

Effectiveness of Computer-Based Scaffolding in the Context of Problem-Based Learning for Stem Education: Bayesian Meta-analysis

Nam Ju Kim¹  · Brian R. Belland² · Andrew E. Walker²

Published online: 29 July 2017

© The Author(s) 2017. This article is an open access publication

Abstract Computer-based scaffolding plays a pivotal role in improving students' higher-order skills in the context of problem-based learning for Science, Technology, Engineering and Mathematics (STEM) education. The effectiveness of computer-based scaffolding has been demonstrated through traditional meta-analyses. However, traditional meta-analyses suffer from small-study effects and a lack of studies covering certain characteristics. This research investigates the effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education through Bayesian meta-analysis (BMA). Specifically, several types of prior distribution information inform Bayesian simulations of studies, and this generates accurate effect size estimates of six moderators (total 24 subcategories) related to the characteristics of computer-based scaffolding and the context of scaffolding utilization. The results of BMA indicated that computer-based scaffolding significantly impacted ($g = 0.385$) cognitive outcomes in problem-based learning in STEM education. Moreover, according to the characteristics and the context of use of scaffolding, the effects of computer-based scaffolding varied with a range of small to moderate values. The result of the BMA contributes to an enhanced understanding of the effect of computer-based scaffolding within problem-based learning.

✉ Nam Ju Kim
namju1001@gmail.com

Brian R. Belland
brian.belland@usu.edu

Andrew E. Walker
andy.walker@usu.edu

¹ Department of Teaching and Learning, University of Miami, Coral Gables, FL, USA

² Department of Instructional Technology and Learning Sciences, Utah State University, Logan, UT, USA

Keywords Computer-based scaffolding · Problem-based learning · Intelligent tutoring systems · STEM education · Bayesian meta-analysis

Introduction

The Next Generation Science Standards promote the use of problem-based learning (PBL), which requires that students construct knowledge to generate solutions to ill-structured, authentic problems (Achieve 2013). Central to student success in such approaches is scaffolding—dynamic support that helps students meaningfully participate in and gain skill at tasks that are beyond their unassisted capabilities (Belland 2014; Hmelo-Silver et al. 2007). When originally defined, instructional scaffolding was delivered in a one-to-one manner by a teacher (Wood et al. 1976). But researchers have considered how to use computer-based tools as scaffolding to overcome the limitations of one-to-one scaffolding such as high student to teacher ratios (Hawkins and Pea 1987). Computer-based scaffolding has been utilized in the context of PBL in Science, Technology, Engineering, Mathematics (STEM) education, and many studies have demonstrated the effect of computer-based scaffolding on students' conceptual knowledge and higher-order skills. However, it is difficult to generalize from the results of individual studies without the use of systematic synthesis methods (e.g., meta-analysis) due to different educational populations in particular contexts. For that reason, several meta-analyses have addressed the effectiveness of computer-based scaffolding (Azevedo and Bernard 1995; Belland et al. 2017; Ma et al. 2014), but none focused specifically on scaffolding in the context of PBL.

This study aims to determine and generalize the effectiveness of computer-based scaffolding utilized in PBL in terms of several characteristics of computer-based scaffolding and its contexts through meta-analysis. In addition, a Bayesian approach was used to address some issues of traditional meta-analysis such as biased results due to publication bias and low statistical power due to a lack of studies included in meta-analysis.

Computer-Based Scaffolding in Problem-Based Learning

PBL is a learner-centered instructional approach that aims to improve students' content knowledge and problem-solving skills through engagement with authentic and ill-structured problems, which have no single correct answer (Hmelo-Silver 2004; Kolodner et al. 2003; Thistlethwaite et al. 2012). Students acquire new knowledge by identifying knowledge gaps between their current level of knowledge and the level of knowledge it would take to address the given problem (Barrows 1996). To accomplish PBL tasks, students need diverse skills including advanced problem-solving skills, critical thinking, and collaborative learning skills (Gallagher et al. 1995). However, students who are new to PBL can struggle due to different levels of background knowledge, learning skills, and motivation. Scaffolding has been utilized to make the tasks in PBL more manageable and accessible (Hmelo-Silver et al. 2007) and to help students improve deep content knowledge and higher-order thinking skills (Belland et al. 2011). Computer-based scaffolding, especially, has had positive impacts on students' cognitive learning outcomes. For example, students can be invited to consider the complexity that is integral to the target skill and spared the burden of addressing complexity that is not (Reiser 2004) through computer-based hints (Leemkuil and de Jong 2012; Li 2001; Schrader and Bastiaens 2012), visualization (Cuevas et al. 2002; Kumar 2005; Linn et al. 2006), question

prompts (Hmelo-Silver and Day 1999; Kramarski and Gutman 2006), and concept mapping (Puntambekar et al. 2003). In addition, computer-based scaffolding can improve students' interest and motivation toward their learning (Clarebout and Elen 2006).

Meta-analyses Related to Computer-Based Scaffolding

Computer-based scaffolding has been found to positively impact many variables across many studies. However, generalizing such results is difficult because the learning environment, population, and experimental condition vary across the studies. Therefore, some scholars tried to combine the results and synthesize information from multiple individual studies through meta-analysis. Belland et al. (2017) conducted a traditional meta-analysis to determine the influence of computer-based scaffolding in the context of problem-centered instructions for STEM education. The overall effect size of scaffolding was $g = 0.46$ (Belland et al. 2017). This result indicated that computer-based scaffolding can help students learn effectively in problem-centered instruction. However, the previous meta-analysis (Belland et al. 2017) did not specify the effects of scaffolding according to scaffolding and students characteristics within the context of problem-based scaffolding. In addition, in the case of conventional instruction, as opposed to problem-centered instructional models, meta-analysis indicated that computer-based scaffolding including intelligent tutoring systems positively impacted students' learning ($g = 0.66$) regardless of instructor's effects, study types, and region (Kulik and Fletcher 2016). Other meta-analyses on the effectiveness of intelligent tutoring systems showed a wide range of effect sizes: $g = 0.41$ among college students (Steenbergen-Hu and Cooper 2014), $g = 0.09$ for K-12 students' mathematical learning (Steenbergen-Hu and Cooper 2013), $d = 0.76$ when compared the effectiveness with human-tutors (VanLehn 2011), and $d = 1.00$ for scaffolding within an early model of intelligent tutoring systems (Anderson et al. 1995). However, no meta-analyses have investigated the effectiveness of computer-based scaffolding in the context of problem-based learning.

Issues of Traditional Meta-analysis

Meta-analyses offer much more in the way of systematicity than traditional research reviews, but some scholars (Biondi-Zoccai et al. 2012; Greco et al. 2013; Koricheva et al. 2013) noted two potential pitfalls that can occur during meta-analysis. One issue is related to small-study effects, which can result in publication bias. In meta-analyses, the effect size is estimated based on observations reported in previous studies. For meta-analyses, these observations should be standardized, but multiple errors can occur because of the size of studies involved. Small studies tend to report larger effect sizes than large studies, and the effect sizes from studies with small sample sizes (i.e., $n < 10$) can be biased (Hedges 1986). For this reason, some scholars concluded that if both small-study effects and large-study effects are included in one data set, researchers should focus on analyzing only large-study effects (Biondi-Zoccai et al. 2012; Greco et al. 2013; Hedges 1986; Koricheva et al. 2013). However, in the case of educational research, large sample sizes are rare.

Another issue is that in traditional meta-analysis, there is no method to include a level of a variable if the number of studies coded at that level is too small. There has been debate on the minimum number of studies that should be included in meta-analysis (Guolo and Varin 2015). In theory, one can conduct meta-analysis with just two studies (Valentine et al. 2010), but in this case, statistical power is largely reduced. In one study that investigated the difference of

statistical power in meta-analysis according to the included number of studies, the range of power in meta-analysis with 10 studies was not beyond 0.2 (random effect model with large heterogeneity), 0.4 (random effect with small heterogeneity), and 0.5 (fixed effect model) (Valentine et al. 2010). However, in educational research, especially PBL, it can be difficult to find a large number of studies on a given topic that all meet inclusion criteria, due to different education populations and levels, learning environments, and outcomes, which in turn make it difficult to arrive at reliable and validated results (Ahn et al. 2012).

Considering the abovementioned issues of meta-analysis, it is worth considering an alternative methodology to address publication bias resulting from small-study effects and a small number of included studies. This study suggests a Bayesian approach to conduct meta-analysis for determining the effect size of computer-based scaffolding in the context of problem-based learning. The explanation of Bayesian meta-analysis will be specified in the next section.

Bayesian Meta-analysis

To address the identified limitation of traditional meta-analysis approaches, one can use a Bayesian approach, which assumes that all parameters come from a superpopulation with its own parameters (Hartung et al. 2011; Higgins et al. 2009). Superpopulation refers to an infinite population of abstractions that has unique characteristics of the population, and the finite population itself is considered to be the sample of superpopulation (Royall 1970). That is, in sample-based inference, the expected value can be obtained from all possible samples of the given finite population. On the other hand, the expected value from superpopulation model-based inference is derived from all possible samples in an infinite population, which means that the number of the units constituting the population is infinite. Under this assumption of superpopulation, the Bayesian approach relies on (a) generating a prior distribution ($\rho(\cdot)$) utilizing data from pre-collected studies that should not be included in Bayesian meta-analysis, (b) estimating the likelihood that the prior distribution is valid based on the observed data ($\rho(\text{data}|\cdot)$), and (c) generating a posterior distribution, which can be calculated by the Bayesian law of probability ($\rho(\cdot|\text{data})$). In short, this approach can provide a more accurate estimate of the treatment effect by adding another component of variability—the prior distribution (Schmid and Mengersen 2013). Prior distribution is defined as a distribution that articulates researchers' beliefs or the results from previous studies about parameters prior to the collection of new data (Raudenbush and Bryk 2002). Prior distributions play a role in summarizing the evidence and determining evidential uncertainty (Spiegelhalter et al. 2004). In the Bayesian model, μ (weighted mean effect size), τ (standard deviation of the between-study variance), and β (study-level covariates) of prior distributions can be important factors in estimating prior distributions (Findley 2011; Sutton and Abrams 2001).

Typically, the prior distribution can be divided into two types (i.e., informative and non-informative prior distribution) according to whether prior information (i.e., μ , β , and τ) about the topic exists or not. When prior information is not available, the Bayesian approach commonly uses non-informative prior distribution. The posterior distribution can be different according to how the between-study variance (τ^2) in the prior distribution is set up (see Table 1). It is important to consider all possible τ^2 (i.e., maximum values of τ^2 and minimum values of τ^2) in non-informative prior distribution.

As seen in Table 1, the purpose of reference prior distribution is to maximize divergence between the variances of studies (i.e., τ^2), and it can generate maximum effects of newly added

Table 1 Reference prior distributions for τ^2 (Spiegelhalter et al. 2004)

Reference	Distribution for τ^2	Characteristics
DuMouchel prior for τ^2	$\tau_0^2 = \frac{S_0}{9}$ and $\tau_0^2 = 9S_0$	<ul style="list-style-type: none"> • A highly dispersed prior distribution for τ^2 • Values of τ^2 to be near 0 • Protection of highly skew toward large values of τ^2
Uniform prior for τ^2	$\tau^2 \sim \text{Uniform}(0, 1000)$	<ul style="list-style-type: none"> • Large value of τ^2
Inverse gamma prior on τ^2	$\frac{1}{\tau^2} \sim \text{Gamma}(.001, .001)$	<ul style="list-style-type: none"> • Greater weight to values of τ near 0

data on the posterior distribution (Sun and Ye 1995). After assuming all possible τ^2 across studies using the above reference prior distributions, one can identify the most suitable prior distribution model using deviance information criteria (Spiegelhalter et al. 2002).

Research Questions

The purpose of this study is to determine the effect of computer-based scaffolding in the context of PBL by addressing the following research questions.

- 1) How does computer-based scaffolding affect students' cognitive learning outcomes in the context of problem-based learning for STEM education?
- 2) How does the effectiveness of computer-based scaffolding vary according to scaffolding intervention?
- 3) How does the effectiveness of computer-based scaffolding vary according to scaffolding customization and its methods?
- 4) How does the effectiveness of computer-based scaffolding vary according to the higher-order skill it is intended to enhance?
- 5) How does the effectiveness of computer-based scaffolding vary according to scaffolding strategies?
- 6) How does the effectiveness of computer-based scaffolding vary according to discipline?

Method

Search Process

The initial databases we used to search for articles were Education Source, PsycInfo, Digital dissertation, CiteSeer, ERIC, PubMed, Academic Search Premier, IEEE, and Google scholar. Those databases were recommended by a librarian and experts who have expertise in the fields related to this study. However, some articles were duplicated across databases or we could not find any article that satisfied our inclusion criteria from the following databases: PubMed, Academic Search Premier, and IEEE. Various combinations of the following search terms were used in the databases listed above: “scaffold, scaffolds, computer-based scaffolding/supports,” “Problem-based learning,” “cognitive tutor,” “intelligent tutoring systems,” “Science, Technology, Engineering, Mathematics,” and subcategories of higher-order thinking skills. The search terms were determined by researchers' consensus and advisory board members' advice.

Inclusion Criteria

The following inclusion criteria were used: studies needed to (a) be published from January 1, 1990, and December 31, 2015; (b) present sufficient information to conduct Bayesian meta-analysis (statistical results revealing the difference between treatments and control group, number of participants, study design); (c) be conducted in the context of problem-based learning within STEM education; (d) clearly reveal which types of scaffolding they used; and (e) address higher-order thinking skills as the intended outcome of the scaffold itself. We found a total of 21 studies with 47 outcomes (see [Appendix](#)). The numbers of studies and outcomes differ because some studies had multiple outcomes with different levels of the moderators, which were used in this meta-analysis (i.e., scaffolding intervention, higher-order thinking skills, scaffolding customization and its method, scaffolding strategy, and disciplines).

Moderators for meta-analysis

Scaffolding Intervention Conceptual scaffolding provides expert hints, concept mapping and/or tools to engage in concept mapping, and visualization depicting concepts to help them identify what to consider when solving the problem (Hannafin et al. 1999). For example, scaffolding in Su (2007) was designed to focus attention on key content and to stay organized in a way to achieve project requirements. Strategic scaffolding helps students identify, find, evaluate information for problem-solving, and guide a suitable approach to solve the problems (Hannafin et al. 1999). An example can be seen in the scaffolding in Rosen and Tager (2014), which enabled students to construct a well-integrated structural representation (e.g., about the benefits of organic milk). Conceptual scaffolding can be distinguished from strategic scaffolding in that conceptual scaffolding helps students consider tasks from different angles through the reorganization and connection of evidence; on the other hand, strategic scaffolding tells students how to use the evidence for problem-solving (Saye and Brush 2002). Metacognitive scaffolding allows students to reflect on their learning process and encourages students to consider possible problem solutions (Hannafin et al. 1999). For example, the reflection sheet in Su and Klein (2010) encouraged students to summarize what they had learned, reflect upon it, and then debrief that information. Motivational scaffolding aims to enhance students' interest, confidence, and collaboration (Jonassen 1999a; Tuckman and Schouwenburg 2004).

Higher-Order Thinking Skills Scaffolding is often designed to enhance higher-order thinking skills (Aleven and Koedinger 2002; Azevedo 2005; Quintana et al. 2005). The definition of higher-order thinking and its subcategories differ according to different scholars. Higher-order thinking can be defined as “challenge and expanded use of the mind” (Newmann 1991, p. 325) and students can enhance higher-order thinking skills through active participation in such activities as making hypotheses, gathering evidence, and generating arguments (Lewis and Smith 1993). According to Bloom's taxonomy (Bloom 1956), higher-order thinking is the stage beyond understanding and declarative knowledge. Therefore, analyzing, synthesizing, and evaluating can be classified as higher-order skills. Analysis means the ability to identify the components of information and ideas, and to establish the relations between elements (Lord and Baviskar 2007). For example, scaffolding in *Bird Watching Learning* (Chen et al. 2003) provided pictures, questions, and other information that learners can use to identify bird species. Synthesis refers to recognition of the patterns of components and the formation of a new whole through creativity. Through this ability learners can formulate a hypothesis or propose alternatives

(Anderson et al. 2001). The mapping software in Toth et al. (2002) can help students formulate scientific statements using hypotheses and data with a summary of information found in web-based materials. Defined as the ability to judge the value of material based on definite criteria, evaluation allows learners to judge the value of data and experimental results and justify conclusions (Krathwohl 2002). For example, several strategies of scaffolding (question prompts, expert advice, and suggestion) in Simons and Klein (2007) provided learners support to rate the reliability and believability of each evidence item and the extent to which they believe the statements that they generated. Based on the above illustrative phrases, critical thinking and logical thinking can be combined to form the critical thinking category of “Analysis” and creative thinking and reflective thinking can be combined to form the critical thinking category of “Synthesis,” and problem-solving skills and decision-making can be combined to form the critical thinking category of “Evaluation” (Bloom 1956; Hershkowitz et al. 2001).

Therefore, in this study, I defined higher-order thinking skills as those cognitive skills that allow students to function at the analysis, synthesis, and evaluation levels of Bloom’s Taxonomy, and in this study, the categorization for intended outcomes can be analysis, synthesis, and evaluation as the variation of higher-order skills.

Scaffolding Customization and Its Methods By effectively controlling the timing and degree of scaffolding, students reach the final learning goal by their own learning strategies and processes (Collins et al. 1989). In this sense, scaffolding customization is defined as the change of scaffolding frequency and its nature based on a dynamic assessment of students’ current abilities (Belland 2014). There are three kinds of scaffolding customization (i.e., fading, adding, fading/adding supports). Fading means that scaffolds are introduced and then pulled away. As an example of fading, *Web-based Inquiry Science Environment* (Raes et al. 2012) faded scaffolding according to the level of students’ learning progress (e.g., full function of scaffolding in the beginning step, but no scaffolding in advanced steps). On the other hand, adding is defined as increasing frequency of scaffolds, reducing the interval of scaffolding, and adding new scaffolding elements as the intervention goes on. In Chang et al. (2001), if students continue to struggle excessively, greater quantities and intensities of scaffolding can be requested through the use of a hint button. Moreover, other type of scaffolding customization is fading/adding, which is defined as increasing or pulling away scaffolds depending on students’ current learning status and their requests. Scaffolding, which neither increased nor decreased regarding its nature or frequency, was categorized as none. According to the meta-analysis reported by Belland et al. (2017), in 65% of the included studies, there was no scaffolding customization. Moreover, Lin et al. (2012) also pointed out a lack of a number of studies adopting fading function (9.3%) in a review of 43 scaffolding-related articles. This means that while many scholars maintained that fading is an important element of scaffolding (Collins et al. 1989; Dillenbourg 2002; Puntambekar and Hübscher 2005; Wood et al. 1976), scaffolding customization has largely been overlooked in scaffolding design.

There are three ways to determine scaffolding fading, adding, and fading/adding: performance-adaptation, self-selection, and fixed time interval. Performance-adaptation means that the frequency and nature of scaffolding can be changed by students’ current learning performance and status. On the other hand, self-selection is defined as the scaffolding customization by students’ own decision to request fading, adding, and both of them. In intelligent tutoring systems that can monitor students’ ability, scaffolding fading is often performance-adapted and adding supports is self-selected. In addition to the bases of performance and self-selection, scaffolding customization can also be fixed, defined as adding or fading after a predetermined number of events or a fixed time interval has passed (Clark et al. 2012). Among the scaffolding

customization methods (i.e., performance-adapted, self-selected, and fixed), performance-adapted scaffolding customization was the most frequent (Belland et al. 2017), but there have been few studies that investigated which scaffolding customization method have the highest effects on students' learning performance in the context of problem-based learning.

Scaffolding Strategies Scaffolding strategies include feedback, question prompts, hints, and expert modeling (Belland 2014; Van de Pol et al. 2010). Feedback is the provision of information regarding the students' performance to the students (Belland 2014). In Siler et al. (2010), a computer tutor that covered experimental design evaluated students' designs and provided feedback about their selection of the variables of interest. Question prompts help students draw inferences from their evidence and encourage their elaborative learning (Ge and Land 2003). Students read question prompts that directed their attention to important problem elements and encouraged them to conduct certain tasks (Ge et al. 2010). Hints are clues or suggestions to help students go forward (Melero et al. 2011). For example, when students tried to change the paragraph text, computer systems showed word definitions and provided audio supports to read these words. Expert modeling presents how experts perform a given task (Pedersen and Liu 2002). In Simons and Klein (2007), when students struggled with balloon design, expert advice was provided to help them distinguish between valuable and useless information. In addition, several types of strategies can be used within one study to satisfy students' different needs according to the contexts (Dennen 2004; Gallimore and Tharp 1990). As an example of several types of strategies, in Butz et al. (2006), students received expert modeling, question prompts, feedback, and hints to solve a real-life problem in their introductory circuits class.

Discipline In this paper, "STEM" refers to two things: (a) the abbreviation of Science, Technology, Engineering, and Mathematics, in which scaffolding was utilized and (b) integrated STEM curricula. Integrated STEM education began with the aim of performance enhancement in science and mathematics education as well as the cultivation of engineers, scientists, and technicians (Kuenzi 2008; Sanders 2009). Application of integrated STEM education increased students' motivation and interests in science learning and contributed to positive attitudes toward a STEM-related area (Bybee 2010). For example, the results of two meta-analyses indicated that the integrative approaches on STEM disciplines showed higher effects ($d > 0.8$) on students' performance than in separate STEM disciplines (Becker and Park 2011; Lam et al. 2008). However, there are few studies investigating the effects of computer-based scaffolding, which has been commonly utilized in each STEM field, in the context of problem-based learning for integrated STEM education. Therefore, it may be worthwhile to investigate the comparison of scaffolding effects between integrated approach in STEM education and each STEM fields. In this regard, integrated STEM education and each STEM discipline (i.e., Science, Technology, Engineering, and Mathematics) are included as discipline moderator.

Table 2 shows the moderators and subcategories of each moderator in this meta-analysis.

Prior Distribution in This Study

In Bayesian analysis, the estimation of the posterior distribution can be substantially affected by how one can set up the prior distribution. Typically, there are three methods to determine the prior distribution. One method is to follow experts' opinions about parameter information related to a certain topic. Experts' opinions reflect the results of existing studies, and it is

Table 2 Moderators in Bayesian meta-analysis

Moderator	Higher-order skills	Scaffolding customization	Scaffolding customization method (SCM)
Subcategories	<ul style="list-style-type: none"> • Analysis • Synthesis • Evaluation 	<ul style="list-style-type: none"> • No fading/adding • Fading • Adding • Fading/adding 	<ul style="list-style-type: none"> • No SCM • Performance-adapted • Fixed time interval • Self-selected
Moderator	Scaffolding strategies	Scaffolding types	Discipline
Subcategories	<ul style="list-style-type: none"> • Feedback • Question prompts • Hints • Expert modeling • Multiple strategies 	<ul style="list-style-type: none"> • Conceptual scaffolding • Metacognitive scaffolding • Strategic scaffolding • Motivational scaffolding 	<ul style="list-style-type: none"> • Science • Technology • Engineering • Mathematics • Integrated STEM

possible that their opinion can represent current trends about the effects of a certain treatment. Unfortunately, there are few summarized experts' opinions regarding the effects on computer-based scaffolding in PBL. Those that do exist can be seen as highly subjective. In this regard, this study excluded the use of expert opinion as a possible prior distribution models. As the second method, one can utilize the results of meta-analysis as a prior distribution. There are two representative meta-analyses related to computer-based scaffolding including intelligent tutoring systems (ITS)—Belland et al. (2017) and Kulik and Fletcher (2016). In the case of Kulik and Fletcher's meta-analysis, their research interests focused on how the effects of computer-based scaffolding including ITS vary in the various contexts of learning environments such as sample size, study duration, and evaluation types. This means that their results did not emphasize the characteristics of ITS. Therefore, it is difficult to utilize these results as prior distribution of this study, which focuses on the characteristics of scaffolding. Recently, a National Science Foundation-funded project (Belland et al. 2017) aimed to synthesize quantitative research on computer-based scaffolding in STEM education. The moderators in this project are overlapped with the many moderators in this study. However, the big difference between the previous TMA and this paper is the learning contexts. This paper only focuses on problem-based learning, but the contexts in the TMA included several problem-centered instructional models (e.g., inquiry-based learning, design-based learning, project-based learning). Such problem-centered instructional models incorporate many different teacher roles, learning goals and processes, student learning strategies, and scaffolding usage patterns (Savery 2006). This makes it difficult to apply the results of TMA into the informative prior distribution in this paper, which only handled problem-based learning.

The last method to set up the prior distribution is to use non-informative prior distribution. If one does not have enough prior information about the parameter θ , one can hope that the selected prior has little influence on the inferences for generating the posterior distribution of the parameter. In other words, non-informative prior distribution only contains minimal information about parameters. For example, if assuming the range of parameter from 0 to 1, one can set up μ (i.e., the mean of parameter) as 0 and the variances of μ as 1. But, if the parameter occurs on an infinite interval, the variance between studies should be large enough ($\tau^2 \rightarrow \infty$) to have little influence on the posterior. These techniques make the prior distribution non-informative. In this paper, several non-informative prior distribution models (i.e., uniform, DuMouchel, and Inverse Gamma), which specify different weighted values of the between-study variance, τ^2 , were used to identify the most suitable model fits for the given data.

Data Coding

All study features were coded by theory-driven constructs regarding scaffolding characteristics, roles, and the contexts of its use. And these codes were validated by experts in the field of scaffolding, problem-based learning, and STEM education. All effect sizes for all features were calculated by Hedges's g , a measure of effect size that is corrected due to the use of weighted standard deviations. In addition, considering the variances between individual studies with the wide range of education population, subject areas, and scaffolding interventions, this study utilized the random effects model assuming that the true effect size may vary from study to study. The kinds of possible quantitative results from the individual studies were F statistics, t statistics, mean differences, and chi-square. With these quantitative results, all effect sizes of computer-based scaffolding corresponding the moderators were calculated using the metan package of STATA 14. Two graduate students who have extensive knowledge of scaffolding and problem-based learning, as well as coding experience for meta-analysis, participated in the coding work. The primary coder selected the candidate studies based on the inclusion criteria and generated initial codes about the moderators (i.e., scaffolding type, scaffolding strategy, scaffolding customization, scaffolding customization methods, higher-order thinking skills, and disciplines) in this study. The second coder also coded the data independently, and then the coding between two coders was compared. When there was inconsistency of coding between the coders, consensus codes were determined through discussion. Inter-rater reliability was calculated using the Krippendorff's alpha statistic when the initial coding was finished. Krippendorff's alpha is used to measure the level of coders' agreement on the values of variables (i.e., nominal, ordinal, and ratio) in the coding rubric by computation of the weighted percent agreement and weighted percent chance agreement (Krippendorff 2004). Krippendorff (2004) recommended 0.667 as the minimum acceptable alpha value to avoid the wrong conclusion from unreliable data. All Krippendorff's alpha values ($\alpha \geq 0.8$) across all moderators were above the minimum standard for reliable data, and this indicated that there was strong agreement between the two coders (see Fig. 1).

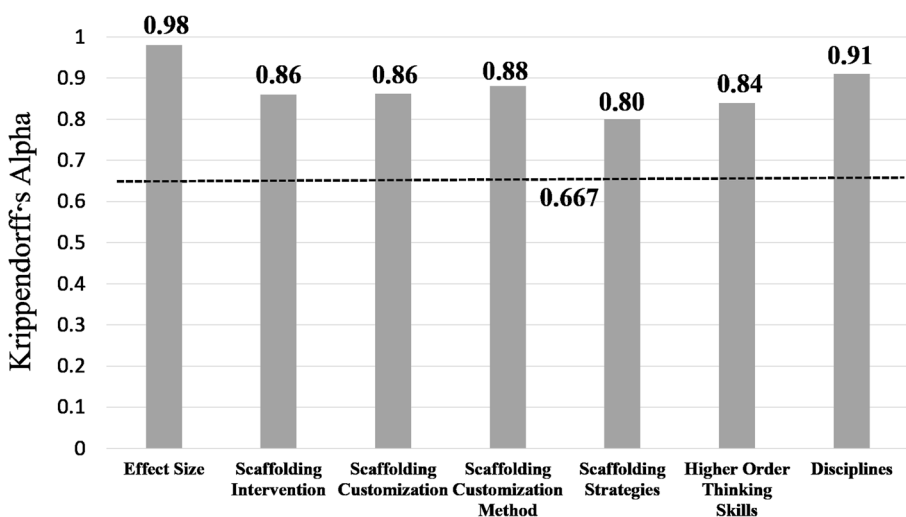


Fig. 1 Krippendorff's alpha for inter-rater reliability (*dotted line* indicates minimum acceptable reliability (Krippendorff 2004))

Data Analysis

For data analysis, STATA 14 and WinBUGS 1.4.3 were utilized. WinBUGS 1.4.3 provides Bayesian estimation including prior distributions options by MCMC and STATA 14 imported the results and codes from WinBUGS and generate graphical representations.

Markov Chain Monte Carlo simulations were used to sample from a probability distribution of the Bayesian model (Dellaportas et al. 2002). Integrating Markov chains can replace unstable or fluctuated initial values of random variables with more accurate values through repetitive linear steps, in which the next state (i.e., value of variable) can be influenced by the current one, not by preceding one (Neal 2000). In this process, 22,000 MCMC iterations for estimation of posterior distribution were generated and 2000 initial iterations were omitted to eliminate initial values that were randomly given. After analysis, deviance information criteria (DIC) was utilized for identifying model fits (Spiegelhalter et al. 2002). The lowest value in DIC means the best model to predict the reproduction of data as the observed data (Spiegelhalter et al. 2004). DuMouchel for “scaffolding customization methods” and uniform prior distributions for all remaining moderators had the smallest value of DIC (see Table 3). Uniform and DuMouchel assume different variances between studies (i.e., τ^2), and the results can differ according to which prior distribution was used, even though the underlying dataset is the same. After MCMC generated the posterior distribution of each moderator, the validation of models was investigated through four types of graphs—trace plots, autocorrelation plots, histogram plots, and density plots.

Observed Data Characteristics The number of observed data points across subcategories within moderators is unbalanced (see Table 4). There was no included study that involved motivation scaffolding; thus, motivation scaffolding could not be included in this paper. Moreover, around 10.6% of the outcomes included in this paper had small sample sizes ($n < 10$), resulting in the possibility of small-study effects. Smaller studies often show larger effect sizes than larger ones, leading to overestimation of treatment effects (Schwarzer et al. 2015).

To investigate empirically if there were small-study effects, I conducted Egger’s regression to test the null hypothesis of “there are no small-study effects” (see Table 5). The result shows that there are small-study effects since the null hypothesis was rejected, $p < 0.05$.

More than 80% of recently published meta-analyses may contain biased results caused by small-study effects (Kicinski 2013). This means that there may be a high probability of having biased results if a traditional meta-analysis approach with this data was employed. However,

Table 3 DIC values of prior distributions

Moderators	Uniform prior Distribution	DuMouchel Prior distribution	Inverse Gamma prior distribution
Scaffolding	520.25	524.85	523.12
Scaffolding intervention	93.19	94.27	95.23
Scaffolding customization	84.58	86.89	87.54
Scaffolding customization methods	89.30	87.26	90.78
Scaffolding strategies	133.22	138.45	139.10
Higher-order thinking	117.90	118.23	119.54
Discipline	89.02	91.54	92.21

Table 4 Number of outcomes according to subcategories

Higher-order skills	<i>N</i>	Scaffolding customization	<i>N</i>	Scaffolding customization methods	<i>N</i>
• Analysis	11	• None	32	• None	32
• Synthesis	14	• Fading	5	• Performance-adapted	8
• Evaluation	22	• Adding	3	• Fixed	4
		• Fading/adding	7	• Self-selected	3
Scaffolding strategies	<i>N</i>	Scaffolding intervention	<i>N</i>	Discipline	<i>N</i>
• Feedback	6	• Conceptual	31	• Science	25
• Question prompts	24	• Metacognitive	4	• Technology	7
• Hints	6	• Strategic	12	• Engineering	3
• Expert modeling	2	• Motivational	0	• Mathematics	5
• Multiple strategies	9			• Integrated STEM	7

using a Bayesian approach can address potential small-study effects by shrinking overweighted effect sizes through interval estimation and the appropriate use of priors (Kay et al. 2016; Kicinski 2013; Mengersen et al. 2016).

Interpretation of Bayesian Meta-analysis

Bayesian inference is based on posterior probability, which is in turn based on the likelihood of the observed data, not point or interval estimation from the frequentist approach (i.e., confidence interval (CI)) because the standard error gets closer to 0 due to the large number of samples generated through MCMC simulations (Robins and Wasserman 2000). The Bayesian 95% credible interval (CrI) is similar in some ways to the 95% CI from the frequentist perspective, but there is a big difference between them in terms of basic principle and interpretation. A 95% confidence interval means the range including the true effect size in 95% of the cases across all possible samples from the same population. This means that the population parameters are fixed and the samples are random in the frequentist approach (Edwards 1992). On the other hand, Bayesian approach regards parameters as random and samples as fixed (Ellison 2004). Therefore, the Bayesian 95% CrI indicates a 95% probability range of values on the posterior distribution (i.e., parameters), which is generated by fitting the predetermined prior distribution including the information of parameters into the observed data. For example, a wider CI means huge standard error caused by little knowledge of effects or small samples, but CrI is a range of true effects on treatments at the level of populations. Therefore, most Bayesians are reluctant to use frequentist hypothesis testing using *p* values because the observed data was not included in the posterior distribution (Babu 2012; Bayarri and Berger 2004; Kaweski and Nickeson 1997). However, many scholars interpret the results

Table 5 Egger's regression test for small-study effects

	Coefficient	Std. Err.	<i>t</i>	<i>p</i>	95% CI	
					Lower	Upper
Slope	.09	.13	0.72	0.48	-.17	.35
Bias	1.13	.52	2.18	0.04	.08	2.17
Test of H_0 : no small-study effects				0.04		

of Bayesian analysis with the perspective of frequentists, and this causes misunderstanding of results (Gelman et al. 2014).

Results

Overall Effect Size of Computer-Based Scaffolding

The general effects of computer-based scaffolding compared to the group who did not receive any scaffolding was $g = 0.385$ (95% CrI = 0.022 to 0.802) based on 20,000 simulated iterations. The range of true effect sizes of computer-based scaffolding is modeled as population statistics from 0.022 to 0.802 with 95% probability, and this means that students who used computer-based scaffolding within the context of PBL showed better learning performance than those who did not at the level of population, which is normally distributed ($\theta = 0.385$, $\sigma = 0.254$, $\tau^2(0,1000)$) (see Fig. 2).

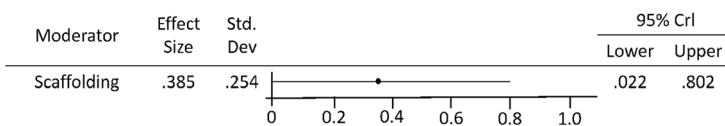
To verify the model with the above results, graphical summaries were generated (see Fig. 3). The pattern of trace was vertical and dense without any fluctuation of data, and this means that the mean parameter was well-mixed. In addition, the autocorrelation plots indicated that the value of autocorrelation closed to 0 as the lag increased. This indicated that the values generated by each lag were independent. Furthermore, histogram and density plots provided evidence supporting a normal distribution of the generated samples.

Subgroup Analysis

The non-informative prior distribution, which was utilized in this study, considers all possible τ^2 (between-study heterogeneity) across the studies, and it justified subgroup analysis to identify potential moderator variables. This study has six moderators (i.e., scaffolding intervention, scaffolding customization, scaffolding customization methods, scaffolding strategies, higher-order thinking, and discipline).

Scaffolding Intervention Students who received computer-based scaffolding showed better cognitive outcomes than those who did not (see Fig. 4). The effect size was largest for metacognitive scaffolding ($g = 0.384$, 95% CrI = 0.17 to 0.785), followed by strategic ($g = 0.345$, 95% CrI = 0.14 to 0.692), and conceptual scaffolding ($g = 0.126$, 95% CrI = 0.003 to 0.260).

The generated samples were well mixed as shown in Fig. 5. In the trace plots, all data were stable and did not show any sharp changes. Moreover, the results of autocorrelation were close to 0 and the histogram and density plots indicated the normal distribution of the simulated data.



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 2 Effect size of scaffolding

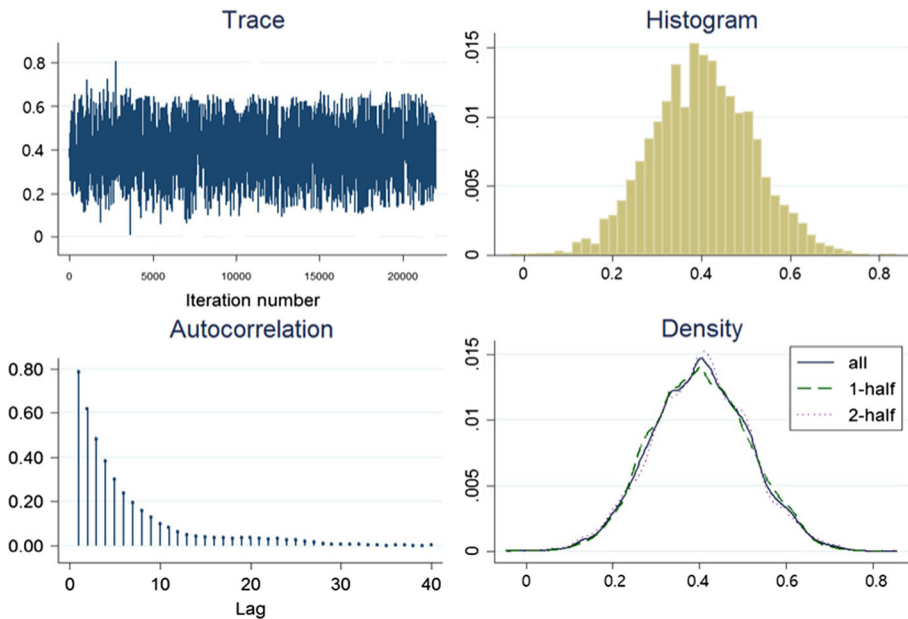
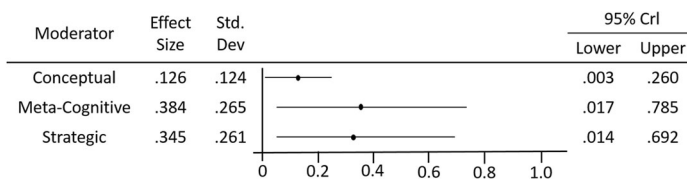


Fig. 3 The graphic summaries for overall effects of computer-based scaffolding

Scaffolding Customization Scaffolding customization means that scaffolding can be added, faded, or added and faded according to students’ current learning status and abilities. Scaffolding can be gradually faded as students gain knowledge and skill or increased when students need more help. Moreover, these fading and adding functions of scaffolding can occur simultaneously. The result indicates that when students received customized scaffolding, the effects of scaffolding were noticeably higher than no customized scaffolding and no scaffolding (see Fig. 6). Among scaffolding customization types, fading/adding showed the highest effect sizes ($g = 0.590$, 95% CrI = 0.155 to 0.985) compared to fading ($g = 0.429$, 95% CrI = 0.015 to 0.865) and adding scaffolding ($g = 0.443$, 95% CrI = 0.012 to 0.957). Each scaffolding customization type showed a wide range of true effect sizes (credible interval) at the level of population, but it is clear that applying scaffolding customization shows better effects on students’ learning performance than no scaffolding customization at the level of population.

There were no issues in terms of convergence, standard errors, and normality of the generated samples by MCMC, and the results of scaffolding customization represented the parameters as shown in Fig. 7.



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 4 Effect size of subcategories within scaffolding intervention

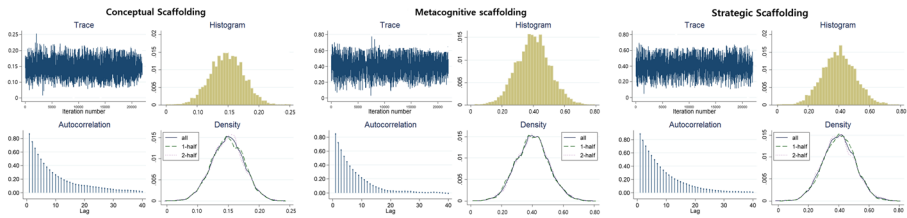


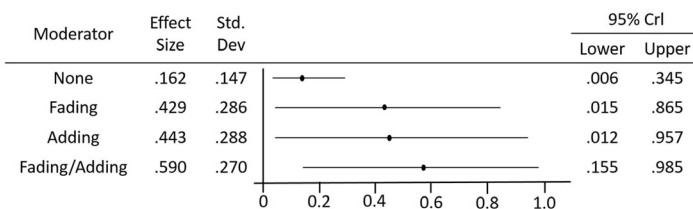
Fig. 5 Graphic summaries for effects of scaffolding intervention

Scaffolding Customization Methods Scaffolding customization (i.e., fading, adding, fading/adding) proceeds according to performance-adaption, self-selection, and fixed time intervals. Scaffolding can be customized based on test scores or formative assessment (performance-adapted). In addition, students themselves can request the fading or adding of scaffolding based on their judgment (self-selected). Unlike scaffolding customization methods by performance-adaption and self-selection, which are happening during students’ learning, scaffolding can be customized by fixed time interval, which is determined by scaffolding designer regardless of students’ performance and decision.

When scaffolding was customized based on self-selection, the effect sizes of scaffolding were higher ($g = 0.519$, 95% CrI = 0.167 to 0.989) than other scaffolding customization methods: fixed time interval ($g = 0.376$, 95% CrI = 0.018 to 0.713) and performance-adaption ($g = 0.434$, 95% CrI = 0.013 to 0.863) as shown in Fig. 8.

None means both no scaffolding customization and no scaffolding customization methods. Nevertheless, the effect sizes between two “Nones” from scaffolding customization ($g = 0.162$) and scaffolding customization methods ($g = 0.122$) were slightly different, although the data for two moderators was exactly the same. One possible reason is the usage of different prior distributions, which assume different variance between the studies (τ^2). The uniform prior distribution utilized in the scaffolding customization model weighted the larger value of τ^2 than the Dumouchel prior distribution in the scaffolding customization methods model. Therefore, none in the category scaffolding customization had wider CrIs than one in the scaffolding customization methods category. The graphical summaries illustrated the well-mixed nature of the MCMC chains (see Fig. 9).

Scaffolding Strategies Scaffolding can assume several forms such as hints, feedback, question prompts, expert modeling, and multi-forms as combination. The results show that the effect size for expert modeling ($g = 0.523$, 95% CrI = 0.030 to 0.979) and feedback ($g = 0.474$, 95% CrI = 0.026 to 0.968) approached a middle level. Other forms (i.e., hints ($g = 0.375$, 95% CrI = 0.013 to 0.742) and multi-forms ($g = 0.340$, 95% CrI = 0.012 to 0.698) also had relatively higher effects on students’ cognitive outcomes than students who did not



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 6 Effect size of subcategories within scaffolding customization

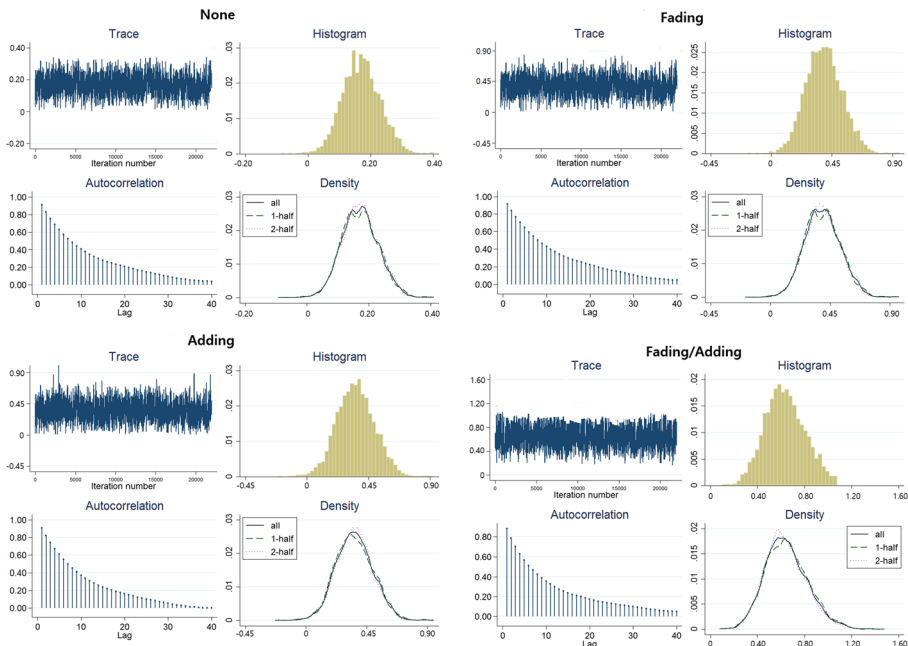
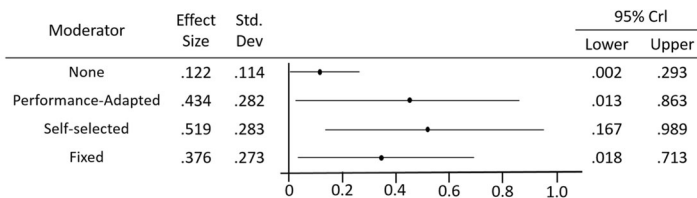


Fig. 7 The graphic summaries for effects of scaffolding customization

receive scaffolding (see Fig. 10). However, in the case of question prompts, there was no evident difference ($g = 0.078$, 95% CrI = 0.003 to 0.156) as compared to the control group in terms of the effect size.

The stability of MCMC chain and the normality of generated samples had no problems, but in the case of autocorrelation, the correlation did not rapidly drop toward 0 as the lag increased (see Fig. 11). This can cause inaccurate prediction of parameters due to the correlation of the values generated from each lag. This issue could be caused by inadequacy of the observed samples for the predetermined prior distribution or by the wrong selection of prior distribution. However, it was clear that the autocorrelation of these categories was evidently reduced and closer to 0 as the lags increased. This means that the values simulated by MCMC accurately predict each parameter of the categories (Geyer 2010).

Higher-Order Thinking Higher-order thinking is one of the important intended outcomes expected by scaffolding. When scaffolding intended to improve students’ analysis ability, which identifies the components of information and ideas and establish the relations between elements, the effect size was highest at the population level ($g = 0.537$, 95% CrI = 0.038 to



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 8 Effect size of subcategories within scaffolding customization methods

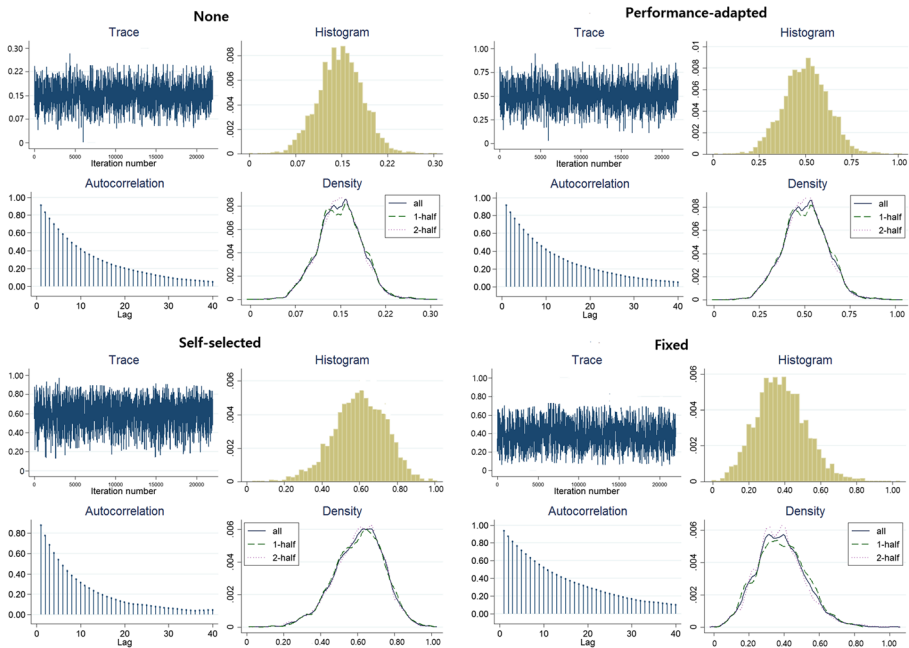
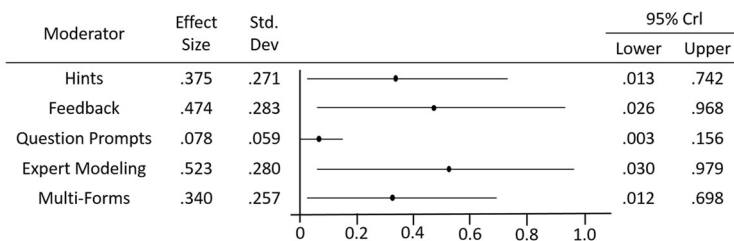


Fig. 9 Graphic summaries for effects of scaffolding customization methods

0.981). However, higher-order thinking skills related to synthesis ($g = 0.156$, 95% CrI = 0.004 to 0.329) and evaluation ($g = 0.147$, 95% CrI = 0.003 to 0.288), which require higher level application of knowledge than analysis ability, were not improved to a large extent by computer-based scaffolding (see Fig. 12).

The graphical summaries provide evidence that outcomes from the samples generated by MCMC represent the parameters of higher-order skills due to well-mixed data (see Fig. 13).

Discipline Scaffolding had higher effect sizes when it was used in Engineering ($g = 0.528$, 95% CrI = 0.025 to 0.983), Technology ($g = 0.379$, 95% CrI = 0.011 to 0.782), and Mathematics education ($g = 0.425$, 95% CrI = 0.024 to 0.883) than when it was utilized in Science ($g = 0.146$, 95% CrI = 0.003 to 0.295) as shown in Fig. 14. In the case of integrated STEM education, which integrated the content and process of disciplines, the effect size was relatively lower ($g = 0.201$, 95% CrI = 0.005 to 0.428).



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 10 Effect size of subcategories within scaffolding strategies

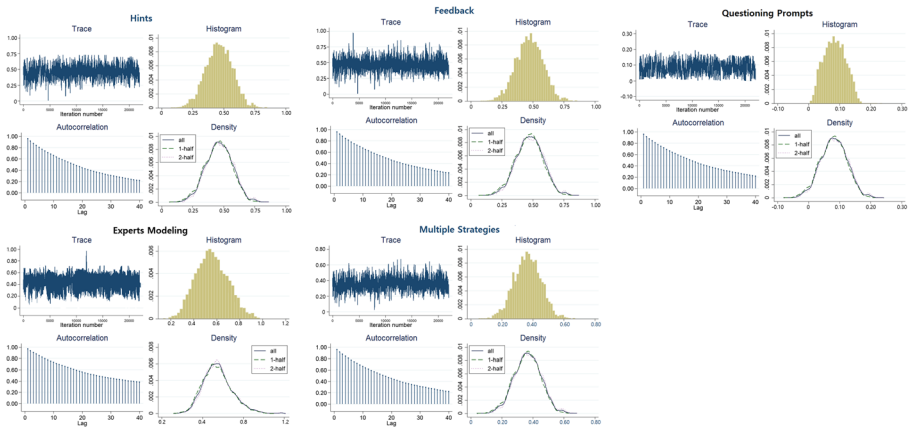


Fig. 11 The graphic summaries for effects of scaffolding strategies

Figure 15 showed that the generated samples were well-mixed by MCMC, and the results of BMA can represent parameters.

Discussion

Overall Effects of Computer-Based Scaffolding

The overall effect size of computer-based scaffolding used in the context of PBL obtained from Bayesian meta-analysis was $g = 0.385$, which would be labeled as a small/moderate effect size by Cohen (1988). This result means that computer-based scaffolding utilized in the context of problem-based learning for STEM education is an effective means to improve students’ higher-order skills and learning outcomes. The overall effect size estimate in this meta-analysis parallels the results from meta-analyses on the effects of computer-presented feedback ($g = 0.35$; Azevedo and Bernard 1995) and on the effectiveness of computer-based supports ($g = 0.37$; Hattie 2008) despite different contexts of scaffolding usages. However, the effect size of computer-based scaffolding in this study ($g = 0.385$) was lower than that obtained in a traditional meta-analysis that covered scaffolding used in the context of a variety of problem-centered instructional models ($g = 0.46$; Belland et al. 2017). One reason why the results between the two meta-analyses may have differed is that the context in which scaffolding was used differed. Another possible reason is that BMA could resolve issues of overweighted effect

Moderator	Effect Size	Std. Dev	95% CrI	
			Lower	Upper
Analysis	.537	.278	.038	.981
Synthesis	.156	.143	.004	.329
Evaluation	.147	.149	.003	.288

Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 12 Effect size of subcategories within higher-order skills

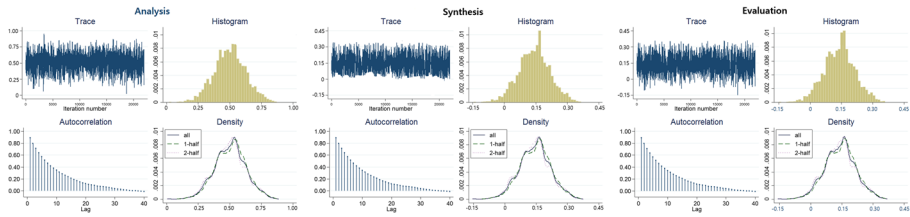


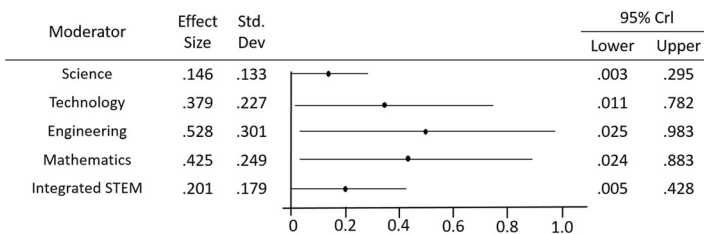
Fig. 13 The graphic summaries for effects of computer-based scaffolding on higher-order skills

size caused by smaller studies, which were included in the traditional meta-analysis (Kay et al. 2016; Kicinski 2013; Mengersen et al. 2016).

Implications for Research

Various scaffolding strategies played a role in improving learning performance in problem-based learning in STEM education. This corresponds with the positive results of hints (Raphael et al. 2008), expert modeling (Pedersen and Liu 2002), and feedback (Rouinfar et al. 2014) from previous empirical studies. However, the results indicated that providing multiple strategies of scaffolding in one scaffolding system is counter-productive. This aligns with research that shows that multiple forms of scaffolding regardless of students’ current needs and learning status can be less effective (Aleven and Koedinger 2002; Azevedo and Hadwin 2005; Baylor and Kim 2005). One interesting result in terms of scaffolding strategies is that question prompts had the smallest effect size ($g = 0.078$) among the scaffolding strategies (e.g., expert modeling, hints, and feedback). One of the main advantages posited for question prompts is to improve students’ metacognition (Chen and Bradshaw 2007; Davis 2003; Scardamalia et al. 1989) by helping students draw inferences from their evidence and encourage their elaborative learning. However, much research has demonstrated that effectiveness of question prompts can vary according to individuals’ abilities (e.g., prior knowledge, cognition and metacognition, and problem-solving skills) (Lee and Chen 2009), and different question prompts (i.e., procedural, elaborative, and reflective prompts) should be provided to different student (Ge et al. 2005). One possible reason why the effect of question prompts to enhance metacognition is smallest is that most studies utilizing question prompts in this meta-analysis did not consider students’ various level of prior knowledge and problem-solving skills.

When scaffolding was intended to help students identify components of information and ideas (i.e., analysis level of higher-order thinking), its effect size was much higher than when supporting the recognition of the patterns of components (synthesis level) and judging the



Note: MCMC iterations = 22,000, Burn-in = 2,000, MCMC samples = 20,000, 95% CrI: 95% Credible Interval

Fig. 14 Effect size of subcategories within discipline

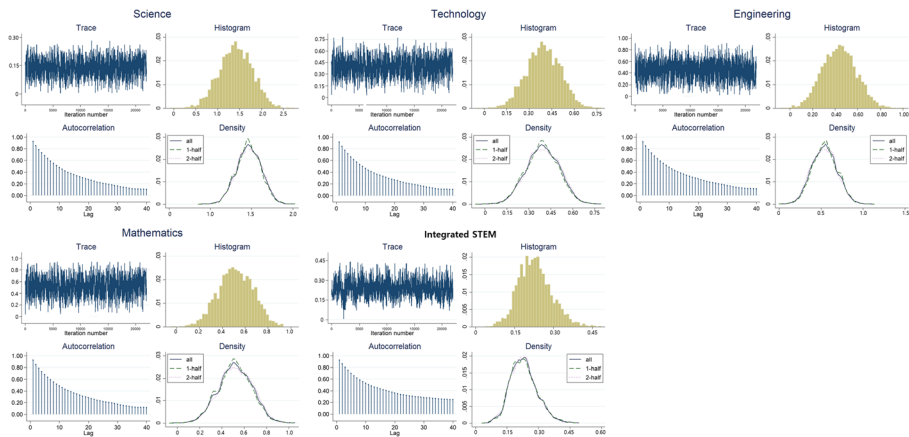


Fig. 15 The graphic summaries for effects of computer-based scaffolding on disciplines

value of data for justifying conclusion (evaluation level). This shows that current technology and designs often struggle to support students' more advanced and complicated learning process in PBL such as "synthesis," "application," and "reflection" (Wu 2010). This supports many scholars' claims that the effectiveness of computer-based scaffolding can be maximized when integrated with one-to-one scaffolding by teachers (Hoffman et al. 2003; Pata et al. 2006). Therefore, computer-based scaffolding can be effectively provided at the level of analysis with current technology in PBL.

The effect of scaffolding was highest ($g = 0.528$) when it was used in the context of Engineering. Mathematics and Technology education have been regarded as pre-requisite subjects for engineering courses in K-12 education (Douglas et al. 2004). Therefore, the high effects of scaffolding across Engineering, Mathematics, and Technology are understandable due to the curriculum sharing and similarity of problem-solving process. However, in the case of problem-based learning for science education, BMA showed a relatively lower effect size ($g = .146$) than for other disciplines (i.e., Technology, Engineering, and Mathematics). In science education, many scholars have engaged in debates on the proper instructional design, the characteristics of scaffolding, and curriculum, and this led to an inconsistency in scaffolding design among science education researchers (Lin et al. 2012). For example, some scholars insisted that science education in PBL requires students' advanced scientific inquiry and higher-order thinking skills (Hoidn and Kärkkäinen 2014), but Zohar and Barzilai (2013) claimed that students' identification of what should be considered for problem-solving by understanding of content knowledge in science education should take precedence over enhancing higher-order thinking skills.

Another interesting finding was the low effect size of computer-based scaffolding in integrated STEM education. Contrary to the results of several meta-analyses on the effectiveness of integrated STEM education (Becker and Park 2011; Lam et al. 2008), the current review indicated that computer-based scaffolding was not very effective at improving students' learning performance in integrated STEM education.

Ill-structured problems in PBL cannot be solved by mere application of existing knowledge because solving this type of problem requires students' advanced thinking process and strategy rather than the recall of information (Beyer 1997). In this sense, Hmelo-Silver and Ferrari

(1997) argued that higher-order thinking skills such as analysis, synthesis, and evaluation of knowledge are essential abilities to accomplish PBL tasks. Higher-order thinking skills allow students to (a) understand the given problematic situation and to identify the gaps in their knowledge to solve the problem and (b) apply knowledge to find the solution to the problem (Hmelo-Silver and Lin 2000). Students can enhance higher-order thinking skills through active participation in such activities as making hypotheses, gathering evidence, and generating arguments (Lewis and Smith 1993). Many studies indicated that computer-based scaffolding can enhance students' higher-order thinking skills (Rosen and Tager 2014; Zydney 2005, 2008). Nevertheless, the results of BMA demonstrated that computer-based scaffolding did not excel at supporting every level of higher-order thinking skills (i.e., analysis, synthesis, and evaluation of knowledge). When scaffolding was intended to help students identify the components of information and ideas (i.e., analysis level of higher-order thinking), its effect size ($g = 0.537$) was much higher than when supporting the recognition of the patterns of components (synthesis level) ($g = 0.156$) and judging the value of data for justifying conclusions (evaluation level) ($g = 0.147$). In other words, computer-based scaffolding was very effective in helping students understand problematic situations and compare information required to find out the solutions (i.e., analysis level). However, computer-based scaffolding is often ineffective at supporting students as they (a) build new knowledge by reorganizing existing information (i.e., synthesis level) and (b) validate their claims based on this new knowledge (i.e., evaluation level). This corresponds with many researchers' claims that students struggle to improve synthesis and evaluation skills in spite of computer-based supports (Hu 2006; Jonassen 1999b). This result has an important implication that scaffolding design for improving students' higher-order thinking should be differentiated according to which level (i.e., analysis, synthesis, and evaluation) of higher-order thinking is addressed in PBL.

Effects of Scaffolding Characteristics in Different Contexts

The computer-based scaffolding interventions included in this paper all came from the context of problem-based learning (PBL). PBL requires students to hypothesize the solution of ill-structured problems, to verify their claims with reasonable evidence, and to reflect on their learning performance. In large part, whether the student can successfully navigate each of these PBL requirements depends on the students' higher-order thinking skills. In this way, students' abilities to reflect on their own thinking (metacognition) are regarded as essential to successful PBL learning. Moreover, students' own learning strategies are important to help make the complicated PBL procedure more accessible and understandable through enhancing self-directed learning and lifelong learning skills (Hmelo-Silver 2004; Schoenfeld 1985). Several studies have demonstrated the importance of metacognition and learning strategies to improve students' learning outcomes in problem-based learning (Downing et al. 2009; Sungur and Tekkaya 2006).

Implications for Scaffolding Design

The results of BMA showed that metacognitive ($g = 0.384$) and strategic scaffolding ($g = 0.345$) led to relatively large effects on students' learning outcomes in the context of problem-based learning. On the other hand, conceptual scaffolding did not show large overall effects ($g = 0.126$) on learning as compared to the control groups. This finding contradicts the

results of previous meta-analyses, which reported better or similar effectiveness of conceptual scaffolding compared to metacognitive and strategic scaffolding in the context of problem-centered instructional approaches (Belland et al. 2015, 2017). This might be due to the fundamental nature of PBL, with a loosely specified end goal presented to students at the beginning of the unit, and less inherent support built into the model than other models like design-based learning and project-based learning (Belland et al. 2017). PBL requires students' higher level of cognition and retention of knowledge (Downing et al. 2009); as such, metacognitive and strategic scaffolding might have a larger impact on students' problem-solving in PBL than conceptual scaffolding.

What this means for scaffolding designers is that, rather than focus on conceptual scaffolding, which is by far the most common ($N = 31$ outcomes in this meta-analysis) scaffolding type used in the context of PBL, a better approach would be to focus on creating metacognitive and strategic scaffolds. This would entail a shift from using scaffolding to inform students of important things to consider when addressing the problem, to bootstrapping a known effective strategy for addressing the problem, and helping students question their own understanding (Hannafin et al. 1999). It is interesting though that, compared to conceptual scaffolding, the scaffolding intervention types that came out of top are in effect either giving students more (in the case of strategic scaffolding) or less (in the case of metacognitive scaffolding) structure to the process with which students can address the problem. That is, for a given problem X , conceptual scaffolding says "here are some important things that should be considered when addressing problem x ," strategic scaffolding says "here's a great strategy to address problem x and here is how you do it," and metacognitive scaffolding says "here is how you can assess the extent to which you addressed problem x well." It is possible that conceptual scaffolding just does not give enough insight into concrete things that students can do to move their investigation forward in the context of PBL. Further research, including design studies and multiple treatment studies are needed to fully see if this is the case.

Furthermore, our findings indicate that in terms of scaffolding form, expert modeling and feedback have the strongest effects. To our knowledge, there are no previously published comprehensive meta-analyses that focus on scaffolding forms. However, one meta-analysis focusing on the scaffolding niche of dynamic assessment found great promise for the tool among students with special needs (Swanson and Lussier 2001). Expert modeling has often been associated with conceptual scaffolding. For example, in *Alien Rescue* (Liu and Bera 2005), an expert states the things he would consider when thinking about the most appropriate planet to choose as a new home for a given alien species. But it could equally well provide an effective framework for strategic or metacognitive scaffolding. For example, after students encounter a problem, an expert could say something like "you know, I saw a problem like this in the past, and here is what I did to address it.... So this is what you need to do to carry out this strategy."

Feedback presents an interesting conundrum for scaffolding in PBL, in that students could potentially take an infinite number of acceptable paths to an acceptable solution to a PBL problem. Therefore, setting up a system of providing useful feedback in the scaffolding sense is difficult. One would likely need to use artificial intelligence technologies that employ automated content analysis of articulations of students' thoughts.

Within PBL, theoretical arguments are often made that fading is necessary (Collins et al. 1989; Dillenbourg 2002; Pea 2004; Puntambekar and Hübscher 2005). In this meta-analysis,

only 10% of included outcomes were associated with fading, only 6% with adding, and 15% with fading/adding, which is consistent with prior research (Belland et al. 2017; Lin et al. 2012). Using a traditional meta-analysis approach would have rendered it difficult to compare the effects of fading, adding, and no customization. However, according to BMA, which overcomes this issue, cognitive effects of scaffolding were highest when fading and adding were combined, ($g = 0.590$). This is a striking finding, in that it goes against the grain of most theoretical arguments in PBL. Most often, when fading is combined with adding in scaffolding, the theoretical basis that drives the design is that of Adaptive Control of Thought—Rational (ACT-R) theory, according to which human knowledge is gained through integrated production from the declarative and procedural memory modules (Anderson et al. 1997). That is, successful performance in learning depends on each student's amount of the existing knowledge and the ability to retrieve and apply the relevant knowledge from their memory for the given tasks. ACT-R has not often been used to inform the design of learning supports for PBL. But it may be worth considering in future design of scaffolding and other supports for PBL. ACT-R also provides justification for self-selected adding (hints). Self-selection was found to lead to the strongest cognitive effects in this meta-analysis.

Limitations and Suggestions for Future Research

Heterogeneity of Computer-Based Scaffolding Across Studies Regarding Its Level and Type

In this study, various strategies of computer-based scaffolding from individual studies were categorized into several forms (i.e., hints, feedback, question prompts, expert modeling, and multi-forms as combination). However, any time one groups instructional strategies into a constrained set of categories, there can be substantial heterogeneity among the strategies that belong to any one category. In addition, according to the theoretical foundations (e.g., knowledge integration, activity theory, and ACT-R) to design and develop computer-based scaffolding, These scaffolds can have different learning goals and interventions. This can cause heterogeneity between studies under the same category, leading to risk of bias (Higgins et al. 2011). To overcome this limitation, a more nuanced study with more specific categorization of scaffolding forms may be beneficial.

Reliability of Results

The results of BMA are estimated based on probabilistic inferences. However, the probability cannot necessarily say what is actually done in the real world, and it is just a prediction. The results of BMA were reasonable with high level of probability, but an exception can occur. Thus, one should not have blind faith in the BMA results. To minimize this limitation, many researchers are trying to develop new techniques of Bayesian analysis, but the frequentist method might theoretically bring more accurate meta-analysis results if the observed data with large samples satisfies all statistical conditions, which do not cause any bias of results. However, there have been few studies to directly compare the results from BMA and TMA, and it might be difficult to tell about which approach among BMA and TMA can bring more reliable results. Therefore, a follow-up study is required to verify the effects of BMA as an alternative to TMA.

The Selection of Prior Distribution

The estimation of posterior distribution is largely influenced by how one set up the prior distribution (Findley 2011; Lewis and Nair 2015). For example, if one generates a prior distribution with a certain value of mean and variance, the distribution of each observed datum can be matched with this predetermined prior distribution. This means that the posterior distribution simulated by the prior distribution and likelihood also has the same pattern of parameters with prior distribution. Therefore, the results of BMA can be different according to how researchers set up the prior distribution. To solve this issue, in the medical field, in which BMA has been commonly used, the use of informative prior distributions rather than non-informative prior distributions has gradually increased due to the accumulated data of treatments obtained from several meta-analyses with same or similar topics and moderators (Klauenberg et al. 2015; Turner et al. 2015). This accumulated data can serve as the standard value of prior distribution for future BMAs, resulting in the possibility of an informative prior distribution. This informative prior distribution predicts the posterior distribution more accurately and effectively than the non-informative prior distribution, which just assumes the normal distribution of parameter with the prior information of a certain treatment. However, in educational research, there are few meta-analyses utilizing informative prior distribution due to a lack of consensus among scholars and prior meta-analysis on certain research topics or educational tools. Hopefully, the results of this study will play a role as the informative prior distribution for the effectiveness of computer-based scaffolding in PBL, which allows following BMAs on this topic to collaboratively build a more robust prior distribution model.

Expansion of Research Scope

This study only included studies on the influence of computer-based scaffolding on cognitive outcomes in the context of problem-based learning in STEM education. However, other forms of scaffolding including one to one scaffolding (e.g., teacher scaffolding) and peer scaffolding are commonly utilized. Therefore, Bayesian meta-analysis with a broad subject of scaffolding (see Table 6) as future research is also needed. It is clear though that such an endeavor would be challenging, as quantitative studies in which (a) teacher scaffolding for example can be compared to a control condition and (b) data for which effect sizes can be calculated are not plentiful. One of the problems has to do with the fact that teacher scaffolding is so dynamic that any given student in a classroom could receive any number of minutes of teacher scaffolding within a 50-min period, while some students may receive none at all (Van de

Table 6 List of independent and dependent variables for future Bayesian meta-analysis

Independent variables	Dependent variables
1. Educational level: from K-12 to adult	1. Content knowledge
2. Target population: students with disabilities, English language learner, low/high-SES students, low/high performing students	2. Motivation
3. Scaffolding format: computer-based scaffolding, teacher scaffolding, peer scaffolding	3. Collaboration skills
4. Learning environment outside of STEM	

Pol et al. 2010; Wood et al. 1976). The same rationale applies to peer scaffolding, though it is less likely that a given student would receive no peer scaffolding within a class period. Nonetheless, performing Bayesian meta-analyses along these lines has the potential to be very valuable.

Moreover, this study could not include motivational scaffolding as subcategory of scaffolding intervention due to a lack of studies with motivational scaffolding in the context of problem-based learning that satisfied our inclusion criteria. Many research has reported the importance and effects of students' motivation in their learning outcomes (Perkins and Salomon 2012), and future research should investigate how and to what extent motivational scaffolding can affect students' learning outcomes in problem-based learning.

Conclusion

This Bayesian meta-analysis demonstrated the effectiveness of computer-based scaffolding on students' learning performance and higher-order thinking skills in the context of problem-based learning for STEM education. Currently, many researchers have been interested in designing and developing computer-based scaffolding. However, various types of scaffolding are often utilized without considering the characteristics of the given learning contexts. For example, the goal of problem-based learning is not to improve students' content knowledge, but to enhance students' advanced problem-solving skills and thinking strategies based on the application of the existing knowledge (Savery 2015). Nevertheless, teacher-based and computer-based supports to improve students' content knowledge are the most widely used scaffolding type in the context of PBL, but its effectiveness on PBL is imperceptible. This indicated that instructional designers should focus more on the metacognitive and strategic scaffolding interventions to improve students' problem-solving skills using their own learning strategies in the learning context of PBL.

The most interesting result in this paper is on the effects of combining adding and fading within scaffolding customization. Adding scaffolding has often been disregarded in the context of PBL. But in PBL, which requires students' different abilities such as information-searching strategy, problem-solving skills, creative thinking, and collaborative learning skills, adding scaffolding should be considered alongside fading as an effective strategy for promoting strong learning outcomes. In addition, when scaffolding customization occurs by students themselves, its effects may be best because self-selected scaffolding can improve students' self-directed learning and motivation toward their learning. According to the results, question prompts are the most prominent strategy of computer-based scaffolding, but its effect was not as strong as many scholars believed (Choi et al. 2008; Ge and Land 2003). The original goal of scaffolding to address this issue might be considered. The main purpose of scaffolding is to help students improve engagement in learning and successfully accomplish the given tasks that are beyond their current abilities (Wood et al. 1976). However, if question prompts in PBL are provided in a complicated and difficult manner, the question prompts can be a hindrance of learning instead of support. Therefore, this paper suggests that more simple and directive supports such as hints and expert modeling are more suitable for PBL.

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

Appendix

Table 7 Summary of included outcomes

Study	ES	Higher-order thinking skills	Scaffolding intervention	Disciplines	Scaffolding customization	Scaffolding customization methods	Scaffolding strategy
Ak 2015	0.50	Analysis	Strategic	Technology	Fading	Self-selected	Question prompts
Ak 2015	0.19	Evaluation	Strategic	Technology	Fading	Self-selected	Question prompts
Ardac and Sezen 2002	0.13	Evaluation	Conceptual	Science	Adding	Performance-adapted	Hints
Ardac and Sezen 2002	0.66	Synthesis	Conceptual	Science	Adding	Performance-adapted	Hints
Belland 2008	0.09	Synthesis	Strategic	Science	None	None	Hints
Belland 2008	0.25	Synthesis	Strategic	Science	None	None	Hints
Barab et al. 2009	0.93	Synthesis	Conceptual	Science	None	None	Question prompts
Barab et al. 2009	1.01	Synthesis	Conceptual	Science	None	None	Question prompts
Barab et al. 2006	1.31	Analysis	Conceptual	Engineering	Fading/adding	Performance-adapted	Multiple strategies
Butz et al. 2006	0.97	Analysis	Conceptual	Engineering	Fading	Performance-adapted	Multiple strategies
Chen et al. 2003	1.06	Analysis	Conceptual	Science	Fading/adding	Performance-adapted	Feedback
Clarebout and Elen 2006	0.48	Evaluation	Strategic	Science	None	None	Feedback
Clarebout and Elen 2006	0.32	Evaluation	Strategic	Science	Adding	Performance-adapted	Feedback
Clark et al. 2012	0.59	Analysis	Conceptual	Science	Fading/adding	Fixed	Question prompts
Clark et al. 2012	0.52	Evaluation	Conceptual	Science	Fading/adding	Fixed	Hints
Ethers and Tan 2004	0.67	Evaluation	Strategic	Mathematics	None	None	Question prompts
Ge et al. 2010	1.83	Synthesis	Conceptual	Science	None	None	Question prompts
Hmelo-Silver and Day 1999	0.83	Synthesis	Conceptual	Science	None	None	Question prompts
Hmelo-Silver and Day 1999	0.82	Analysis	Conceptual	Science	None	None	Question prompts
Raes et al. 2012	0.02	Analysis	Conceptual	Science	Fading	Self-selected	Question prompts
Rosen and Tager 2014	0.65	Analysis	Strategic	Engineering	None	None	Feedback
Siler et al. 2010	0.80	Analysis	Strategic	Science	Fading/adding	Performance-adapted	Feedback
Simons and Klein 2007	0.64	Evaluation	Conceptual	Mathematics	Fading	Fixed	Multiple strategies
Simons and Klein 2007	0.36	Evaluation	Conceptual	Mathematics	None	None	Multiple strategies
Simons and Klein 2007	0.82	Evaluation	Conceptual	Mathematics	Fading/adding	Fixed	Multiple strategies
Simons and Klein 2007	0.99	Evaluation	Conceptual	Mathematics	None	None	Multiple strategies

Table 7 (continued)

Study	ES	Higher-order thinking skills	Scaffolding intervention	Disciplines	Scaffolding customization	Scaffolding customization methods	Scaffolding strategy
Stark 2013	0.78	Analysis	Conceptual	Integrated	None	None	Expert modeling
Stark 2013	0.13	Evaluation	Conceptual	Integrated	None	None	Question prompts
Su and Klein 2010	0.14	Evaluation	Conceptual	Technology	None	None	Question prompts
Su and Klein 2010	0.33	Evaluation	Conceptual	Technology	None	None	Question prompts
Su and Klein 2010	0.34	Evaluation	Metacognitive	Technology	None	None	Multiple strategies
Su and Klein 2010	0.05	Evaluation	Metacognitive	Technology	None	None	Multiple strategies
Su and Klein 2010	0.12	Evaluation	Metacognitive	Integrated	None	None	Multiple strategies
Su and Klein 2010	0.71	Evaluation	Conceptual	Technology	None	None	Question prompts
Su 2007	0.15	Evaluation	Conceptual	Integrated	None	None	Question prompts
Su 2007	0.24	Evaluation	Conceptual	Integrated	None	None	Question prompts
Su 2007	0.17	Evaluation	Conceptual	Integrated	None	None	Question prompts
Su 2007	0.11	Evaluation	Conceptual	Integrated	None	None	Question prompts
Toth et al. 2002	1.06	Synthesis	Strategic	Science	None	None	Hints
Toth et al. 2002	1.03	Synthesis	Metacognitive	Science	None	None	Expert modeling
Weusijana et al. 2004	0.55	Analysis	Strategic	Science	Fading/adding	Performance-adapted	Feedback
Zydney 2008	0.52	Synthesis	Conceptual	Science	None	None	Question prompts
Zydney 2008	0.03	Synthesis	Conceptual	Science	None	None	Question prompts
Zydney 2008	-0.03	Synthesis	Conceptual	Science	None	None	Question prompts
Zydney 2008	0.47	Evaluation	Conceptual	Science	None	None	Question prompts
Zydney 2008	0.03	Synthesis	Conceptual	Science	None	None	Question prompts
Zydney 2008	-0.18	Synthesis	Conceptual	Science	None	None	Question prompts

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Asterisks indicate articles included in meta-analysis

- Achieve, I. (2013). *Next generation science standards*. Washington: The National Academies Press.
- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: methodological strengths and weaknesses. *Review of Educational Research*, 82(4), 436–476.
- *Ak, Ş. (2015). The role of technology-based scaffolding in problem-based online asynchronous discussion. *British Journal of Educational Technology*, 47(4), 680–693. [10.1111/bjelt.12254](https://doi.org/10.1111/bjelt.12254).
- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26(2), 147–179. doi:[10.1016/S0364-0213\(02\)00061-7](https://doi.org/10.1016/S0364-0213(02)00061-7).
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: a theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4), 439–462. doi:[10.1207/s15327051hci1204_5](https://doi.org/10.1207/s15327051hci1204_5).
- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Allyn & Bacon.
- *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. [10.1023/A:1013995314094](https://doi.org/10.1023/A:1013995314094).
- Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, 40(4), 193–197.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111–127.
- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition—implications for the design of computer-based scaffolds. *Instructional Science*, 33(5–6), 367–379. doi:[10.1007/s11251-005-1272-9](https://doi.org/10.1007/s11251-005-1272-9).
- Babu, G. J. (2012). Bayesian and frequentist approaches. In Online Proceedings of the Astronomical Data Analysis Conference.
- *Barab, S.A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S.J., & Warren, S. (2009). Transformational play as a curricular scaffold: using videogames to support science education. *Journal of Science Education and Technology*, 18, 305–320. [10.1007/s10956-009-9171-5](https://doi.org/10.1007/s10956-009-9171-5).
- Barrows, H. S. (1996). Problem-based learning in medicine and beyond: a brief overview. *New Directions for Teaching and Learning*, 1996(68), 3–12.
- Bayarri, M.J., & Berger, J.O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, 58–80.
- Baylor, A. L., & Kim, Y. (2005). Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education*, 15(2), 95–115.
- Becker, K., & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: a preliminary meta-analysis. *Journal of STEM Education: Innovations and Research*, 12(5/6), 23.
- *Belland, B. R. (2008). Supporting middle school students' construction of evidence-based arguments: Impact of and student interactions with computer-based argumentation scaffolds (Order No. 3330215). Available from ProQuest Dissertations & Theses Full Text. (304502316). Retrieved from <http://search.proquest.com/docview/304502316?accountid=14761>. Accessed 1 Nov 2016.
- Belland, B. R. (2014). Scaffolding: Definition, current debates, and future directions. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 505–518). New York: Springer.
- Belland, B. R., Glazewski, K. D., & Richardson, J. C. (2011). Problem-based learning and argumentation: Testing a scaffolding framework to support middle school students' creation of evidence-based arguments. *Instructional Science*, 39(5), 667–694. doi:[10.1007/s11251-010-9148-z](https://doi.org/10.1007/s11251-010-9148-z).
- Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Journal of Educational Technology & Society*, 18(1), 183–197 Retrieved from <http://www.jstor.org/stable/jeductechsoci.18.1.183>. Accessed 1 Nov 2016.

- Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research*, 87(2), 309–344. doi:10.3102/0034654316670999.
- Beyer, B. K. (1997). *Improving student thinking: a comprehensive approach*. Needham Heights: Allyn & Bacon.
- Biondi-Zoccai, G., Agostoni, P., Abbate, A., D'Ascenzo, F., & Modena, M. G. (2012). Potential pitfalls of meta-analyses of observational studies in cardiovascular research. *Journal of the American College of Cardiology*, 59(3), 292–293.
- Bloom, B. S. (1956). *Handbook I, cognitive domain. Taxonomy of educational objectives: the classification of educational goals*. New York: Longman.
- *Butz, B.P., Duarte, M., & Miller, S.M. (2006). An intelligent tutoring system for circuit analysis. *IEEE Transactions on Education*, 49, 216–223. 10.1109/TE.2006.872407.
- Bybee, R. W. (2010). What is STEM education? *Science*, 329(5995), 996–996.
- 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:10.1046/j.1365-2729.2001.00156.x.
- Chen, C. H., & Bradshaw, A. C. (2007). The effect of web-based question prompts on scaffolding knowledge integration and ill-structured problem solving. *Journal of Research on Technology in Education*, 39(4), 359–375.
- *Chen, Y.S., Kao, T.C., & Sheu, J.P. (2003). A mobile learning system for scaffolding bird watching learning. *Journal of Computer Assisted Learning*, 19, 347–359. 10.1046/j.0266-4909.2003.00036.x.
- Choi, I., Land, S. M., & Turgeon, A. (2008). Instructor modeling and online question prompts for supporting peer-questioning during online discussion. *Journal of Educational Technology Systems*, 36(3), 255–275.
- *Clarebout, G., & Elen, J. (2006). Open learning environments and the impact of a pedagogical agent. *Journal of Educational Computing Research*, 35, 211–226. 10.2190/3UL1-4756-H837-2704.
- *Clark, D.B., Touchman, S., Martinez-Garza, M., Ramirez-Marin, F., & Skjerping Drews, T. (2012). Bilingual language supports in online science inquiry environments. *Computers & Education*, 58(4), 1207–1224. 10.1016/j.compedu.2011.11.019.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: 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: Erlbaum.
- Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: use of diagrams in computer-based training environments. *Instructional Science*, 30(6), 433–464. doi:10.1023/A:1020516301541.
- 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.
- Dellaportas, P., Forster, J. J., & Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing*, 12(1), 27–36.
- Dennen, V. P. (2004). Cognitive apprenticeship in educational practice: research on scaffolding, modeling, mentoring, and coaching as instructional strategies. *Handbook of Research on Educational Communications and Technology*, 2, 813–828.
- 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: Open Universiteit Nederland.
- Douglas, J., Iverson, E., & Kalyandurg, C. (2004). *Engineering in the K-12 classroom: an analysis of current practices and guidelines for the future*. Washington: American Society for Engineering Education.
- Downing, K., Kwong, T., Chan, S. W., Lam, T. F., & Downing, W. K. (2009). Problem-based learning and the development of metacognition. *Higher Education*, 57(5), 609–621.
- Edwards, A. W. F. (1992). *Likelihood*. Baltimore: Johns Hopkins University Press.
- Ellison, A. M. (2004). Bayesian inference in ecology. *Ecology Letters*, 7(6), 509–520.
- *Etheris, A. I., & Tan, S. C. (2004). Computer-supported collaborative problem solving and anchored instruction in a mathematics classroom: an exploratory study. *International Journal of Learning Technology*, 1, 16–36. 10.1504/ijlt.2004.003680.
- Findley, J.L. (2011). *An examination of the efficacy of classical and Bayesian meta-analysis approaches for addressing important meta-analysis objectives* (Doctoral Dissertation). New York: City University of New York. Retrieved from ProQuest Dissertations & Theses Full Text. (Publication Number 3469784).
- Gallagher, S. A., Sher, B. T., Stepien, W. J., & Workman, D. (1995). Implementing problem-based learning in science classrooms. *School Science and Mathematics*, 95(3), 136–146.
- Gallimore, R., & Tharp, R. (1990). Teaching mind in society: teaching, schooling, and literate discourse. *Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology*. (175–205).
- 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 and Development*, 51(1), 21–38. doi:10.1007/BF02504515.

- Ge, X., Chen, C. H., & Davis, K. A. (2005). Scaffolding novice instructional designers' problem-solving processes using question prompts in a web-based learning environment. *Journal of educational computing research*, 33(2), 219–248. doi:[10.2190/5f6j-hhvf-2u2b-8t3g](https://doi.org/10.2190/5f6j-hhvf-2u2b-8t3g).
- *Ge, X., Planas, L. G., & Er, N. (2010). A cognitive support system to scaffold students' problem-based learning in a web-based learning environment. *Interdisciplinary Journal of Problem-Based Learning*, 4(1), 30–56. doi:[10.7771/1541-5015.1093](https://doi.org/10.7771/1541-5015.1093).
- Gelman, A., Carlin, J. B., Stern, H.S., & Rubin, D.B. (2014). *Bayesian data analysis* (vol 2). London: Taylor & Francis.
- Geyer, C. J. (2010). Introduction to Markov chain Monte Carlo. In A. Gelman, S. Brooks, G. Jones, & X. L. Meng (Eds.), *In handbook of Markov chain Monte Carlo: methods and applications* (pp. 3–47). London: CRC Press.
- Greco, T., Zangrillo, A., Biondi-Zoccai, G., & Landoni, G. (2013). Meta-analysis: pitfalls and hints. *Heart, Lung and Vessels*, 5(4), 219–225.
- Guolo, A., & Varin, C. (2015). Random-effects meta-analysis: the number of studies matters. *Statistical Methods in Medical Research*, 0962280215583568, 1–19. doi:[10.1177/0962280215583568](https://doi.org/10.1177/0962280215583568).
- 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: Lawrence Erlbaum Associates.
- Hartung, J., Knapp, G., & Sinha, B. K. (2011). *Statistical meta-analysis with applications* (Vol. 738). Hoboken: John Wiley & Sons.
- Hattie, J. (2008). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hawkins, J., & Pea, R. D. (1987). Tools for bridging the cultures of everyday and scientific thinking. *Journal of Research in Science Teaching*, 24(4), 291–307.
- Hedges, L. V. (1986). Issues in meta-analysis. *Review of Research in Education*, 13, 353–398. doi:[10.2307/1167228](https://doi.org/10.2307/1167228).
- Hershkowitz, R., Schwarz, B.B., & Dreyfus, T. (2001). Abstraction in context: Epistemic actions. *Journal for Research in Mathematics Education*, 195–222.
- Higgins, J., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 137–159.
- Higgins, J., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., et al. (2011). The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928. doi:[10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928).
- 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](https://doi.org/10.1023/B:EDPR.0000034022.16470.f3).
- *Hmelo-Silver, C. E., & Day, R. S. (1999). Contextualized questioning to scaffold learning from simulations. *Computers & Education*, 32, 151–164. doi:[10.1016/S0360-1315\(98\)00062-1](https://doi.org/10.1016/S0360-1315(98)00062-1).
- Hmelo-Silver, C. E., & Ferrari, M. (1997). The problem-based learning tutorial: Cultivating higher order thinking skills. *Journal for the Education of the Gifted*, 20, 401–422. doi:[10.1177/016235329702000405](https://doi.org/10.1177/016235329702000405).
- Hmelo-Silver, C.E., & Lin, X. (2000). Becoming self-directed learners: strategy development in problem-based learning. *Problem-based learning: A research perspective on learning interactions*, 227–250.
- 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](https://doi.org/10.1080/00461520701263368).
- Hoffman, J. L., Wu, H.-K., Krajcik, J. S., & Soloway, E. (2003). The nature of middle school learners' science content understandings with the use of on-line resources. *Journal of Research in Science Teaching*, 40(3), 323–346.
- Hoidn, S., & Kärkkäinen, K. (2014). *Promoting skills for innovation in higher education: a literature review on the effectiveness of problem-based learning and of teaching behaviors*. OECD Education Working Papers, No. 100. Paris: OECD Publishing.
- Hu, D. (2006). *The effects of scaffolding on the performance of students in computer-based concept linking and retention of comprehension* (Unpublished doctoral dissertation). Blacksburg: Virginia Tech.
- Jonassen, D. H. (1999a). Designing constructivist learning environments. *Instructional Design Theories and Models: A New Paradigm of Instructional Theory*, 2, 215–239.
- Jonassen, D. H. (1999b). *Learning with technology: a constructivist perspective*. Englewood Cliffs: Prentice-Hall, Inc..
- Kaweski, D., & Nickeson, M. (1997). C-SHRP Bayesian modeling: a user's guide. Ottawa: Transportation Association of Canada. Retrieved from /view.aspx?id=474116.
- Kay, M., Nelson, G.L., & Hekler, E.B. (2016). Researcher-centered design of statistics: why Bayesian statistics better fit the culture and incentives of HCI. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4521–4532). ACM.
- Kicinski, M. (2013). Publication bias in recent meta-analyses. *PLoS One*, 8(11), e81823.
- Klauenberg, K., Walzel, M., Ebert, B., & Elster, C. (2015). Informative prior distributions for ELISA analyses. *Biostatistics*, online first.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (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](https://doi.org/10.1207/S15327809JLS1204_2).

- Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of meta-analysis in ecology and evolution*. Princeton: Princeton University Press.
- 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.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory Into Practice*, 41(4), 212–218.
- Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Thousand Oaks: Sage Publications.
- Kuenzi, J. J. (2008). Science, technology, engineering, and mathematics (stem) education: background, federal policy, and legislative action. CRS Report for Congress. Retrieved from <http://digitalcommons.unl.edu/crsdocs/35/>. Accessed 1 Nov 2016.
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, 86(1), 42–78. doi:10.3102/0034654315581420.
- Kumar, A. N. (2005). In *Proceedings of the 36th SIGCSE technical symposium on computer science education* (pp. 216–220). New York: ACM. 10.1145/1047344.1047422.
- Lam, P., Doverspike, D., Zhao, J., Zhe, J., & Menzemer, C. (2008). An evaluation of a STEM program for middle school students on learning disability related IEPs. *Journal of STEM education*, 9(1&2), 21–29.
- Lee, C. Y., & Chen, M. P. (2009). A computer game as a context for non-routine mathematical problem solving: the effects of type of question prompt and level of prior knowledge. *Computers & Education*, 52(3), 530–542. doi:10.1016/j.compedu.2008.10.008.
- Leemkuil, H., & de Jong, T. (2012). Adaptive advice in learning with a computer-based knowledge management simulation game. *Academy of Management Learning & Education*, 11, 653–665. doi:10.5465/amle.2010.0141.
- Lewis, M. G., & Nair, N. S. (2015). Review of applications of Bayesian meta-analysis in systematic reviews. *Global Journal of Medicine and Public Health*, 4(1), 1–9.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, 32(3), 131–137.
- Li, S. (2001). *Contingent scaffolding strategies in computer-based learning environments* (Doctoral Dissertation). Bloomington: Indiana University. Retrieved from Retrieved from ProQuest Dissertations & Theses Full Text. (Publication Number 3024299).
- 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(2), 437–455.
- Linn, M. C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, 313(5790), 1049–1050.
- Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research and Development*, 53(1), 5–21 Retrieved from <http://www.jstor.org/stable/30220414>. Accessed 1 Nov 2016.
- Lord, T., & Baviskar, S. (2007). Moving students from information recitation to information understanding-exploiting Bloom's taxonomy in creating science questions. *Journal of College Science Teaching*, 36(5), 40.
- 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.
- Melero, J., Hernández-Leo, D., & Blat, J. (2011). A review of scaffolding approaches in game-based learning environments. *Proceedings of the European Conference on Games Based Learning*, (717–724).
- Mengersen, K. L., Drovandi, C. C., Robert, C. P., Pyne, D. B., & Gore, C. J. (2016). Bayesian estimation of small effects in exercise and sports science. *PLoS One*, 11(4), e0147311.
- Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2), 249–265.
- Newmann, F. M. (1991). Promoting higher order thinking in social studies: overview of a study of 16 high school departments. *Theory & Research in Social Education*, 19(4), 324–340.
- Pata, K., Lehtinen, E., & Sarapuu, T. (2006). Inter-relations of tutor's and peers' scaffolding and decision-making discourse acts. *Instructional Science: An International Journal of Learning and Cognition*, 34(4), 313–341.
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *The Journal of the Learning Sciences*, 13(3), 423–451.
- Pedersen, S., & Liu, M. (2002). The effects of modeling expert cognitive strategies during problem-based learning. *Journal of Educational Computing Research*, 26(4), 353–380.
- Perkins, D. N., & Salomon, G. (2012). Knowledge to go: a motivational and dispositional view of transfer. *Educational Psychologist*, 47(3), 248–258.
- Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: what have we gained and what have we missed? *Educational Psychologist*, 40, 1–12. doi:10.1207/s15326985ep4001_1.

- Puntambekar, S., Stylianou, A., & Hübscher, R. (2003). Improving navigation and learning in hypertext environments with navigable concept maps. *Human-Computer Interaction, 18*, 395–428. doi:10.1207/S15327051HCI1804_3.
- Quintana, C., 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/s15326985sep4004_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. 10.1016/j.compedu.2011.11.010.
- Raphael, L. M., Pressley, M., & Mohan, L. (2008). Engaging instruction in middle school classrooms: an observational study of nine teachers. *The Elementary School Journal, 109*(1), 61–81.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: applications and data analysis methods* (vol 1). Newbury Park: Sage.
- Reiser, B. J. (2004). Scaffolding complex learning: the mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13*(3), 273–304. doi:10.1207/s15327809jls1303_2.
- Robins, J., & Wasserman, L. (2000). Conditioning, likelihood, and coherence: a review of some foundational concepts. *Journal of the American Statistical Association, 95*(452), 1340–1346.
- *Rosen, Y., & Tager, M. (2014). Making student thinking visible through a concept map in computer-based assessment of critical thinking. *Journal of Educational Computing Research, 50*, 249–270. 10.2190/EC.50.2.f.
- Rouinfar, A., Agra, E., Murray, J., Larson, A., Loschky, L.C., & Rebello, N.S. (2014). Influence of visual cueing on students' eye movements while solving physics problems. In *Proceedings of the symposium on eye tracking research and applications* (pp. 191–194). New York: ACM. 10.1145/2578153.2578181.
- Royall, R. M. (1970). Finite population sampling—on labels in estimation. *The Annals of Mathematical Statistics, 41*(5), 1774–1779.
- Sanders, M. (2009). Stem, stem education, stemmania. *The Technology Teacher, 68*(4), 20–26.
- Savery, J. R. (2006). Overview of problem-based learning: definitions and distinctions. *Interdisciplinary Journal of Problem-Based Learning, 1*(1), 9–20. doi:10.7771/1541-5015.1002.
- Savery, J. R. (2015). Overview of problem-based learning: definitions and distinctions. In A. Walker, H. Leary, C. E. Hmelo-Silver, P. A. Ertmer, & P. A. (Eds.), *Essential readings in problem-based learning: exploring and extending the legacy of Howard S. Barrows* (pp. 5–15). West Lafayette: Purdue University Press.
- Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments. *Educational Technology Research and Development, 50*(3), 77–96. doi:10.1007/BF02505026.
- Scardamalia, M., Bereiter, C., McLean, R. S., Swallow, J., & Woodruff, E. (1989). Computer-supported intentional learning environments. *Journal of Educational Computing Research, 5*(1), 51–68.
- Schmid, C. H., & Mengersen, K. (2013). Bayesian meta-analysis. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *The handbook of meta-analysis in ecology and evolution* (pp. 174–191). Princeton: Princeton University Press.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando: Academic Press.
- 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 & Distance Learning, 13*(3), 206–227.
- Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). *Meta-analysis with R*. New York, NY: Springer.
- *Siler, S., Mowery, D., Magaro, C., Willows, K., & Klahr, D. (2010). Comparison of a computer-based to hands-on lesson in experimental design. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent Tutoring Systems* (pp. 408–410). Berlin: Springer.
- *Simons, K. D., & Klein, J. D. (2007). The impact of scaffolding and student achievement levels in a problem-based learning environment. *Instructional Science, 35*(1), 41–72. 10.1007/s11251-006-9002-5.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583–639.
- Spiegelhalter, D.J., Abrams, K.R., & Myles, J.P. (2004). Bayesian approaches to clinical trials and health-care evaluation (vol 13). Chichester: John Wiley & Sons.
- *Stark, D.M. (2013). *Ill-structured problems, scaffolding and problem-solving ability of novice nursing students* (PhD Thesis). Minneapolis: Capella University. Retrieved from ProQuest Dissertations & Theses Full Text. (Publication Number 3553778).
- 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.

- *Su, Y. (2007). *The impact of scaffolding type and prior knowledge in a hypermedia, problem-based learning environment* (PhD Thesis). Tempe: Arizona State University. Retrieved from ProQuest Dissertations & Theses Full Text. (Publication Number 3288016). Retrieved from <http://dist.lib.usu.edu/login?url=https://search.proquest.com/docview/304896264?accountid=14761>. Accessed 1 Nov 2016.
- *Su, Y., & Klein, J.D. (2010). Using scaffolds in problem-based hypermedia. *Journal of Educational Multimedia and Hypermedia*, 19(3), 327–347.
- Sun, D., & Ye, K. (1995). Reference prior Bayesian analysis for normal mean products. *Journal of the American Statistical Association*, 90(430), 589–597.
- Sungur, S., & Tekkaya, C. (2006). Effects of problem-based learning and traditional instruction on self-regulated learning. *The Journal of Educational Research*, 99(5), 307–320.
- Sutton, A. J., & Abrams, K. R. (2001). Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research*, 10(4), 277–303.
- Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, 71(2), 321–363.
- 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
- *Toth, E.E., Suthers, D.D., & Lesgold, A.M. (2002). “Mapping to know”: the effects of representational guidance and reflective assessment on scientific inquiry. *Science Education*, 86, 264–286. 10.1002/sce.10004.
- Tuckman, B. W., & Schouwenburg, H. C. (2004). Behavioral interventions for reducing procrastination among university students. In H. C. Schouwenburg, C. H. Lay, T. A. Pychyl, & J. R. Ferrari (Eds.), *Counseling the procrastinator in academic settings* (pp. 91–103). Washington: American Psychological Association.
- Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., & Higgins, J. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, 34(6), 984–998.
- Valentine, J. C., Pigott, T. D., & Rothstein, H. R. (2010). You need? A primer on statistical power for meta-analysis. *Journal of Educational and Behavioral Statistics*, 35(3), 215–247. doi:10.3102/1076998609346961.
- Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: a decade of research. *Educational Psychology Review*, 22(3), 271–296.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- *Weusijana, B.K.A., Riesbeck, C.K., & Walsh, J.T., Jr (2004). Fostering reflection with Socratic Tutoring Software: results of using inquiry teaching strategies with web-based HCI techniques. In Proceedings of the 6th International Conference on Learning Sciences (pp. 561–567). Santa Monica, California: International Society of the Learning Sciences. Retrieved from <http://dl.acm.org/citation.cfm?id=1149126.1149195>. Accessed 1 Nov 2016.
- Wood, D., Bruner, J. S., & 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.
- Wu, H. (2010). *Scaffolding in technology-enhanced science education* (PhD Thesis). College Station: Texas A&M University. Retrieved from ProQuest Dissertations & Theses Full Text. (Publication Number 741013672).
- Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education: current and future directions. *Studies in Science Education*, 49(2), 121–169.
- Zydney, J. M. (2005). Eighth-grade students defining complex problems: the effectiveness of scaffolding in a multimedia program. *Journal of Educational Multimedia and Hypermedia*, 14, 61–90.
- *Zydney, J.M. (2008). Cognitive tools for scaffolding students defining an ill-structured problem. *Journal of Educational Computing Research*, 38, 353–385. 10.2190/EC.38.4.a.