator. *Research Synthesis Methods*, *7*, 420–432. doi:10.1002/jrsm.1204

Hadwin, A. F., Wozney, L., & Pontin, O. (2005). Scaffolding the appropriation of self-regulatory activity: A socio-cultural analysis of changes in teacher–student discourse about a graduate research portfolio. *Instructional Science*, *33*, 413–450. doi:10.1007/s11251-005-1274-7

Hannafin, M., Land, S., & Oliver, K. (1999). Open-ended learning environments: Foundations, methods, and models. In C. M. Reigeluth (Ed.), *Instructional design theories and models: Volume II: A new paradigm of instructional theory* (pp. 115–140). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

Harman, W. G., Boden, C., Karpenski, J., & Muchowicz, N. (2016). No Child Left Behind: A postmortem for Illinois. *Education Policy Analysis Archives*, *24*(48), 1–24. doi:10.14507/epaa.v24.2186

Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal of Research in Science Teaching*, *24*, 291–307. doi:10.1002/tea.3660240404

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*, 77–89. doi:10.1080/19312450709336664

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*, 490–499. doi:10.1037/0033-2909.92.2.490

Heinrich, S., Knight, V., Collins, B. C., & Spriggs, A. D. (2016). Embedded simultaneous prompting procedure to teach STEM content to high school students with moderate disabilities in an inclusive setting. *Education and Training in Autism and Developmental Disabilities*, *51*, 41–54.

Helle, L., Tynjälä, P., & Olkinuora, E. (2006). Project-based learning in post-secondary education: Theory, practice and rubber sling shots. *Higher Education*, *51*, 287–314. doi:10.1007/s10734-004-6386-5

Hernandez, P. R., Schultz, P. W., Estrada, M., Woodcock, A., & Chance, R. C. (2013). Sustaining optimal motivation: A longitudinal analysis of interventions to broaden participation of underrepresented students in STEM. *Journal of Educational Psychology*, *105*, 89–107. doi:10.1037/a0029691

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., . . . Sterne, J. A. C. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, *343*, d5928. doi:10.1136/bmj.d5928

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*, 172–177. doi:10.1111/j.1750-8606.2008.00061.x

Hmelo-Silver, C. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*, 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C., Duncan, R., & Chinn, C. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, *42*, 99–107. doi:10.1080/00461520701263368

Israel, M., Maynard, K., & Williamson, P. (2013). Promoting literacy-embedded, authentic stem instruction for students with disabilities and other struggling learners. *Teaching Exceptional Children*, *45*, 18–25. doi:10.1177/004005991304500402

Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., & Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Computers & Education*, *82*, 263–279. doi:10.1016/j.compedu.2014.11.022

Jansen, J. P., Crawford, B., Bergman, G., & Stam, W. (2008). Bayesian meta-analysis of multiple treatment comparisons: An introduction to mixed treatment comparisons. *Value in Health*, *11*, 956–964. doi:10.1111/j.1524-4733.2008.00347.x

Jansen, J. P., Fleurence, R., Devine, B., Itzler, R., Barrett, A., Hawkins, N., . . .Cappelleri, J. C. (2011). Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: Report of the ISPOR task force on indirect treatment comparisons good research practices: Part 1. *Value in Health*, *14*, 417–428. doi:10.1016/j.jval.2011.04.002

Jeong, A., & Joung, S. (2007). Scaffolding collaborative argumentation in asynchronous discussions with message constraints and message labels. *Computers & Education*, *48*, 427–445. doi:10.1016/j.compedu.2005.02.002

Jonassen, D. (2003). Using cognitive tools to represent problems. *Journal of Research on Technology in Education*, *35*, 362–381. doi:10.1080/15391523.2003.10782391

Jonassen, D. (2011). *Learning to solve problems: A handbook for designing problem-solving learning environments*. New York, NY: Routledge.

K-12 Computer Science Framework Steering Committee. (2016). *K–12 computer science framework*. Retrieved from http://www.k12cs.org

Kalaian, H. A., Mullan, P. B., & Kasim, R. M. (1999). What can studies of problem-based learning tell us? Synthesizing and modeling PBL effects on National Board of Medical Examination performance: Hierarchical linear modeling meta-analytic approach. *Advances in Health Sciences Education*, *4*, 209–221. doi:10.1023/A:1009871001258

Kay, M., Nelson, G. L., & Hekler, E. B. (2016). Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 4521–4532). San Jose, CA: Association for Computing Machinery.

Keys, C. W., & Bryan, L. A. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching*, *38*, 631–645. doi:10.1002/tea.1023

Kim, M. C., Hannafin, M. J., & Bryan, L. A. (2007). Technology-enhanced inquiry tools in science education: An emerging pedagogical framework for classroom practice. *Science Education*, *91*, 1010–1030. doi:10.1002/sce.20219

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. doi:10.1207/s15326985ep4102_1

Koedinger, K., & Corbett, A. (2006). Cognitive tutors: Technology bringing learning sciences to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). Cambridge, England: Cambridge University Press.

Kokkelenberg, E. C., & Sinha, E. (2010). Who succeeds in STEM studies? An analysis of Binghamton University undergraduate students. *Economics of Education Review*, *29*, 935–946. doi:10.1016/j.econedurev.2010.06.016

Kolodner, J., Camp, P., Crismond, D., Fasse, B., Gray, J., Holbrook, J., . . .Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting Learning by Design™ into practice. *Journal of the Learning Sciences*, *12*, 495–547. doi:10.1207/S15327809JLS1204_2

Korganci, N., Miron, C., Dafinei, A., & Antohe, S. (2014). Comparison of generating concept maps and using concept maps on students' achievement. *eLearning and Software for Education*, *2*, 287–293. doi:10.12753/2066-026X-14-098

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, *7*, 313–350. doi:10.1080/10508406.1998.9672057

Krajcik, J., Codere, S., Dahsah, C., Bayer, R., & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education*, *25*, 157–175. doi:10.1007/s10972-014-9383-2

Kramarski, B., & Zeichner, O. (2001). Using technology to enhance mathematical reasoning: Effects of feedback and self-regulation learning. *Educational Media International*, *38*, 77–82. doi:10.1080/09523980110041458

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*, 411–433. doi:10.1111/j.1468-2958.2004.tb00738.x

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*, 109–113. doi:10.1080/00461520701263376

Lin, T.-C., Hsu, Y.-S., Lin, S.-S., Changlai, M.-L., Yang, K.-Y., & Lai, T.-L. (2012). A review of empirical evidence on scaffolding for science education. *International Journal of Science and Mathematics Education*, *10*, 437–455. doi:10.1007/s10763-011-9322-z

Little, R. J. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *American Statistician*, *60*, 213–223. doi:10.1198/000313006X117837

Liu, M. (2004). Examining the performance and attitudes of sixth graders during their use of a problem-based hypermedia learning environment. *Computers in Human Behavior*, *20*, 357–379. doi:10.1016/S0747-5632(03)00052-9

Lumley, T. (2002). Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*, *21*, 2313–2324. doi:10.1002/sim.1201

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337. doi:10.1023/A:1008929526011

Lutz, S. L., Guthrie, J. T., & Davis, M. H. (2006). Scaffolding for engagement in elementary school reading instruction. *Journal of Educational Research*, *100*, 3–20. doi:10.3200/JOER.100.1.3-20

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*, 901–918. doi:10.1037/a0037123

Magana, A. J. (2014). Learning strategies and multimedia techniques for scaffolding size and scale cognition. *Computers & Education*, *72*, 367–377. doi:10.1016/j.compedu.2013.11.012

Mahardale, J. W., & Lee, C. B. (2013). Understanding how social and epistemic scripts perpetuate intersubjectivity through patterns of interactions. *Interactive Learning Environments*, *21*, 68–88. doi:10.1080/10494820.2010.547204

Marx, R., Blumenfeld, P., Krajcik, J., Fishman, B., Soloway, E., Geier, R., & Tal, R. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching*, *41*, 1063–1080. doi:10.1002/tea.20039

McLaughlin, M., & Overturf, B. J. (2012). The Common Core: Insights into the K–5 standards. *The Reading Teacher*, *66*, 153–164. doi:10.1002/TRTR.01115

McNeish, D., & Dumas, D. (2017). Nonlinear growth models as measurement models: A second-order growth curve model for measuring potential. *Multivariate Behavioral Research*, *52*, 61–85. doi:10.1080/00273171.2016.1253451

Mengersen, K. L., Drovandi, C. C., Robert, C. P., Pyne, D. B., & Gore, C. J. (2016). Bayesian estimation of small effects in exercise and sports science. *PLoS One*, *11*(4), e0147311. doi:10.1371/journal.pone.0147311

Mills, E., Thorlund, K., & Ioannidis, J. P. A. (2013). Demystifying trial networks and network meta-analysis. *British Medical Journal*, *346*, f2914. doi:10.1136/bmj.f2914

Molina, R., Borror, J., & Desir, C. (2016). Supporting STEM success with elementary students of color in a low-income community. *Distance Learning*, *13*(2), 19–25.

Mortimer, E. F., & Wertsch, J. V. (2003). The architecture and dynamics of intersubjectivity in science classrooms. *Mind, Culture, and Activity*, *10*, 230–244. doi:10.1207/s15327884mca1003_5

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*, 98–129.

Nariman, N., & Chrispeels, J. (2015). PBL in the era of reform standards: Challenges and benefits perceived by teachers in one elementary school. *Interdisciplinary Journal of Problem-Based Learning*, *10*(1). doi:10.7771/1541-5015.1521

National Center for Science and Engineering Statistics, National Science Foundation. (2013). *Women, minorities, and persons with disabilities in science and engineering: 2013*. Retrieved from http://www.nsf.gov/statistics/wmpd/2013/pdf/nsf13304_digest.pdf

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards*. Retrieved from http://www.corestandards.org/the-standards

Nie, B., Cai, J., & Moyer, J. C. (2009). How a standards-based mathematics curriculum differs from a traditional curriculum: With a focus on intended treatments of the ideas of variable. *ZDM*, *41*, 777–792. doi:10.1007/s11858-009-0197-1

Obama, B. (2016). *2016 State of the Union*. Retrieved from https://www.whitehouse.gov/sotu

Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh—A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities." *Journal of Learning Disabilities*, *31*, 370–373. doi:10.1177/002221949803100406

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175. doi:10.1207/s1532690xci0102_1

Pfahl, D., Laitenberger, O., Ruhe, G., Dorsch, J., & Krivobokova, T. (2004). Evaluating the learning effectiveness of using simulations in software project management education: Results from a twice replicated experiment. *Information and Software Technology*, *46*, 127–147. doi:10.1016/S0950-5849(03)00115-0

Pifarre, M., & Cobos, R. (2010). Promoting metacognitive skills through peer scaffolding in a CSCL environment. *International Journal of Computer-Supported Collaborative Learning*, *5*, 237–253. doi:10.1007/s11412-010-9084-6

Polikoff, M. S. (2015). How well aligned are textbooks to the common core standards in mathematics? *American Educational Research Journal*, *52*, 1185–1211. doi:10.3102/0002831215584435

Price, H. E. (2016). Assessing U.S. public school quality: The advantages of combining internal "consumer ratings" with external NCLB ratings. *Educational Policy*, *30*, 403–433. doi:10.1177/0895904814551273

Puhan, M. A., Schünemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., . . . Guyatt, G. H. (2014). A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *British Medical Journal*, *349*, g5630. doi:10.1136/bmj.g5630

Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, *40*, 1–12. doi:10.1207/s15326985ep4001_1

Puntambekar, S., & Kolodner, J. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, *42*, 185–217. doi:10.1002/tea.20048

Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human–Computer Interaction*, *18*, 395–428. doi:10.1207/S15327051HCI1804_3

Quintana, C., Reiser, B., Davis, E., Krajcik, J., Fretz, E., Duncan, R., . . .Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, *13*, 337–386. doi:10.1207/s15327809jls1303_4

Radford, J., Bosanquet, P., Webster, R., & Blatchford, P. (2015). Scaffolding learning for independence: Clarifying teacher and teaching assistant roles for children with special educational needs. *Learning and Instruction*, *36*, 1–10. doi:10.1016/j.learninstruc.2014.10.005

Raudenbush, S. (2009). Analyzing effect sizes: Random effect models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295–315). New York, NY: Russell Sage Foundation.

Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences*, *13*, 273–304. doi:10.1207/s15327809jls1303_2

Riegle-Crumb, C., & King, B. (2010). Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, *39*, 656–664. doi:10.3102/0013189X10391657

Rienties, B., Giesbers, B., Tempelaar, D., Lygo-Baker, S., Segers, M., & Gijselaers, W. (2012). The role of scaffolding and motivation in CSCL. *Computers & Education*, *59*, 893–906. doi:10.1016/j.compedu.2012.04.010

Roscoe, R. D., Segedy, J. R., Sulcer, B., Jeong, H., & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education*, *62*, 286–297. doi:10.1016/j.compedu.2012.11.008

Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods*, *3*, 80–97. doi:10.1002/jrsm.1037

Salanti, G., Giovane, C. D., Chaimani, A., Caldwell, D. M., & Higgins, J. P. T. (2014). Evaluating the quality of evidence from a network meta-analysis. *PLoS One*, *9*, e99682. doi:10.1371/journal.pone.0099682

Salanti, G., Higgins, J. P. T., Ades, A., & Ioannidis, J. P. (2008). Evaluation of networks of randomized trials. *Statistical Methods in Medical Research*, *17*, 279–301. doi:10.1177/0962280207080643

Silk, E., Schunn, C., & Cary, M. (2009). The impact of an engineering design curriculum on science reasoning in an urban setting. *Journal of Science Education and Technology*, *18*, 209–223. doi:10.1007/s10956-009-9144-8

Songer, N. B., Lee, H.-S., & McDonald, S. (2003). Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*, *87*, 490–516. doi:10.1002/sce.10085

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*, 583–639. doi:10.1111/1467-9868.00353

Srinivasan, M., Wilkes, M., Stevenson, F., Nguyen, T., & Slavin, S. (2007). Comparing problem-based learning with case-based learning: Effects of a major curricular shift at two institutions. *Academic Medicine*, *82*, 74–82. doi:10.1097/01.ACM.0000249963.93776.aa

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, *105*, 970–987. doi:10.1037/a0032447

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, *106*, 331–347. doi:10.1037/a0034752

Stettler, C., Wandel, S., Allemann, S., Kastrati, A., Morice, M. C., Schömig, A., . . . Dirsken, M. T. (2007). Outcomes associated with drug-eluting and bare-metal stents: a collaborative network meta-analysis. *Lancet*, *370*, 937–948. doi:10.1016/S0140-6736(07)61444-5

Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities*, *31*, 344–364. doi:10.1177/002221949803100404

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement*, *14*, 29–35. doi:10.1111/j.1745-3992.1995.tb00865.x

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education*, *3*, 285–289. doi:10.4300/JGME-D-11-00147.1

Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, *10*, 277–303. doi:10.1177/096228020101000404

Swanson, H. L., & Deshler, D. (2003). Instructing adolescents with learning disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities*, *36*, 124–135. doi:10.1177/002221940303600205

Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, *71*, 321–363. doi:10.3102/00346543071002321

Tan, S. C., Loong, D. H. W., & So, K. L. (2005). Fostering scientific inquiry in schools through science research course and computer-supported collaborative learning (CSCL). *International Journal of Learning Technology*, *1*, 273–292. doi:10.1504/IJLT.2005.006518

Thistlethwaite, J. E., Davies, D., Ekeocha, S., Kidd, J. M., Macdougall, C., Matthews, P., . . . Clay, D. (2012). The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical Teacher*, *34*, e421–e444. doi:10.3109/0142159X.2012.680939

Thompson, J. (2014). *Bayesian analysis with STATA* (1st ed.). College Station, TX: Stata Press.

Tiekstra, M., Minnaert, A., & Hessels, M. G. P. (2016). A review scrutinising the consequential validity of dynamic assessment. *Educational Psychology*, *36*, 112–137. doi:10.1080/01443410.2014.915930

U.S. Department of Education, Institute of Education Sciences, & National Center for Education Evaluation and Regional Assistance. (2006). *Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide*. Washington, DC: Council for Excellence in Government.

van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, *22*, 271–296. doi:10.1007/s10648-010-9127-6

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*, 197–221. doi:10.1080/00461520.2011.611369

Walker, A. E., Belland, B. R., Kim, N. J., & Piland, J. (2017). *Examining computer based scaffolding research quality through a risk of bias lens*. Paper presented at the 2017 Annual Meeting of the American Educational Research Association, San Antonio, TX.

Walker, A. E., & Leary, H. (2009). A problem based learning meta analysis: Differences across problem types, implementation types, disciplines, and assessment levels. *Interdisciplinary Journal of Problem-Based Learning*, *3*, 12–43. doi:10.7771/1541-5015.1061

White, W. A. T. (1988). A meta-analysis of the effects of direct instruction in special education. *Education & Treatment of Children*, *11*, 364–374.

Wiggins, G., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Williams, J., Thomas, O., Ernst, J. V., & Kaui, T. M. (2015). Special populations at-risk for dropping out of school: A discipline-based analysis of STEM educators. *Journal of STEM Education: Innovations and Research*, *16*, 41–45.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*, 89–100. doi:10.1111/j.1469-7610.1976.tb00381.x

Xin, Y. P., Tzur, R., Hord, C., Liu, J., Park, J. Y., & Si, L. (2017). An intelligent tutor-assisted mathematics intervention program for students with learning difficulties. *Learning Disability Quarterly*, *40*, 4–16. doi:10.1177/0731948716648740

Yadav, A., Subedi, D., Lundeberg, M. A., & Bunting, C. F. (2011). Problem-based learning: Influence on students' learning in an electrical engineering course. *Journal of Engineering Education*, *100*, 253–280. doi:10.1002/j.2168-9830.2011.tb00013.x

Authors

BRIAN R. BELLAND is an associate professor in the Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322, USA; email: *brian.belland@usu.edu*. His research interests center on the use of computer-based scaffolding to enhance middle and high school students' argumentation and problem-solving abilities during problem-based units in science. He also is interested in leveraging what is known throughout the computer-based scaffolding literature to design more effective scaffolding.

ANDREW E. WALKER is an associate professor in the Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322, USA; email: *andy.walker@usu.edu*. His research involves exploring problem-centered pedagogies like problem-based learning; meta-analysis techniques including traditional, network, and Bayesian meta-analysis; and leveraging how both of these traditions can help inform technology teacher professional development.

NAM JU KIM earned his PhD in the Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322, USA and is beginning as an assistant professor of applied learning sciences at the University of Miami in Fall 2017; email: *namju1001@gmail.com*. His research interests include the utilization of immersive technologies and problem-based learning to improve K–12 students' content knowledge and higher order thinking skills in STEM education. He also has a broad background in methodology with advanced statistical methods.