Review of Educational Research
April 2017, Vol. 87, No. 2, pp. 309–344
DOI: 10.3102/0034654316670999

Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis

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

Computer-based scaffolding assists students as they generate solutions to complex problems, goals, or tasks, helping increase and integrate their higher order skills in the process. However, despite decades of research on scaffolding in STEM (science, technology, engineering, and mathematics) education, no existing comprehensive meta-analysis has synthesized the results of these studies. This review addresses that need by synthesizing the results of 144 experimental studies (333 outcomes) on the effects of computer-based scaffolding designed to assist the full range of STEM learners (primary through adult education) as they navigated ill-structured, problem-centered curricula. Results of our random effect meta-analysis (a) indicate that computer-based scaffolding showed a consistently positive ($\bar{g}$ = 0.46) effect on cognitive outcomes across various contexts of use, scaffolding characteristics, and levels of assessment and (b) shed light on many scaffolding debates, including the roles of customization (i.e., fading and adding) and context-specific support. Specifically, scaffolding's influence on cognitive outcomes did not vary on the basis of context-specificity, presence or absence of scaffolding change, and logic by which scaffolding change is implemented. Scaffolding's influence was greatest when measured at the principles level and among adult learners. Still scaffolding's effect was substantial and significantly greater than zero across all age groups and assessment levels. These results suggest that scaffolding is a highly effective intervention across levels of different characteristics and can largely be designed in many different ways while still being highly effective.

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

Computer-based scaffolding assists students as they generate solutions to complex and ill-structured problems, goals, or tasks, helping students enhance domain knowledge and higher order thinking skills (Wood, Bruner, & Ross, 1976). Given the shift to problem-centered models of instruction prompted by the Next Generation Science Standards and the Common Core (McLaughlin & Overturf, 2012), scaffolding has grown in importance in science, technology, engineering, and mathematics (STEM) education. The increased importance has led to an increase in primary research that indicates that scaffolding has a positive impact on student learning. Although there are meta-analyses on scaffolding types, such as dynamic assessment (Swanson & Lussier, 2001), scaffolding in intelligent tutoring systems (Ma, Adesope, Nesbit, & Liu, 2014; VanLehn, 2011), scaffolding for students with learning disabilities (Swanson & Deshler, 2003), and a pilot meta-analysis on a wider swath of computer-based scaffolding (Belland, Walker, Olsen, & Leary, 2015), there are no comprehensive meta-analyses on computer-based scaffolding. Thus, it is difficult to design scaffolding-enhanced learning environments that provide the greatest student success. The purpose of this article is to conduct a comprehensive meta-analysis of computer-based scaffolding in STEM education.

Promotion of Critical Thinking Abilities and Deep Content Knowledge

The widespread adoption of the Common Core State Standards and the Next Generation Science Standards has prompted an increased focus on methods to increase critical thinking skills (Alexander, 2014; Kettler, 2014; Murphy, Rowe, Ramani, & Silverman, 2014; Stage, Asturias, Cheuk, Daro, & Hampton, 2013) and deep content knowledge (Scruggs, Brigham, & Mastropieri, 2013; Stage et al., 2013) among all K–12 students. Methods designed to help students learn critical thinking skills include (a) teaching critical thinking skills explicitly and either stopping there (general critical thinking skills) or inviting students to think critically about a topic (infusion), (b) involving students in subject matter instruction without making critical thinking skills explicit (immersion), or (c) a combination of general and either infusion or immersion (mixed; Ennis, 1989). An early meta-analysis of these critical thinking approaches indicated that immersion led to a statistically lower average effect size ($\bar{g} = 0.09$) than the remaining approaches (Abrami et al., 2008), but a more comprehensive follow-up found no differences between them (Abrami et al., 2015). A relatively small effect ($\bar{g} = 0.18$) of immersion interventions to promote critical thinking skills was found by others (Niu, Behar-Horenstein, & Garvan, 2013), so the evidence appears to be mixed. Some of the variance in findings may be attributable to limitations of the Ennis (1989) framework. It is possible to immerse students in meaningful content instruction and provide nonexplicit support for the development of critical thinking skills. Such an approach can be found in problem-centered instructional models paired with scaffolding (Wood et al., 1976).

Instructional Scaffolding Used in the Context of Problem-Centered Instruction

To reach more students and help them learn how to use cross-disciplinary approaches to address authentic problems, recent initiatives have encouraged (a) the use of problem-centered models of instruction in science (National Research

Council, 2011) and (b) the integration of science with the rest of STEM (Achieve, 2013; National Research Council, 2012; Next Generation Science Standards, 2013). Problem-centered instructional approaches used in STEM education include problem-based learning, modeling/visualization, case-based learning, design-based learning, project-based learning, inquiry-based learning, and problem solving. At the center of all such approaches are ill-structured, authentic problems, defined as problems with no clear goal or path to the goal, and which relate to students' communities and/or lives (Barab, Squire, & Dueber, 2000; Hung & Chen, 2007; Jonassen, 2011). Problem-centered instructional approaches can be considered contexts of scaffolding use, as scaffolding is often present in the context of the former. Sometimes, scaffolding takes the form of one-to-one support provided by a more capable other. Centering instruction on authentic problems while also allowing for extensive student–teacher and student–student dialogue and one-to-one mentoring led to a statistically stronger effect ($\bar{g} = 0.57$) on critical thinking skills than authentic instruction ($\bar{g} = 0.25$) or dialogue ($\bar{g} = 0.23$) by itself, or authentic instruction combined with dialogue ($\bar{g} = 0.32$; Abrami et al., 2015). Other times, scaffolding is delivered via computer-based tools. A recent pilot meta-analysis found no significant difference in cognitive outcomes when computer-based scaffolding was used in the context of two problem-centered approaches—inquiry-based learning and problem solving (Belland et al., 2015). A more comprehensive meta-analysis that covers a wider swath of literature and more problem-centered instructional models is needed.

Scaffolding Components

To facilitate problem-centered instructional models, one needs to provide scaffolding (Hmelo-Silver, Duncan, & Chinn, 2007). Scaffolding originally referred to contingent support from a more capable other that helped toddlers solve complex problems and to gain valuable skills while doing so (Wood et al., 1976). In terms of overall approach, scaffolding encompassed three key characteristics: contingency, intersubjectivity, and transfer of responsibility (Wood et al., 1976). Contingency meant that teachers dynamically assessed students' current abilities through questioning or observation and provided just the right amount of support. Scaffolders then continued to engage in dynamic assessment throughout the scaffolding process, adding and fading support as needed, eventually fading the support completely when students could complete the target task unassisted. Contingency also meant that teachers could provide a tailored strategy using either a generic or a context-specific approach based on what dynamic assessment indicated was needed. Intersubjectivity meant that students needed to be able to recognize a successful solution to the problem that they were addressing (Wood et al., 1976). Without intersubjectivity, students would not be able to take on more responsibility until eventually able to perform the task independently (Wood et al., 1976). Transfer of responsibility meant that successful scaffolding would help students learn to complete the target tasks independently.

Scaffolding strategies include recruitment, controlling frustration, highlighting critical problem features, questioning, modeling expert processes, providing feedback, task structuring, direction maintenance, and demonstration (van de Pol, Volman, & Beishuizen, 2010; Wood et al., 1976). The exact combination of strategies that were deployed typically depended on the needs uncovered through dynamic assessment. Furthermore, teachers provided such support on a context-specific or a generic basis, based on a determination of what students needed. Context-specific scaffolding incorporated content knowledge, whereas generic scaffolding did not.

One-to-one scaffolding soon began to be used among many populations (van de Pol et al., 2010) and across contexts (Palincsar & Brown, 1984). Although very effective, it was not practical as a sole source of support in K–12 classrooms, since large class sizes impede teachers from working one-to-one with students on a large scale. Researchers soon considered how computers could provide scaffolding (Hawkins & Pea, 1987).

Existing Meta-Analyses of Computer-Based Scaffolding

Evidence indicates that computer-based scaffolding is highly effective in promoting cognitive outcomes. For example, a pilot meta-analysis of computer-based scaffolding indicated that computer-based scaffolding led to an average effect of $\bar{g} = 0.53$ (Belland et al., 2015). There also have been several meta-analyses of intelligent tutoring systems, which combine scaffolding with some additional elements, such as adaptivity of content presentation. A meta-analysis of intelligent tutoring systems indicated that step-based intelligent tutoring systems led to an average effect of $ES = 0.76$ versus control and that substep-based intelligent tutoring systems led to an average effect of $ES = 0.40$ (VanLehn, 2011). Other meta-analyses have found varied average effects for intelligent tutoring systems, including $\bar{g} = 0.41$ among students of various levels (Ma et al., 2014), $\bar{g} = 0.09$ among K–12 students engaged in mathematics learning (Steenbergen-Hu & Cooper, 2013), and $\bar{g} = 0.37$ among college students (Steenbergen-Hu & Cooper, 2014). But clearly, there are many scaffolding types that have not been addressed through meta-analysis, or at least not through a comprehensive meta-analysis technique.

Remaining Questions About Computer-Based Scaffolding

Computer-based scaffolding employs many of the same strategies proposed in the original scaffolding definition (Wood et al., 1976). However, the way that it packages and deploys those strategies is very different due to the need to program computer-based scaffolding prior to student use. This results in considerably less contingency. For example, in one-to-one teacher scaffolding, teachers can dynamically assess student understanding and select exactly those strategies that fit the target students' current needs. Within a given lesson, a teacher could in theory dynamically produce thousands of different combinations and versions of scaffolding messages. Computer-based scaffolding needs to be programmed ahead of time, and so scaffolding messages and strategies need to be packaged ahead of time.

One way to think about this is in terms of conceptual, strategic, metacognitive, and motivation scaffolding (Hannafin, Land, & Oliver, 1999). Conceptual scaffolding suggests things to consider when addressing the problem (Hannafin et al., 1999). Strategic scaffolding bootstraps a target strategy, such as argumentation, problem solving, or evaluation (Hannafin et al., 1999). Metacognitive scaffolding helps students question their understanding and evaluate their progress (Hannafin et al., 1999;

Quintana, Zhang, & Krajcik, 2005). Motivation scaffolding supports motivational variables such as students' self-efficacy, autonomy, connectedness, mastery goals, and perceptions of the value of the target task (Belland, Kim, & Hannafin, 2013). A pilot meta-analysis compared conceptual and metacognitive scaffolding, finding that conceptual scaffolding led to stronger effects (Belland et al., 2015). It is an open question as to whether there are differences in cognitive outcomes based on a broader array of scaffolding types, but this can be addressed through meta-analysis.

Next, designers of computer-based scaffolding need to choose whether to embed target content in the scaffolding strategies (context-specific scaffolding) or to use generic scaffolding strategies (McNeill & Krajcik, 2009). This choice is often informed by the theoretical model that drives the scaffolding design. When driven by adaptive control of thought–rational (Koedinger, & Aleven, 2007; VanLehn, 2011) or knowledge integration (Linn, 2000), scaffolding tends to be context-specific. When driven by cultural–historical activity theory (Leont'ev, 1974; Luria, 1976), scaffolding can be either context-specific or generic. According to a pilot meta-analysis of research on computer-based scaffolding in STEM education, there was no difference in cognitive outcomes between generic and context-specific scaffolding (Belland et al., 2015). But it is worthwhile to see if that trend holds in a comprehensive meta-analysis of research on computer-based scaffolding in STEM education.

Developers of computer-based scaffolding often tried to mimic the adding and fading of scaffolding inherent in one-to-one scaffolding, and many have argued that scaffolding must be faded to be called scaffolding (Pea, 2004; Puntambekar & Hubscher, 2005). Scaffolding informed by different theoretical traditions is often implemented differently. In scaffolding informed by adaptive control of thought–rational, scaffolding is almost universally added and faded (Koedinger & Aleven, 2007). In scaffolding informed by knowledge integration and cultural historical activity theory, scaffolding is sometimes faded, but rarely added. Another variation in the implementation of fading and adding in computer-based scaffolding is an expansion in the bases by which fading and adding is performed. Whereas in one-to-one scaffolding, fading and adding are always performed on the basis of dynamic assessment, in computer-based scaffolding, such is often done also according to a fixed schedule or self-selection. A pilot meta-analysis of computer-based scaffolding in STEM education indicated that cognitive outcomes were superior when scaffolding was not faded versus when it was faded on a fixed schedule (Belland, Walker, Olsen, & Leary, 2015). A more comprehensive meta-analysis that covers more variations of fading and adding and fading/adding bases is needed to fully understand scaffolding customization and customization bases.

As the formats of scaffolding expanded, so did the intended learning outcomes and populations targeted by scaffolding. What was once an intervention designed to help toddlers develop problem-solving skills through one-to-one interaction with a teacher was now a multifaceted intervention that targeted diverse learning outcomes among learner populations that were diverse in age, subject matter, learning skill, and demographic characteristics. Targeted learning outcomes of scaffolding now included deep content learning (Davis & Linn, 2000; Linn, 2000), argumentation ability (Hong, Lin, Wang, Chen, & Yang, 2013; Jeong & Joung, 2007; McNeill & Krajcik, 2009), and problem-solving ability (Ge & Land, 2003; González-Calero, Arnau, Puig, & Arevalillo-Herráez, 2015; Kim & Hannafin, 2011). To assess these different forms of learning, it is necessary to use different assessments.

One way to consider this is through reference to the assessment framework of Sugrue (1995), who categorized assessments into concept, principles, and application levels. Concept-level assessments measure students' ability to recall or understanding of target content. Principles-level assessments measure the ability to predict what would happen in a hypothetical situation. Application-level instruments assess the ability to apply principles and processes to solving a novel problem. Previous meta-analyses of intelligent tutoring systems (Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2014; VanLehn, 2011) did not use the classification of assessments into concept, principles, and application as a moderator. Steenbergen-Hu and Cooper (2013) found no difference in the effect of intelligent tutoring systems based on whether the assessments were course-related data, researcher-made tests, or standardized tests. A pilot meta-analysis indicated that there was no difference on the basis of assessment levels, but the number of studies at each assessment level was low, contributing to wide confidence intervals (Belland et al., 2015). There has not yet been a comprehensive effort to explore assessment levels as a moderator in scaffolding outcomes.

With the expansion of the scaffolding metaphor, targeted learner populations now included students ranging from elementary school (Hong et al., 2013) to graduate school (Hadwin, Wozney, & Pontin, 2005), and everyone in between. Furthermore, what was once targeted to middle-class children now served students from various demographic subgroups. It is natural to question whether scaffolding's effectiveness varies based on these learner attributes. This is an empirical question that can be addressed through meta-analysis.

The intended learning outcomes of scaffolding range widely, including cognitive (Reiser, 2004), motivational (Moos & Azevedo, 2008), and metacognitive outcomes (Quintana et al., 2005). For example, cognitive outcomes can include increased problem-solving and argumentation abilities and deep content knowledge (Davis & Linn, 2000; Ge & Land, 2003), motivational outcomes of scaffolding can include enhanced self-efficacy and engagement (Alias, 2012; Moos & Azevedo, 2008), and metacognitive outcomes can include increased knowledge of what one knows and enhanced ability to assess one's processes (Quintana et al., 2005). However, due to the scope of the article, we decided to include only studies that measured cognitive outcomes. The treatment also needed to meet the definition of scaffolding as proposed by Wood et al. (1976) and be used while students engaged with ill-structured problems. Furthermore, the studies needed to include a control condition, be published between 1993 and 2014, and include enough information to calculate effect size.

The purpose of this meta-analysis was to guide the future design of computer-based scaffolding by addressing the following six research questions. First, what is the impact of providing computer-based scaffolding to students engaged in ill-structured problem solving in STEM education? Second, to what extent do learner characteristics moderate cognitive student outcomes in STEM education, including (a) how does education level moderate cognitive student outcomes and (b) how does education population moderate cognitive student outcomes? Third, how does context of scaffolding use moderate cognitive student outcomes? Fourth, to what extent does assessment level moderate cognitive student outcomes in STEM education? Fifth, to what extent do scaffolding characteristics moderate cognitive student outcomes in STEM education, with subquestions including (a) how does scaffolding change (fading, adding, fading/adding, or none) moderate cognitive student outcomes, (b) how does scaffolding logic (performance-based, self-selected, fixed, or none) moderate cognitive student outcomes, (c) how does scaffolding scope (generic or context-specific) moderate cognitive student outcomes, and (d) how does scaffolding intervention (conceptual, strategic, metacognitive, or motivational) moderate cognitive student outcomes? Finally, to what extent does scaffolding study quality moderate cognitive student outcomes in STEM education, including (a) how does study design (random, group random, quasi-experimental) moderate cognitive student outcomes, (b) how does reliability reporting moderate cognitive student outcomes, and (c) how does validity reporting moderate cognitive student outcomes?

Method

Literature Search Procedure

We engaged in three search efforts. Initially, we searched Education Source, PsychINFO, Digital Dissertations, CiteSeer, Proquest, ERIC, PubMed, Academic Search Premier, IEEE, and Google Scholar databases using various combinations of the following search terms: *scaffold**, *computer**, *tutor**, *intelligent tutoring system**, and *cognitive tutor**. To increase results in engineering and mathematics as well as underserved populations, we next conducted hand searches of *Computer Applications in Engineering Education, Journal of Geoscience Education, Journal of Professional Issues in Engineering Education and Practice, International Journal of Mathematical Education in Science and Technology, Journal for Research in Mathematics Education*, and *The Journal of Special Education*. Some of these journals were suggested by advisory board members. Others were journals where articles on scaffolding in mathematics or engineering education or articles including underserved populations were found previously. Last, we searched the reference lists of included studies for referrals to other primary research (see Figure 1).

Inclusion Criteria

To be included in this meta-analysis, studies had to (a) be published between January 1, 1993, and December 31, 2014; (b) have a control condition in which students received an educational intervention but did not receive scaffolding; (c) measure cognitive outcomes; (d) provide sufficient information for calculating effect sizes; (e) provide assistance or scaffolds as defined by Wood et al. (1976) to learners who were (f) engaged in STEM problems that were ill-structured. Problems had to incorporate at least one of the following ill-structured problem elements: (a) unknown or uncertain problem features, (b) multiple solution paths to multiple or no viable solution, (c) multiple criteria for success resulting in ambiguity about appropriate concepts or procedures, or (d) judgments or personal opinion (Jonassen, 2000).

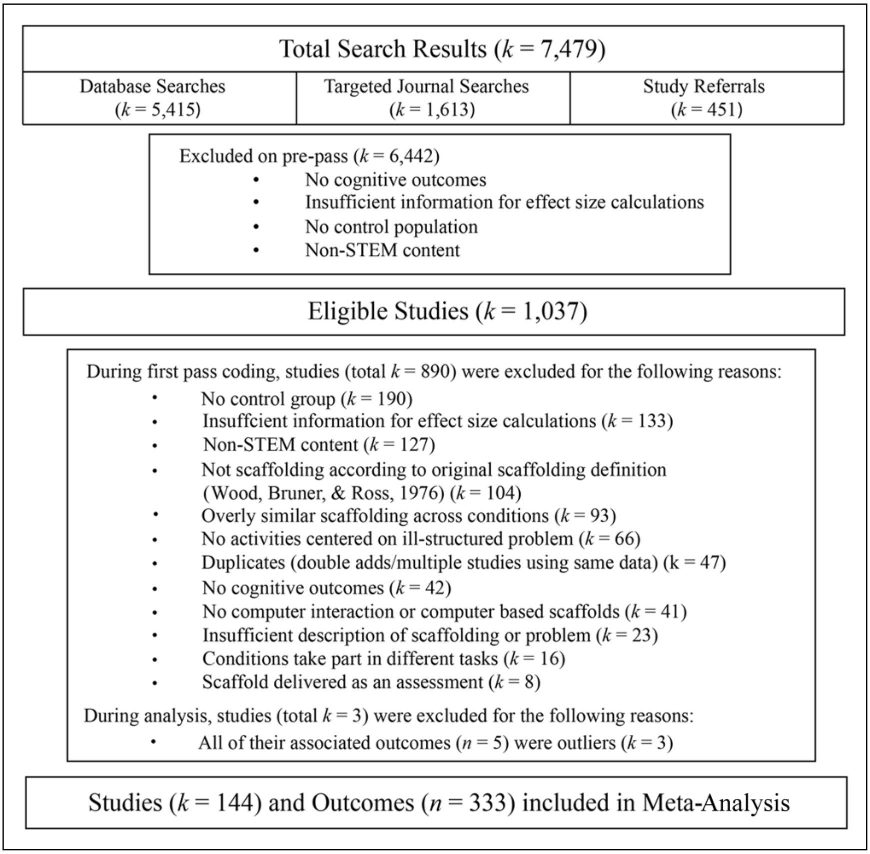

FIGURE 1. *Search and exclusion process. k refers to number of studies, and n refers to number of outcomes. The final set of 144 included studies were associated with 333 outcomes.*

Study Feature Coding

A robust set of features was coded from a mix of theoretically defined constructs and categories that were emergent as part of the research process. All included studies used a treatment-comparison design. Effect sizes were calculated for each outcome using a free online tool (ESFREE: http://esfree.usu.edu/). When possible, we chose calculations that took into account pretest measures (e.g., analysis of covariance F statistic, pre–post change score means with intraclass correlations). All reported effect sizes used the Hedges' g calculation.

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

From a theoretical perspective, scaffolding began with young children (Wood et al., 1976), but it was quickly apparent that scaffolding had branched out to the full spectrum of educational contexts. Recommendations about appropriate tasks and pedagogical approaches vary between these learners to the extent that whole fields of theory, such as developmental psychology (Piaget, 1947) and andragogy (Knowles, 1984), among others, have emerged. When a study included students at multiple education levels, we applied the code corresponding to the largest number of participants.

Education Population (Traditional, Low-Income, Underrepresented, High-Performing, Underperforming)

In addition to the participant education level, we also coded participant characteristics such as prior knowledge (high-performing, underperforming) or socio-economic status. When coding for prior knowledge, many studies assessed students' knowledge prior to the intervention, split them into underperforming and high-performing groups and then reported posttest scores as outcomes (Su & Klein, 2010). In another example, Ross and Bruce (2009) asked teachers to use a set of test results to identify students at the bottom quartile of their grade level and used that as their sampling frame. Sometimes, students were not broken down by performance levels on the pretest but the case was made that the entire school population could be classified as high-achieving based on the school's national ranking on an academic achievement test and the student body's performance in academic competitions (Tan, Loong, & So, 2005). When 33% or more of the student population qualified for free/reduced lunch, received Pell grants, and/or the family income was 125% of the poverty level of a family of its size, student population was coded as low-income.

Student population was coded as underrepresented based on either ethnicity/race or gender when a large portion of participants was typically not represented within a given discipline. For example, participants in Rieber, Tzeng, and Tribble (2004) were over 90% female and received scaffolding in the area of physics. In another example, education population in Siegel (2006) was coded as underrepresented, since 86% (42% African American, 34% Hispanic, 4% Native American, 1% Filipino, 1% Pacific Islander/Other) of the participants were learners from races/ethnicities not proportionally represented in STEM fields.

Assessment Level (Concept, Principles, Application)

This category borrows from Sugrue (1995), an assessment framework for problem-solving contexts used in prior meta-analyses (Dochy, Segers, Van den Bossche, & Gijbels, 2003; Gijbels, Dochy, Van den Bossche, & Segers, 2005). Concept assessments are about facts and ideas, such as asking children to recall essential terms and ideas from the lesson (Ulicsak, 2004). At the principles level, learners must understand the concepts but also the relationships between two or more concepts. Moreno and Mayer (2005) measured principles-level knowledge by assessing whether students could design plants in alien environments that varied in terms of temperature, soil nutrients, and water. Finally, at the application level, learners use what they know at the concept and principles level to solve a holistic and authentic problem. The application-level assessment items in Kramarski and Gutman (2006) required students to use higher order thinking skills to transfer their concept and principle-level knowledge to solve other complex, "real-life" problems.

Sometimes we encountered studies that employed data collection instruments that included items at multiple levels (e.g., concept and principles). When scores were broken down into scale scores, we kept each scale separate and associated such with the appropriate assessment level, such as in Parchman, Ellis, Christinaz, and Vogel (2000), who used the Navy Personnel Research and Development Center test. The scores were broken into the following subscales, which were classified according to the assessment level in parentheses: knowledge of definitions (concept), knowledge of symbols (concept), qualitative knowledge (principles), and quantitative knowledge (principles). When scores were not broken out according to assessment level, we coded the most frequently occurring set of assessment items.

Context of Scaffolding Use (Problem-Based, Case-Based, Design-Based, Project-Based, Inquiry-Based, Modeling/Visualization, Problem Solving)

Scaffolding is not a good fit for traditional pedagogies and is often used alongside a variety of problem-centered instructional models. Differences among these approaches lie in what comes before and after problem presentation. In case-based learning, content needed to address the problem is delivered to students before engagement with the problem, often via lecture (Srinivasan, Wilkes, Stevenson, Nguyen, & Slavin, 2007; Thistlethwaite et al., 2012). For example, Zhang, Chen, and Reid (2000) described principles of effective research designs before illustrating them with example cases. In other problem-centered models, content is typically learned after presentation of the problem.

These models also differ in what students need to produce. In problem-based learning, students produce a conceptual solution to the problem (Hmelo-Silver, 2004). For example, Zydney (2008) presented learners with a complex pollution problem first, then provided resources and scaffolding to aid them in working toward recommending a solution. In project-based learning, students produce an artifact (e.g., video; Krajcik et al., 1998). For example, Aydin and Cagiltay (2012) asked students to conceptualize, create, and then collect data on the performance of their own microwave filters. In this case, the filter itself is an artifact. In design-based learning, students design a product (e.g., a levee) that can address a problem (Kolodner et al., 2003). Inquiry-based learning typically invites students to ask, and set up an experiment to address, questions (Keys & Bryan, 2001). For example, X. Lin and Lehman (1999) asked students to design and engage in simulated experiments on pill bug behavior by manipulating environmental factors. Afterward, they were asked to draw conclusions.

In modeling/visualization, the focus is on students making visual models that represent relationships among underlying variables (Lesh & Harel, 2003) or by presenting these kinds of visuals to students. For example, Linn and Eylon (2000) showed students animations that reveal mass and volume as independent constructs. When no specific pedagogy was identified, studies were coded as problem-solving. Instruction centered on authentic, ill-structured problems but did not involve the processes or goals of prominent problem-centered instructional models. For example, Katai (2011) helped students solve sample recursion problems in computer science by reorganizing students' code to highlight key features and provide step-by-step output when tested.

Scaffolding Change (None, Fading, Adding, Fading/Adding)

Theorists have often argued that scaffolding needs to be removed (or faded) over time based on continuous assessment of the student's growing knowledge and skillset. However, early theoretical efforts (Wood et al., 1976) suggested a broader range of scaffolding change than just fading. Parallel to this broader conceptualization of scaffolding change, we observed the entire range of studies, including interventions that withdrew (fading) support, increased (adding) support, and did both (fading/adding) in addition to studies that made no changes over time. When coding scaffolding change, we looked both for changes in frequency or interval of scaffolding as well as changes to the underlying nature of the scaffolds. As an example of fading, in Chen, Chen, and Chen (2013), all students were provided with a partially completed expert concept map to begin with but then would lose parts of that map over time. In contrast, in Chang, Sung, and Chen (2001), learners were invited to create a concept map but some scaffolds were constant (e.g., prompting them to reflect on their progress), while other scaffolds could be added at the learner's discretion by pressing a hint or "expert concept" button. These hints would begin by only providing a partial description of the linkages between concepts and then progress to a show a more complete description or even later (after a half hour of constructing their own concept map) would finally reveal the expert concept map (Chang et al., 2001).

It is important to note that in both the fading example (Chen et al., 2013) and the adding example (Chang et al., 2001), the frequency and the nature of the scaffolding only moved in one direction (increasing support or decreasing support). This is distinguished from other cases where support was decreased and increased (fading/adding). The SE-Coach provided feedback and self-explanation prompts to students based on a continuous assessment of the students' actions and domain knowledge (Conati & Vanlehn, 2000). In it, both the frequency of support as well as the underlying nature of scaffolding was continuously adjusted according to student ability. Scaffolds that neither increased nor decreased, in terms of nature or frequency, over the duration of the intervention were labelled as none.

Scaffolding Logic (None, Performance Adapted, Self-Selected, Fixed)

Scaffolding logic is especially important in the context of computer-based scaffolding because it also speaks to some of the technological constraints of designing a computer tutor and to the ways in which researchers have worked around those technological deficiencies. In contrast to early scaffolding literature that detailed how human tutors extended scaffolding to young children (Wood et al., 1976) by continuously assessing both learners and their solution trajectories, computer-based scaffolding research has included many examples of scaffolding logic such as none, fixed, performance adapted, and self-selected. Performance-adapted and self-selected scaffolding describe scaffolding logic that is happening during the intervention, while fixed scaffolding logic denotes that the decision of when to add or fade scaffolding was made during the design of the intervention.

Furthermore, scaffolding logic indicates who is making the decision to add or fade scaffolding. In contrast to the other scaffolding logic-coding categories, self-selected scaffolding logic describes scaffolding interventions that left up to learners to determine what scaffolding they want and when they want it. As an example of performance-adapted logic, Conati and Vanlehn (2000) asked students to look at example problems and explain them. Based on those explanations and their use of the interface, a probabilistic model of their understanding was built, which in turn drove what scaffolding prompts the students received and when. Self-selected scaffolding logic is seen in Chang et al. (2001), where the onus of what type of scaffolding and when to receive that scaffolding is left up to the student through the use of several feedback and hint buttons.

The practical realities of scaffolding at scale, however, result in a variety of approaches including fixed scaffolding logic where changes in the scaffolding was decided at specific predefined moments in the intervention or after a set amount of time. For instance, Raes, Schellens, De Wever, and Vanderhoven (2012) used fixed logic to progress from a full set of scaffolds to a less supportive version of the scaffolding (e.g., no sentence starters) at the middle of the project and a least supportive version of the scaffolds at the end of the project (e.g., no sources were provided; Raes et al., 2012). When scaffolds were consistent throughout the intervention, logic was coded as none.

Scaffolding Scope (Generic, Specific)

Scaffolding scope denotes the presence or absence of content within the scaffold. A generic scaffold can be used in a variety of units and contexts without changes in the scaffold itself. As an example, generic question prompts might ask students, "How do I define the problem? . . . What are my reasons/arguments for my proposed solution? . . . Am I on the right track?" (Stark, 2013, p. 50). On the other hand, a specific scaffold contains content elements that would need to be modified if applied to any other content area such as the conceptual question prompt "for an object floating in water, which force or forces are acting on it" (Reid, Zhang, & Chen, 2003, p. 12). In cases where there was a lack of explicit evidence in the text to guide the coding process, we made inferences based on other contextual clues. For example, in Deters (2009), the scaffolds took the form of metacognitive question prompts but lacked examples. In this case, the two coders inferred that the code was generic since the scaffolding prompted students to reflect on their own thinking and progress toward the solution. It is also important to note that some theoretical roots, such as activity theory, allow for either generic or specific scaffolding (Belland, 2011).

Scaffolding Intervention (Conceptual, Metacognitive, Strategic, Motivation)

Conceptual scaffolds indicate things to consider when investigating the problem. For example, a question prompt in Zydney (2008) asked learners to describe the relationship between their clients' goals and activities and an ongoing acid rain problem. Kramarski and Gutman (2006) presented a series of metacognitive prompts aimed at promoting self-regulated learning. These included framing the problem, reflecting on what they already knew that could help, selecting and justifying the use of appropriate problem-solving strategies, and finally reflecting on their problem-solving process and solution. A strategic scaffold called ALGEBAR helped pairs of students bootstrap problem-solving strategies as they modeled a word problem, represented their model in symbolic (algebra) notation, and then solved equations (Looi & Lim, 2009). Motivation scaffolds aim to positively affect variables such as students' perceptions of autonomy and self-efficacy. For example, Schrader and Bastiaens (2012) delivered motivation scaffolds via a pedagogical agent to encourage learners to keep trying and persevere during rigorous problem-solving tasks.

Study Design (Random, Group Random, Quasi-Experimental)

Randomized control trials represent a high standard in quantitative research. Researchers, practitioners, and policy makers respect random designs because they offer the best chance at an equal playing field for groups, are more sensitive to detecting real differences, and when there are several, they tend to converge better on underlying population statistics. Yet they also do not capture much of what happens, especially in educational settings. We chose a simplified version of the Campbell Collaboration to code for study design (Shadish & Myers, 2004). Random designs include the random assignment of students to two or more treatments. Some studies (e.g., Zhang, Chen, Sun, & Reid, 2004) first categorized students by ability such as high, medium, and low but were still coded as random designs if students from those ability groups were then randomly assigned to a condition. In group random designs, random assignment of all students from an intact group is made to a single condition. For example, Fund (2007) randomly assigned 16 entire classes of students from three different schools to five different treatment groups. Quasi-experimental designs include a range of research, such as purposeful assignment based on a survey of learners' scientific beliefs (Linn & Eylon, 2000). In all cases, studies had to include a control. Pre-experimental or nonexperimental designs such as pretest and posttest only were not included.

Reliability/Validity Reporting (None, Attempt, Strong)

In educational research settings, studies often fail to report reliability statistics, and metrics or descriptions of validity are even more rare (Belland, French, & Ertmer, 2009). We thus fell back on the nature of reliability and validity reporting. Strong reporting, such as Cronbach's alpha scores for pretest and posttest reliability (Osman & Lee, 2013) included a description of analysis techniques as well as results. Reliability/validity reporting was coded as attempt when authors only made reference to (a) prior samples/studies (e.g., Ardac and Sezen, 2002, report reliability from a pilot sample) or (b) an approach but no results. Osman and Lee (2013), for example, described a content validity analysis done with lecturers and teachers but did not describe what they found or changes as a result. Studies failing to describe the stability or alignment of their instruments to intended constructs in any way were coded as None.

Coding Process

Four coders with expertise in scaffolding, meta-analysis, or both coded studies. Working independently, two researchers coded each article as described above. The two coders then came to consensus, and consensus codes were used in all meta-analytic analyses. Each coding pair included one professor and one graduate student. Pairs alternated for a total of four possible pairs.

To ensure consistency in interpretation of coding criteria, we used Krippendorff's alpha to measure interrater reliability after initial coding (and

TABLE 1

Krippendorff's alpha for interrater reliability before coming to consensus

Code	Scale type	Krippendorff's alpha
Assessment level	Nominal	.677
Context of scaffolding use	Nominal	.731
Education level	Ordinal	.975
Education population	Nominal	.875
Effect size calculation	Ratio	.995
Reliability reporting	Ordinal	.697
Scaffolding change	Nominal	.758
Scaffolding intervention	Nominal	.716
Scaffolding logic	Nominal	.704
Scaffolding scope	Nominal	.707
Study design	Nominal	.798
Validity reporting	Ordinal	.735

before coming to consensus) because it (a) is robust for the full range of data (nominal, ordinal, and ratio) used in the coding rubric and (b) adjusts for chance agreement (Krippendorff, 2004). All alphas were greater than .667 (see Table 1), which represents the minimum standard for acceptable reliability (Krippendorff, 2004). Two coders were drawn from a pool of four, and 333 data points were used for the interrater reliability analysis. Validity of coding categories was addressed by means of a content validity check with experts specific to scaffolding, meta-analysis, and each of the STEM disciplines.

Meta-Analytic Procedures/Statistical Analyses

Given the wide range of research participants, subject areas, scaffolding interventions, and study measures, it is unlikely that each outcome represents an approximation of a single true effect size. Thus, we utilized a random effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009) for our study. Analyses were conducted using the metan package of STATA 14.

Publication Bias

There are several ways to detect and mitigate the risk of publication bias, defined as the existence of unpublished primary research studies that, if found, would alter the overall effect size. These include visual inspection of a funnel plot, the trim and fill approach (Borenstein et al., 2009), and Egger's regression test (Egger, Smith, Schneider, & Minder, 1997). All such strategies examine the distribution of effect size estimates relative to standard error and assess whether there is symmetry. Many advocate using a combination of approaches (Borenstein et al., 2009).

We examined evidence of publication bias in a funnel plot showing the relationship between the standard deviation and the effect size (see Figure 2). Among coded outcomes, there were five outliers, having very high (z scores above 3.0 or $g \geq 2.34$ in Figure 2) effect sizes. We excluded all five outlier outcomes (square-shaped

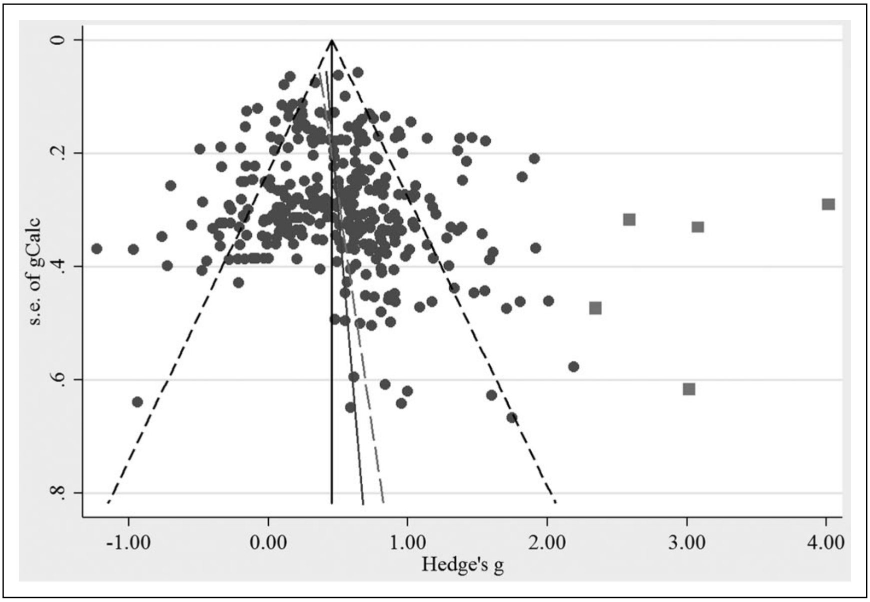

FIGURE 2. *Funnel plot with pseudo 95% confidence limits.*

estimates in Figure 2) and their associated studies ($k = 3$) after further examination of their characteristics (Bernard et al., 2004). Four of the five asked control learners to engage in complex problem solving without any sort of support. In essence, these studies were comparing learners in their zone of proximal development with learners exposed to an excess of cognitive load. Figure 2 shows the funnel plot when adding and deleting the outcomes (square shape). The fitted (dashed) line corresponds to the regression test of the funnel plot asymmetry with outliers and potential publication bias. The fitted (solid) line represents a symmetrical plot after removing outlier outcomes. The funnel plot suggests that there is no publication bias.

To verify this interpretation, we conducted a follow-up Egger's regression test and found no evidence of publication bias (see Table 2). We also used trim-and-fill analysis (Duval & Tweedie, 2000) to compare the observed value and adjusted value as simulating a perfect symmetry but there was no significant difference between observed and adjusted effect sizes. Before examining evidence of publication bias, we had 338 outcomes from 147 studies. By deleting the five outlier outcomes, we brought the number of included outcomes to 333. In so doing, three articles were deleted, bringing the total number of included studies to 144.

Effect Size Dependency

Slightly more than half of the studies had multiple treatment conditions ($k = 79$) or had outcomes at more than one assessment level ($k = 37$), resulting in 333 effect size calculations from the 144 included studies with control groups used in an average of 2.3 comparisons. Including multiple outcomes from the same study may

TABLE 2

Egger's regression analysis results for publication bias

Coefficient	Standard error	*n*	*t*	*p*	95% CI Lower	95% CI Upper
0.336	0.242	333	1.390	.166	−0.141	0.812

Note: n refers to the number of outcomes; CI = confidence interval.

threaten the validity of meta-analytic results by reducing estimates of variance and/or by giving more weight to studies that produced more outcomes. Excluding individual outcomes risks omitting valuable data, or aggregating them in inappropriate ways. We chose to employ a mixed approach. First, we reduced the total number of included effect size outcomes by 35% from 515 to 333 by creating composites (Borenstein et al., 2009) of outcomes from the same study where all coded attributes were identical. Next, we implemented robust variance estimation to empirically test the dependence between remaining outcomes and their study of origin (Hedges, Tipton, & Johnson, 2010). Robust variance estimation attempts to model varying levels of dependence using rho values from 0 (*completely independent*) to 1.0 (*completely dependent*). Our analysis indicates that those extremes do not change subsequent estimates of effect size, or Tau^2 tests of heterogeneity. Since underlying data show no dependency, we report the outcomes as independent to avoid losing the nuance of data with different outcome features.

Results

Impact of Providing Computer-Based Scaffolding to Students Engaged in Ill-Structured Problem Solving

Three hundred thirty-three outcomes across 144 studies were included in the meta-analysis (see Supplementary Table S1 for bibliographic details of included studies and Supplementary Table S2 for coding results according to each outcome; the supplementary tables are available in the online version of the journal). Sixty-five studies had a single outcome and 79 studies included more than one outcome.

The overall mean effect size (see Figure 3) is greater than 0 at a statistically significant level, z = 18.19, $p < .01$, suggesting that students who receive computer-based scaffolding do better on cognitive tests than students who do not receive scaffolds. For the overall effect size, a test for heterogeneity (Q = 1096.96, I^2 = 69.7%, $p < .01$) indicates differences between effect size estimates, which justifies grouping across outcomes in an effort to estimate the overall effect of scaffolding. For each subgroup analysis, the same corpus (n = 333) of outcomes, with the same underlying heterogeneity, are utilized.

Do Learner Characteristics Moderate Cognitive Student Outcomes?

Education Level

Figure 3 contains the number of outcomes (n) and a numerical effect size (Hedges' $\bar{g}$) estimate. As can be seen in Figure 3, Hedges' $\bar{g}$ estimates were

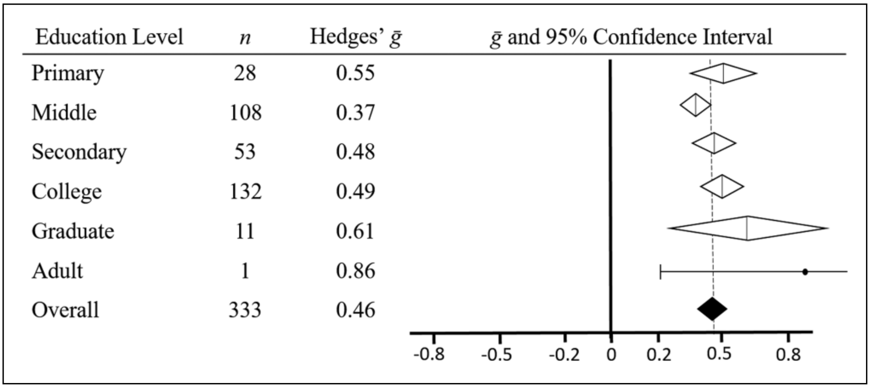

FIGURE 3. *Comparison of effect size according to education level. n refers to the number of outcomes.*

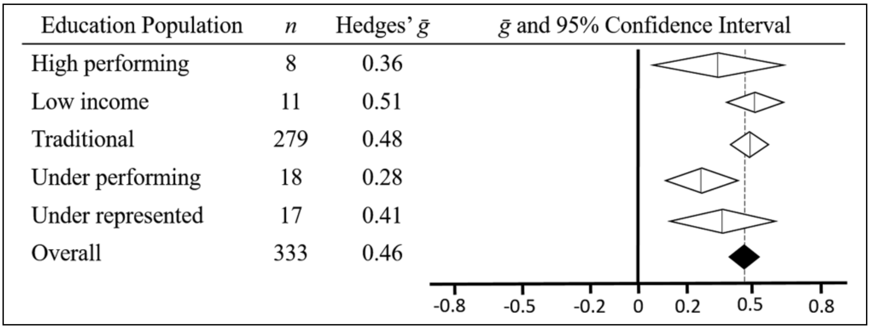

FIGURE 4. *Comparison of effect size according to education population. n refers to the number of outcomes.*

significantly greater than zero and substantial across all education levels, suggesting that scaffolding improves learning for a wide range of students. Hedges' ḡ and confidence intervals are plotted as diamonds; in each diamond, the apex is the Hedges' ḡ estimate, and the diagonals extend in each direction to the upper and the lower limits of the 95% confidence interval. Figures produced in response to other moderator analyses follow this same pattern. The effect size estimate among adult learners was higher than that among college, secondary, middle level, and primary students, $p < .01$. However, caution is warranted, as the effect size estimate for adult learners is based on one outcome.

Education Population

There were wide variations in effect size estimates according to education population subgroups (see Figure 4). The traditional student group accounts for the largest number of outcomes, coming in just above the overall mean.

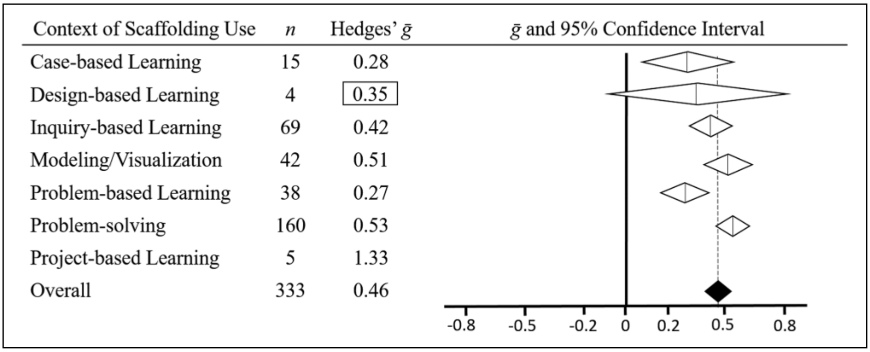

FIGURE 5. *Comparison of effect size according to context of scaffolding use. n refers to the number of outcomes. The effect of design-based learning is not statistically greater than zero, p = .14. Project-based learning (ḡ = 1.33) has an estimate and confidence interval so high that it does not show on our −0.8 to 0.8 scale.*

The estimate for low-income learners is also relatively large. On the contrary, underperforming learners have a small effect size. The difference between traditional and underperforming was significant, $z = 2.29$, $p < .05$.

How Does Context of Scaffolding Use Moderate Cognitive Student Outcomes?

Scaffolds were used alongside several different problem-based instructional models. Hedges' ḡ estimates were significantly greater than zero for all contexts of scaffolding use except design-based learning (see Figure 5), perhaps due to a small sample size for that outcome. Scaffolding's effect size was higher when used in the context of project-based learning than when used in the context of modeling/visualization, $z = 4.69$, $p < .01$, problem solving, $z = 5.09$, $p < .01$, case-based learning, $z = 5.36$, $p < .01$, inquiry-based learning, $z = 5.74$, $p < .01$, design-based learning, $z = 3.90$, $p < .01$, and problem-based learning, $z = 6.08$, $p < .01$. When used in the context of problem solving, scaffolding had a higher effect size than when used in the context of problem-based learning, $z = 2.74$, $p < .01$.

Does Assessment Level Moderate Cognitive Student Outcomes?

Hedges' ḡ estimates were significantly greater than zero across all assessment levels; thus, scaffolding positively influences learning for a variety of assessment types (see Figure 6). Scaffolding's influence was greater when measured at the principles level than when measured at the concept level, $z = 2.17$, $p < .05$.

Do Scaffolding Characteristics Moderate Cognitive Student Outcomes?

Hedges' ḡ estimates were significantly greater than zero across scaffolding customization types (see Figure 7). Differences among effect size estimates were not statistically significant, $p > .05$. In 64.9% of included outcomes, scaffolding did not change over time. The remainder adjusted scaffolding on the basis of learner performance, self-selection, and fixed schedule. Hedges' ḡ estimates were significantly greater than zero across scaffolding logic (see Figure 8). There were no differences in

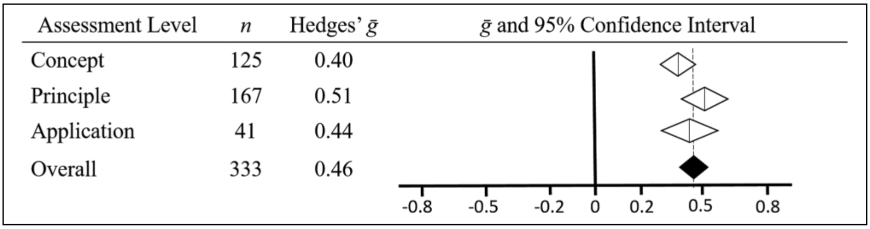

FIGURE 6. *Comparison of effect size according to assessment level. n refers to the number of outcomes.*

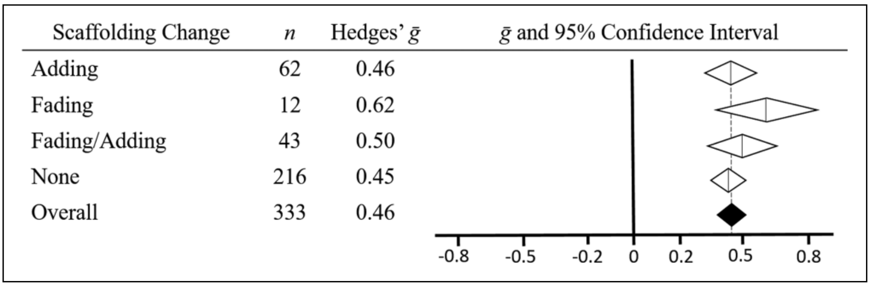

FIGURE 7. *Comparison of effect size according to scaffolding change. n refers to the number of outcomes.*

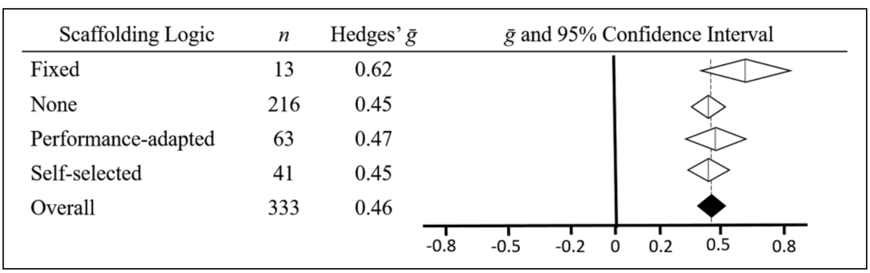

FIGURE 8. *Comparison of effect size according to the basis by which scaffolding was added, faded, or added/faded. n refers to the number of outcomes.*

effect size on the basis of scaffolding logic, $p > .05$. Figure 9 illustrates that generic and context-specific scaffolding were associated with similar cognitive learning outcomes. Each effect size estimate was significantly greater than zero, but the two strategies were not significantly different from each other, $z = -0.281$, $p = .778$. Scaffolding associated with the vast majority of outcomes (82%) was context-specific.

Scaffolding Intervention (Conceptual, Strategic, Metacognitive, or Motivation)

Hedges' $\bar{g}$ estimates were significantly greater than zero across all scaffolding intervention types except for motivation scaffolds (see Figure 10), which means

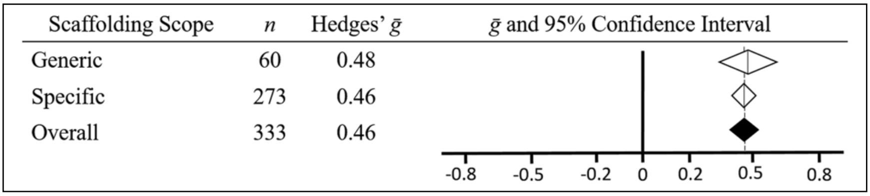

FIGURE 9. *Comparison of effect size according to scaffolding scope. n refers to the number of outcomes.*

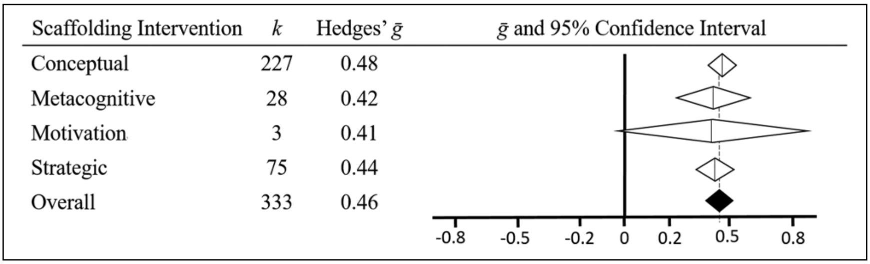

FIGURE 10. *Comparison of effect size according to scaffolding intervention. n refers to the number of outcomes.*

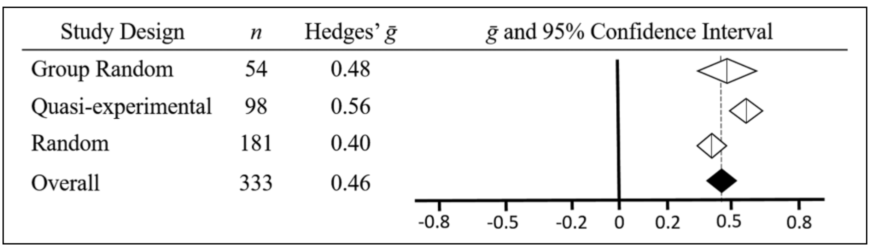

FIGURE 11. *Comparison of effect size according to study design. n refers to the number of outcomes.*

that conceptual, metacognitive, and strategic scaffolds all improve cognitive outcomes. Despite the range in effect sizes, there were no statistically significant differences among the scaffolding intervention types, $p > .05$.

Does Scaffolding Study Quality Moderate Cognitive Student Outcomes?

Study Design

Hedges' $\bar{g}$ estimates were significantly greater than zero and substantial across all included study designs (quasi-experimental, group random, and random), which indicates that each of these study designs has the capacity to allow for the detection of the cognitive outcomes of scaffolding (see Figure 11). When

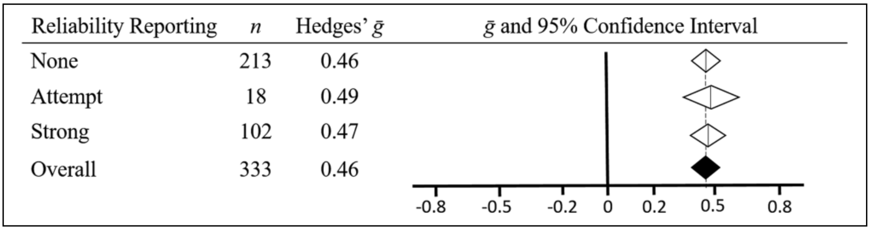

FIGURE 12. *Comparison of effect size according to reliability reporting. n refers to the number of outcomes.*

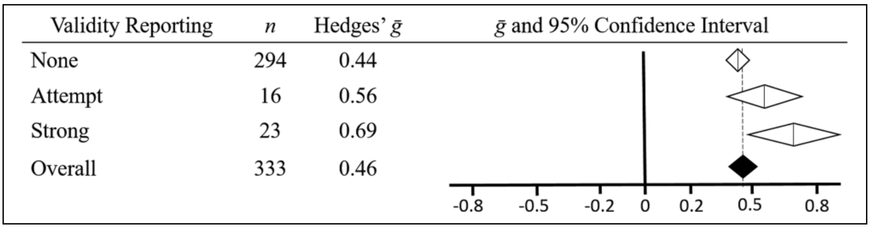

FIGURE 13. *Comparison of effect size according to validity reporting. n refers to the number of outcomes.*

scaffolding was studied using a quasi-experimental design, the effect size estimate was higher than when using a random design, $z = 2.95$, $p < .01$.

Reliability and Validity Reporting

Hedges' $\bar{g}$ estimates were significantly greater than zero across all levels of reliability reporting (see Figure 12). Notably, for 64% of outcomes, there was no reliability reporting at all. There were no differences among levels of reliability reporting, $p > .05$. Effect size estimates were all significantly greater than zero across validity reporting categories (see Figure 13). The effect size estimate when there was strong validity reporting was significantly higher than when there was no validity reporting, $z = 2.27$, $p < .5$.

Discussion

When interpreting effect sizes, one should refer to (a) effect sizes of similar interventions targeting similar outcomes, (b) the gain in the target outcomes that one would see among target learners without an intervention, and (c) practical significance (Durlak, 2009; Hill, Bloom, Black, & Lipsey, 2008; Vacha-Haase & Thompson, 2004). The effect size estimates for scaffolding at the assessment levels of concept, principles, and application were 0.40, 0.51, and 0.44, respectively. Critical thinking outcomes can be seen as cognitive learning outcomes that are measured at the principles and application level. The effect of scaffolding at the principles and application levels compares favorably to the effect size of interventions designed to enhance critical thinking skills among a wide range of learners

(*ES* = 0.34; Abrami et al., 2008) and that of such interventions used among college and graduate students (*ES* = 0.19; Niu et al., 2013). The effect of scaffolding across all assessment levels is also higher than the effect size (*ES* = 0.33) found in a synthesis of 25 meta-analyses on the effect of computer-assisted instruction on cognitive outcomes (Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011).

The overall effect size among elementary school students, middle school students, and high school students were 0.55, 0.37, and 0.48, respectively. The range in terms of effect sizes of average gains on standardized mathematics exams over the course of a year during elementary school was 0.56 to 1.14; the range for middle school was 0.3 to 0.41; the range for high school was 0.01 to 0.22 (Hill et al., 2008). When one considers that the scaffolding treatments in our meta-analysis were considerably shorter than 1 year, the effect size estimate for scaffolding is substantial, especially among high school and middle school students.

One can gauge the practical importance of the results in terms of percentile gains that one would see in a given control student if computer-based scaffolding is used (Lipsey et al., 2012). On average, the use of computer-based scaffolding would bring students who were at the 50th percentile to the 68th percentile (Albanese, 2000). Other strategies that are often proposed for reducing performance gaps include out of school programs (Lauer et al., 2006) and mentoring programs (DuBois, Holloway, Valentine, & Cooper, 2002). In a meta-analysis of out of school programs at the K–12 level, the effect size estimate of out of school programs was 0.16 (Lauer et al., 2006), which would bring students who were at the 50th percentile to the 56th percentile. In a meta-analysis of mentoring programs, the effect size estimate was 0.18 (DuBois et al., 2002), which would bring students who were at the 50th percentile to the 57th percentile. The effect size in this study implies that computer-based scaffolding has the potential to result in a greater reduction in STEM performance gaps, a very important priority.

This meta-analysis also responds to persistent questions in the scaffolding literature, shows where scaffolding's effect is strong, and suggests areas where further research is needed.

Persistent Debates in the Scaffolding Literature

Utility of Fading

There has been consensus among researchers that fading is a necessary component of scaffolding; however, few authors include fading in their scaffolding interventions (Collins, Brown, & Newman, 1989; McNeill, Lizotte, Krajcik, & Marx, 2006; Pea, 2004; Puntambekar & Hubscher, 2005). Only 16.5% of the 333 outcomes in this study included fading, confirming the dearth of this strategy in scaffolding studies (T.-C. Lin et al., 2012). Most studies that did include fading, adding, or fading and adding involved intelligent tutoring systems. Notably, this meta-analysis indicated that including fading did not lead to an effect that was statistically significantly different from the effect when no fading, adding, or fading and adding were employed. This finding differs from our pilot meta-analysis work, in which studies that did not fade scaffolding had higher effect sizes than studies that did fade scaffolding (Belland et al., 2015). But given the attention that fading has been given by scaffolding researchers (Pea, 2004; Puntambekar & Hubscher, 2005), one would expect fading to lead to a significantly higher effect size estimate than not-fading. One argument is that not-fading can lead to overscripting, defined as providing scaffolding when it is in fact unneeded (Dillenbourg, 2002). Overscripting is said to lead to poor motivation and interference with students' cognitive processes (Dillenbourg, 2002). Our finding of no difference in effect sizes between scaffolding that includes fading and scaffolding that includes adding, adding/fading, or no customization suggests that overscripting may not occur or does not negatively affect cognitive outcomes.

Investigation of scaffolding logic indicated no differences based on whether scaffolding change was performance-adapted, fixed, self-selected, or if there was no scaffolding customization at all. This finding may indicate that fading as defined in the scaffolding literature is not necessary to promote student learning and performance (Belland, 2011). Further research is needed to disentangle these results. For example, if more empirical studies can be included that incorporate fading and adding, confidence intervals would likely narrow, and significant differences might emerge. Increasing the number of studies using each form of scaffolding adjustment—performance-based, self-selected, and fixed—may also help determine which adjustment logic is most effective and if any are more effective than no scaffold adjustment.

Generic Versus Context-Specific

Scaffolding researchers often argue whether scaffolding should be generic or context-specific (Davis, 2003; McNeill & Krajcik, 2009; Reiser, 2004). These debates have roots in debates about the domain specificity of problem-solving approaches (for an overview, see Perkins & Salomon, 1989), as well as the idea that students may require context-specific or generic scaffolding depending on the skill that is being supported. The vast majority of outcomes (273 out of 333, or 82%) were associated with context-specific scaffolds. Yet there was no statistically significant difference in effect size estimate between the two approaches. The effect size estimates were so close as to render it unlikely that a significant difference would emerge if more studies on generic scaffolding were found. Even if a significant difference emerged, it would have little practical importance, as the magnitude of the difference would likely be on the order of 0.01 standard deviations. This result implies that scaffolding designers can choose to use generic or context-specific scaffolding depending on the learning needs of the target learners, the nature of the skill to be learned, and scalability considerations, and can do so with confidence that learning goals will be met effectively.

Expansion of the Scaffolding Metaphor

Educational Level

Scaffolding has expanded not only in terms of who or what can provide scaffolding but also in terms of education level and targeted learning outcomes. Scaffolding leads to statistically and practically significant effect sizes among a wide range of education populations, including primary, middle, secondary, college, graduate, and adult—remarkable for a technique that emerged from use with a preschool audience. The highest point estimates of scaffolding's effect on cognitive outcomes were found in graduate and adult education. This means that scaffolding's strongest effects are in populations the furthest from the target learner population in the original scaffolding definition. Although the effect size was lowest for middle school, it is important to note that an effect size of 0.37 (a) would be labeled small to medium by Cohen's (1988) guidelines, (b) is similar to the average effect size found among interventions to promote critical thinking (Abrami et al., 2008), and (c) is higher than the average effect size (*ES* = 0.18) of the strongest educational technology applications for mathematics education in a meta-analysis by Cheung and Slavin (2013).

As important as available data that we examined are the data that are missing from studies not in the literature. Scaffolding's roots are with preschool samples. There are technology-based learning tools associated with this education level but we were unable to find primary research in this population that met our inclusion criteria.

Targeted Learning Outcome

In its original definition, scaffolding was intended to enhance problem-solving skill (Wood et al., 1976). One would measure the outcome of scaffolding in its original form using principles-level or application-level assessments (Sugrue, 1995). We found that scaffolding led to an effect size that was statistically greater than zero across all three assessment levels—concept, principles, and application. Furthermore, the effect size estimates for all three were above 0.40, which is considerably higher than the mean effect size of educational technology applications in mathematics education (Cheung & Slavin, 2013). Future research should use techniques like metaregression to examine the relationship between the targeted learning outcome, assessment level, and scaffolding strategies being used. Such an examination was beyond the scope of this article.

When the effects of problem-centered instructional models implemented without the use of computer-based scaffolding are synthesized through meta-analysis, effect size estimates are not always statistically greater than zero across assessment levels. For example, meta-analyses have indicated that problem-based learning leads to effect sizes that are significantly greater than zero when learning outcomes are assessed at the principles level but not at the concept or application levels (Gijbels et al., 2005), or at the principles and application levels, but not at the concept level (Walker & Leary, 2009). Scaffolding helps problem-centered instructional models go from simply enhancing principles- and/or application-level outcomes, to also enhancing concept-level outcomes. This outcome is important in that it is often necessary to have pertinent content knowledge to be able to apply problem-solving strategies to new situations (Perkins & Salomon, 1989). Scaffolding in these contexts also has the potential to help problem-centered instructional models overcome criticisms that they do not lead to adequate content learning (Kirschner, Sweller, & Clark, 2006).

Areas in Which More Empirical Work Is Needed

As noted previously, when the effect size estimate was not significantly different from zero, the sample size was very small (i.e., three or fewer outcomes were used to calculate the effect size estimate). These circumstances included motivation scaffolding ($n = 3$), design-based learning ($n = 3$), and the adult population ($n = 1$). It is not surprising that these cases exhibited very wide confidence intervals, and it is also important to urge caution in interpreting their effect size estimates.

Scaffolding for Students With Learning Disabilities

One-to-one scaffolding has long been used to support students with learning disabilities, helping such students achieve at a high level and often facilitating their effective inclusion in mainstreamed classrooms (Palincsar, 1998; Stone, 1998). Employing scaffolding among students with learning disabilities encourages them to adopt responsibility for high-level tasks and skills, which is often the opposite of what schooling encourages among students with learning disabilities (Biemiller & Meichenbaum, 1998). However, in this context, scaffolding largely takes the form of one-to-one scaffolding, rather than computer-based scaffolding (Stone, 1998); studies on one-to-one scaffolding among students with learning disabilities did not meet the inclusion criteria, and thus were excluded. It is important to conduct studies on computer-based scaffolding among students with special needs to explore whether this tool is promising for students with special needs.

Design-Based Learning

The effect size estimate for design-based learning was not significantly greater than zero, although caution is warranted due to the small sample size. Design-based learning has been posited as an approach that can facilitate the integration of science and engineering in education (Doppelt, Mehalik, Schunn, Silk, & Krysinski, 2008; Kolodner et al., 2003). The Next Generation Science Standards encourages the integration of science and engineering in education, as well as the use of authentic problems in school (National Science Board, 2010; Next Generation Science Standards, 2013). Further research on scaffolding in the context of design-based learning is needed so as to have a more precise effect size estimate and to learn what scaffolding elements lead to the strongest outcomes when used with this instructional model.

Project-Based Learning

The effect size estimate of project-based learning was derived from three outcomes, which warrant caution. Yet it was statistically higher than all other contexts of use. Future research should investigate if the effect size estimate remains consistent as more empirical research is added.

Motivation Scaffolding

Much recent research has highlighted the role of socioemotional support in advancing student learning outcomes (Belland et al., 2013; Perkins & Salomon, 2012). Few outcomes of motivation scaffolding met our inclusion criteria, most notably that the outcomes be cognitive, which caused the corresponding confidence interval to overlap with zero. Further efforts to measure cognitive outcomes from motivation scaffolding should cause the confidence interval to narrow.

Limitations and Suggestions for Future Research

Meta-analyses are a good way to synthesize results from quantitative research on a topic, but they cannot include the results of all empirical research (Cooper, Hedges, & Valentine, 2009). There were many studies on computer-based scaffolding that were either qualitative, or were quantitative but did not employ control groups, and thus needed to be eliminated from consideration in this meta-analysis. Thus, our effect size estimates do not reflect all empirical research on computer-based scaffolding. However, large amounts of quantitative work of the type that can be included in meta-analyses typically emerge once a research area matures. That we were able to include 144 studies despite the rigorous application of our inclusion and exclusion criteria suggests that computer-based scaffolding in STEM education is a mature research area. Thus, meta-analyses can help identify important trends in the literature and suggest avenues for future research.

Although our coding scheme was robust, it could not reflect perfectly every construct of interest in the scaffolding literature. For example, many intelligent tutoring systems incorporate performance-based fading and self-selected hints (Koedinger & Aleven, 2007). The coding scheme was set up to assign a single value for scaffolding logic. In these studies, we deemed scaffolding adjustment to be performance-based using the rationale that performance-based fading was always provided, whereas students may choose not to self-select hints. In future studies, it may be useful to identify the logic separately for each type of scaffolding adjustment. This type of coding would allow for a closer depiction of the nature of scaffolding interventions as well as how combinations of different scaffolding adjustment methods influence learning although it would require more complicated analyses and introduce additional dependency issues. Common to all meta-analyses is the issue of what actually occurred as opposed to what is described in the publication. Coding levels like "none" for scaffolding change, "traditional" for research populations, or "problem solving" for context of use indicate a lack of description about alternative options as much as a positive identification of a study feature.

The inclusion of such a wide variety of literature could be seen as a limitation. For example, scaffolding in intelligent tutoring systems and scaffolding based in knowledge integration and activity theory utilize different strategies and are grounded in different assumptions about learning. However, we only included studies in which students engaged with ill-structured problems and in which the scaffolding intervention was used to extend and enhance student capabilities to allow them to address the problems. Thus, much of the scaffolding literature was not included, such as studies that investigated the influence of interventions that did not require students to engage with ill-structured problems. Furthermore, if the intervention was provided before engagement with the problem, or was otherwise not consistent with our definition for scaffolding, the study was excluded.

In this sense, included studies were all similar. Moreover, by including a wide swath of literature, we were able to include a large variation of different scaffolding strategies and provide a comprehensive overview of computer-based scaffolding and scaffolding contexts associated with the strongest learning outcomes. Future syntheses as well as primary research studies might investigate interactions and dependencies between promising variables, taking care that such investigations are theoretically driven.

Scaffolding has led to promising results in subject areas outside of STEM, including social studies (Brush & Saye, 2001; Nussbaum, 2002; Saye & Brush, 2002), language arts (Proctor, Dalton, & Grisham, 2007), and teacher education (Chua, Tan, & Liu, 2015). As noted previously, scaffolding is only part of the intelligent tutoring system approach. A meta-analysis indicated that intelligent tutoring systems led to an average effect of $\bar{g}$ = 0.34 in language and literacy and $\bar{g}$ = 0.63 in humanities and social science (Ma et al., 2014). Another meta-analysis calculated a point estimate for intelligent tutoring systems in college-level business education of $\bar{g}$ = 0.16 (Steenbergen-Hu & Cooper, 2014). The average effect of scaffolding in STEM education ($\bar{g}$ = 0.46) calculated in this review is near the midpoint of the calculated effect size estimates for intelligent tutoring systems outside of STEM. Including studies from outside of STEM was outside the scope of this review, but future meta-analyses should synthesize work on a broader range of scaffolding outside of STEM education, as many different types of learning are critical to the generation of a well-educated, productive, and civic-minded citizenry (Guyotte, Sochacka, Costantino, Walther, & Kellam, 2014; Stearns, 1994).

Conclusion

This meta-analysis indicates that computer-based scaffolding in STEM disciplines is highly efficacious, leading to an average effect size of $\bar{g}$ = 0.46. Strong outcomes were consistent across a wide range of learner populations, contexts of use, and scaffolding characteristics. But this study also addresses many persistent questions in the scaffolding literature. Notably, we found that there was no difference in effect sizes (a) among scaffolding with fading, adding, fading/adding, or no fading or adding; (b) on the basis of scaffold customization logic; and (c) on the basis of context specificity. Furthermore, we found that scaffolding influences cognitive outcomes at the concept, principles, and application levels. Scaffolding has expanded considerably in terms of learner population and targeted learner outcome, leading to the strongest cognitive outcomes among the learner population furthest from the original scaffolding learner population (i.e., adults), as opposed to the original population of preschoolers.

Note

This research was supported by the National Science Foundation under REESE Grant 1251782. Any opinions, findings, or conclusions are those of the authors and do not necessarily represent official positions of the Foundation.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, *85*, 275–314. doi:10.3102/0034654314551063

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A Stage 1 meta-analysis. *Review of Educational Research*, *78*, 1102–1134. doi:10.3102/0034654308326084

Albanese, M. (2000). Problem-based learning: Why curricula are likely to show little effect on knowledge and clinical skills. *Medical Education*, *34*, 729–738. doi:10.1046/j.1365-2923.2000.00753.x

Alexander, P. A. (2014). Thinking critically and analytically about critical-analytic thinking: An introduction. *Educational Psychology Review*, *26*, 469–476. doi:10.1007/s10648-014-9283-1

Alias, N. A. (2012). Design of a motivational scaffold for the Malaysian e-Learning environment. *Journal of Educational Technology & Society*, *15*, 137–151.

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

Aydin, E., & Cagiltay, N. (2012). A new RF and microwave engineering course enriched with advanced technologies. *Computer Applications in Engineering Education*, *20*, 634–645. doi:10.1002/cae.20432

Barab, S. A., Squire, K. D., & Dueber, W. (2000). A co-evolutionary model for supporting the emergence of authenticity. *Educational Technology Research & Development*, *48*(2), 37–62. doi:10.1007/BF02313400

Belland, B. R. (2011). Distributed cognition as a lens to understand the effects of scaffolds: The role of transfer of responsibility. *Educational Psychology Review*, *23*, 577–600. doi:10.1007/s10648-011-9176-5

Belland, B. R., French, B. F., & Ertmer, P. A. (2009). Validity and problem-based learning research: A review of instruments used to assess intended learning outcomes. *Interdisciplinary Journal of Problem-Based Learning*, *3*, 59–89. doi:10.7771/1541-5015.1059

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

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., . . . Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, *74*, 379–439. doi:10.3102/00346543074003379

Biemiller, A., & Meichenbaum, D. (1998). The consequences of negative scaffolding for students who learn slowly: A commentary on C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities." *Journal of Learning Disabilities*, *31*, 365–369. doi:10.1177/002221949803100405

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (1st ed.). Hoboken, NJ: Wiley.

Brush, T., & Saye, J. (2001). The use of embedded scaffolds with hypermedia-supported student-centered learning. *Journal of Educational Multimedia and Hypermedia*, *10*, 333–356.

Chang, K. E., Sung, Y. T., & Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, *17*, 21–33. doi:10.1111/j.1365-2729.2001.00156.x

Chen, H.-H., Chen, Y.-J., & Chen, K.-J. (2013). The design and effect of a scaffolded concept mapping strategy on learning performance in an undergraduate database course. *IEEE Transactions on Education*, *56*, 300–307. doi:10.1109/TE.2012.2217747

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. doi:10.1016/j.edurev.2013.01.001

Chua, B. L., Tan, O. S., & Liu, W. C. (2015). Using technology to scaffold problem-based learning in teacher education: Its tensions and implications for educational leaders. In C. Koh (Ed.), *Motivation, leadership and curriculum design* (pp. 119–135). Singapore: Springer.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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: Erlbaum.

Conati, C., & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 389–415.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *Journal of the Learning Sciences*, *12*, 91–142. doi:10.1207/S15327809JLS1201_4

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

Deters, K. M. (2009). *Investigating a computerized scaffolding software for student designed science investigations* (Doctoral dissertation). Avaliable from ProQuest Dissertations & Theses database. (UMI No. 304942256)

Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. Dillenbourg & G. Kanselaar (Eds.), *Three worlds of CSCL. Can we support CSCL?* (pp. 61–91). Heerlen, Netherlands: Open Universiteit Nederland.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction*, *13*, 533–568. doi:10.1016/S0959-4752(02)00025-7

Doppelt, Y., Mehalik, M. M., Schunn, C. D., Silk, E., & Krysinski, D. (2008). Engagement and achievements: A case study of design-based learning in a science context. *Journal of Technology Education*, *19*(2), 21–38.

DuBois, D. L., Holloway, B. E., Valentine, J. C., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology*, *30*, 157–197. doi:10.1023/A:1014628810714

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*, 917–928. doi:10.1093/jpepsy/jsp004

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. doi:10.1111/j.0006-341X.2000.00455.x

Egger, M., Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629–634. doi:10.1136/bmj.315.7109.629

Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, *18*(3), 4–10. doi:10.3102/0013189X018003004

Fund, Z. (2007). The effects of scaffolded computerized science problem-solving on achievement outcomes: A comparative study of support programs. *Journal of Computer Assisted Learning*, *23*, 410–424. doi:10.1111/j.1365-2729.2007.00226.x

Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research & Development*, *51*(1), 21–38. doi:10.1007/bf02504515

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

González-Calero, J. A., Arnau, D., Puig, L., & Arevalillo-Herráez, M. (2015). Intensive scaffolding in an intelligent tutoring system for the learning of algebraic word problem solving. *British Journal of Educational Technology*, *46*, 1189–1200. doi:10.1111/bjet.12183

Guyotte, K. W., Sochacka, N. W., Costantino, T. E., Walther, J., & Kellam, N. N. (2014). STEAM as social practice: Cultivating creativity in transdisciplinary spaces. *Art Education*, *67*(6), 12–19.

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: Erlbaum.

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

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39–65. doi:10.1002/jrsm.5

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. E. (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. E., Duncan, R. G., & Chinn, C. A. (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

Hong, Z.-R., Lin, H., Wang, H.-H., Chen, H.-T., & Yang, K.-K. (2013). Promoting and scaffolding elementary school students' attitudes toward science and argumentation through a science and society intervention. *International Journal of Science Education*, *35*, 1625–1648. doi:10.1080/09500693.2012.734935

Hung, D., & Chen, D. (2007). Context-process authenticity in learning: implications for identity enculturation and boundary crossing. *Educational Technology Research & Development*, *55*(2), 147–167. doi:10.1007/s11423-006-9008-3

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. (2000). Toward a design theory of problem solving. *Educational Technology Research & Development*, *48*(4), 63–85. doi:10.1007/BF02300500

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

Katai, Z. (2011). Multi-sensory method for teaching-learning recursion. *Computer Applications in Engineering Education*, *19*, 234–243. doi:10.1002/cae.20305

Kettler, T. (2014). Critical thinking skills among elementary school students comparing identified gifted and general education student performance. *Gifted Child Quarterly*, *58*, 127–136. doi:10.1177/0016986214522508

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., & Hannafin, M. (2011). Scaffolding 6th graders' problem solving in technology-enhanced science classrooms: A qualitative case study. *Instructional Science*, *39*, 255–282. doi:10.1007/s11251-010-9127-4

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

Knowles, M. S. (1984). *Andragogy in action. Applying modern principles of adult education*. San Francisco, CA: Jossey-Bass.

Koedinger, K., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239–264. doi:10.1007/s10648-007-9049-0

Kolodner, J. L., Camp, P. J., 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(tm) into practice. *Journal of the Learning Sciences*, *12*, 495–547. doi:10.1207/S15327809JLS1204_2

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

Kramarski, B., & Gutman, M. (2006). How can self-regulated learning be supported in mathematical E-learning environments? *Journal of Computer Assisted Learning*, *22*, 24–33. doi:10.1111/j.1365-2729.2006.00157.x

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, *76*, 275–313. doi:10.3102/00346543076002275

Leont'ev, A. N. (1974). The problem of activity in psychology. *Soviet Psychology*, *13*(2), 4–33. doi:10.2753/RPO1061-040513024

Lesh, R., & Harel, G. (2003). Problem solving, modeling, and local conceptual development. *Mathematical Thinking and Learning*, *5*, 157–189. doi:10.1080/10986065.2003.9679998

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

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, *36*, 837–858. doi:10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U

Linn, M. C. (2000). Designing the knowledge integration environment. *International Journal of Science Education*, *22*, 781–796. doi:10.1080/095006900412275

Linn, M. C., & Eylon, B.-S. (2000). Knowledge integration and displaced volume. *Journal of Science Education and Technology*, *9*, 287–310. doi:10.1023/A:1009451808539

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., . . . Busick, M. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Retrieved from http://eric.ed.gov/?id=ED537446

Looi, C.-K., & Lim, K.-S. (2009). From bar diagrams to letter-symbolic algebra: A technology-enabled bridging. *Journal of Computer Assisted Learning*, *25*, 358–374. doi:10.1111/jca.2009.25.issue-410.1111/j.1365-2729.2009.00313.x

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations* (M. Cole, Ed.; M. Lopez-Morillas & L. Solotaroff, Trans.). Cambridge, MA: Harvard University Press.

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

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

McNeill, K. L., & Krajcik, J. (2009). Synergy between teacher practices and curricular scaffolds to support students in using domain-specific and domain-general knowledge in writing arguments to explain phenomena. *Journal of the Learning Sciences*, *18*, 416–460. doi:10.1080/10508400903013488

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*, 153–191. doi:10.1207/s15327809jls1502_1

Moos, D. C., & Azevedo, R. (2008). Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior*, *24*, 1686–1706. doi:10.1016/j.chb.2007.07.001

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*, 117–128. doi:10.1037/0022-0663.97.1.117

Murphy, P. K., Rowe, M. L., Ramani, G., & Silverman, R. (2014). Promoting critical-analytic thinking in children and adolescents at home and in school. *Educational Psychology Review*, *26*, 561–578. doi:10.1007/s10648-014-9281-3

National Research Council. (2011). *Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics*. Washington, DC: National Academies Press.

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

National Science Board. (2010). *Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital*. Arlington, VA: National Science Foundation.

Next Generation Science Standards. (2013). Retrieved from http://www.nextgenscience.org/next-generation-science-standards

Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, *9*, 114–128. doi:10.1016/j.edurev.2012.12.002

Nussbaum, E. M. (2002). Scaffolding argumentation in the social studies classroom. *Social Studies*, *93*, 79–83. doi:10.1080/00377990209599887

Osman, K., & Lee, T. T. (2013). Impact of interactive multimedia module with pedagogical agents on students' understanding and motivation in the learning of electrochemistry. *International Journal of Science and Mathematics Education*, *12*, 395–421. doi:10.1007/s10763-013-9407-y

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

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology*, *12*, 73–87. doi:10.1207/S15327876MP1201_4

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, *13*, 423–451. doi:10.1207/s15327809jls1303_6

Perkins, D., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, *18*, 16–25. doi:10.3102/0013189X018001016

Perkins, D., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist*, *47*, 248–258. doi:10.1080/00461520.2012.693354

Piaget, J. (1947). *Le jugement et le raisonnement chez* l'enfant [Judgment and reasoning in the child]. Paris, France: Delachaux & Niestlé.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39*, 71–93. doi:10.1080/10862960709336758

Puntambekar, S., & Hubscher, 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

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*, 235–244. doi:10.1207/s15326985ep4004_5

Raes, A., Schellens, T., De Wever, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education*, *59*, 82–94. doi:10.1016/j.compedu.2011.11.010

Reid, D. J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, *19*, 9–20. doi:10.1046/j.0266-4909.2003.00002.x

Reiser, B. J. (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

Rieber, L. P., Tzeng, S.-C., & Tribble, K. (2004). Discovery learning, representation, and explanation within a computer-based simulation: Finding the right mix. *Learning and Instruction*, *14*, 307–323. doi:10.1016/j.learninstruc.2004.06.008

Ross, J. A., & Bruce, C. D. (2009). Student achievement effects of technology-supported remediation of understanding of fractions. *International Journal of Mathematical Education in Science and Technology*, *40*, 713–727. doi:10.1080/00207390902971999

Saye, J., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research & Development*, *50*(3), 77–96. doi:10.1007/BF02505026

Schrader, C., & Bastiaens, T. (2012). Learning in educational computer games for novices: The impact of support provision types on virtual presence, cognitive load, and learning outcomes. *International Review of Research in Open and Distributed Learning*, *13*, 206–227.

Scruggs, T. E., Brigham, F. J., & Mastropieri, M. A. (2013). Common Core Science Standards: Implications for students with learning disabilities. *Learning Disabilities Research & Practice*, *28*, 49–57. doi:10.1111/ldrp.12002

Shadish, W. R., & Myers, D. (2004). *Campbell collaboration research design policy brief*. Oslo, Norway: Campbell Collaboration Methods.

Siegel, M. A. (2006). High school students' decision making about sustainability. *Environmental Education Research*, *12*, 201–215. doi:10.1080/13504620600689003

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

Stage, E. K., Asturias, H., Cheuk, T., Daro, P. A., & Hampton, S. B. (2013). Opportunities and challenges in Next Generation Standards. *Science*, *340*, 276–277. doi:10.1126/science.1234011

Stark, D. M. (2013). *Ill-structured problems, scaffolding and problem-solving ability of novice nursing students* (Doctoral dissertation). Available from ProQuest Dissertations & Theses database. (UMI No. 3553778)

Stearns, P. N. (1994). *Meaning over memory: Recasting the teaching of culture and history*. Chapel Hill: University of North Carolina Press.

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

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

Su, Y., & Klein, J. (2010). Using scaffolds in problem-based hypermedia. *Journal of Educational Multimedia and Hypermedia*, *19*, 327–347.

Sugrue, B. (1995). A theory-based framework for assessing domain-specific problem-solving ability. *Educational Measurement: Issues and Practice*, *14*(3), 29–35. doi:10.1111/j.1745-3992.1995.tb00865.x

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

Tamim, R., Bernard, R., Borokhovski, E., Abrami, P., & Schmid, R. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*, *81*, 4–28. doi:10.3102/0034654310393361

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.

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

Ulicsak, M. H. (2004). "How did it know we weren't talking?" An investigation into the impact of self-assessments and feedback in a group activity. *Journal of Computer Assisted Learning*, *20*, 205–211. doi:10.1111/j.1365-2729.2004.00083.x

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*, 473–481. doi:10.1037/0022-0167.51.4.473

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., & 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

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

Zhang, J., Chen, Q., & Reid, D. J. (2000). Simulation-based scientific discovery learning: a research on the effects of experimental support and learners' reasoning ability. In *Proceedings IFIP World Computer Congress, Educational Uses of Communication and Information Technologies* (pp. 344–351). *Beijing, China: ICEUT.*

Zhang, J., Chen, Q., Sun, Y., & Reid, D. J. (2004). Triple scheme of learning support design for scientific discovery learning based on computer simulation: Experimental research. *Journal of Computer Assisted Learning*, *20*, 269–282. doi:10.1111/j.1365-2729.2004.00062.x

Zydney, J. M. (2008). Cognitive tools for scaffolding students defining an ill-structured problem. *Journal of Educational Computing Research*, *38*, 353–385. doi:10.2190/EC.38.4.a

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; 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; 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 is a PhD student in the Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322; 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.

MASON LEFLER is a former middle school teacher and current PhD student in the Department of Instructional Technology and Learning Sciences, Utah State University, 2830 Old Main Hill, Logan, UT 84322; email: *masonlefler@hotmail.com*. His research interests include computerized learning environments, computer-based scaffolding, problem-centered instruction, formative assessment, and data analytics in the classroom.