Customized Scaffolding for Pre-service Teachers' Problem-Solving in STEM

Brian R. Belland, ChanMin Kim, & Eunseo Lee

The Pennsylvania State University

As calls increase that all students learn computer science (Grover & Pea, 2013; K-12 Computer Science Framework Steering Committee, 2016; Lye & Koh, 2014), it is critical to investigate how to prepare teachers to succeed in this environment. Simply telling teachers to teach computer science is not likely to be effective if they do not already have the necessary content knowledge. In addition, doing so would impinge on teachers' autonomy, which in turn could lead to low motivation (Deci & Ryan, 2000). Rather, it is important to help teachers learn computer science content in an effective, problem-centered manner (Lye & Koh, 2014). This, in turn, can help model effective methods of teaching computer science. But it is not enough to simply give authentic problems to students. To the contrary, one needs to provide support for problem solving. This can be done with scaffolding, defined as dynamic support provided by teachers, computer-based tools, or peers, that helps students perform and gain skill at performing problem solving tasks (Wood et al. 1976). In particular, when students are challenged with ill-structured problems with no clear solutions or solution path, scaffolding can address students' learning and performance needs based on conceptual, metacognitive, strategic, and motivational perspectives (Hannafin et al., 1999). Scaffolding has been integrated into instruction in a wide range of educational levels (e.g., K-12, university, graduate, and adult) and disciplines (e.g., science, technology, engineering, mathematics, social studies) (Duffy & Azevedo, 2015; Poitras & Lajoie, 2014). This paper focuses on identifying the most effective combinations of scaffolding features and contexts of use within computer science education at the college and graduate school levels. This is done through clustering an expansion of a scaffolding meta-analysis dataset (Belland, Walker, Kim, et al., 2017).

Scaffolding

Meta-analyses of between-subjects and within-subjects effects indicated that students who use scaffolding in the context of problem solving outperform students who receive lecture (Belland et al., 2015; Belland, Walker, & Kim, 2017; Belland, Walker, Kim, et al., 2017). While students from all education populations and education levels benefit from scaffolding, there are significant differences in effect sizes resulting from scaffolding across education levels and learner populations (Belland, Walker, & Kim, 2017; Belland, Walker, Kim, et al., 2017). This suggests that, within published research on scaffolding, scaffolding features are not optimized to individual student characteristics. Furthermore, scaffolding features often are used in different combinations, and such different combinations lead to different effects in different fields. Prior work focused on calibrating classifiers to accurately predict expected effect size of scaffolding schemes for students (Kim et al., 2017). This paper extends the line of research by exploring and describing possible relationships among student groups with different scaffolding effect sizes and scaffolding characteristics variables using two-step clustering. Because scaffolding feature combinations and the context in which they are used, and the efficiacy of such, often vary according to discipline and education level, in this paper, we focus on clustering scaffolding research in the technology and engineering fields at the college and graduate school levels.

Literature Review

Problem-centered Instruction and Scaffolding

With the rise of automation and the explosion of information access, the workforce of the future requires workers who can think critically about information and address complex problems (National Academies of Sciences, Engineering, and Medicine, 2017). A promising way to help students gain these skills is through engaging them in ill-structured problem solving with the assistance of scaffolding. In its original form, scaffolding was delivered one-on-one by a teacher who dynamically assessed the student's performance characteristics (Wood et al., 1976). Computer-based scaffolding can help overcome the issue of high student-teacher ratios in K-12 contexts, which precludes teachers from working one-on-one with the same student for an entire class period (Hawkins & Pea, 1987; Saye & Brush, 2002). Computer-based scaffolding is highly effective, with students who address authentic problems while supported by computer-based scaffolding performing at least 0.4 standard deviations better than students who receive lecture across concept, principles, and application level assessments (Belland, Walker, Kim, et al., 2017).

One way to conceptualize scaffolding is as a tool that aids current performance and enables skill gain by removing complexity that is not, and highlights complexity that is, central to learning the target skill (Reiser, 2004). Scaffolding does this through a combination of cognitive and motivational support (Belland et al., 2013). Scaffolding can be seen as taking three forms - conceptual (help with things to consider while solving the problem), strategic (bootstrapping an overall strategy for solving the problem, and metacognitive (helping students question their own understanding and processes) (Hannafin et al., 1999).

Variation in Scaffolding Effectiveness

While highly effective as a whole, some learner populations benefit more from computer-based scaffolding than others (Belland, Walker, & Kim, 2017; Belland, Walker, Kim, et al., 2017). For example, the within-subjects effects resulting from scaffolding have the highest probability of the best among college- and graduate-level learners, and among elementary special education students (Belland, Walker, & Kim, 2017). The between-subjects effects resulting from scaffolding are stronger among traditional students than among lower-achieving students, and stronger among adult learners than among elementary, middle, secondary, college, and graduate-level learners (Belland, Walker, Kim, et al., 2017).

Possible reasons for variation in scaffolding effectiveness. As Wood et al. (1976) noted, scaffolding needs to be “generated by the interaction of the tutor’s two theories” – namely “a theory of the task or problem and how it may be completed” and “a theory of the performance characteristics of [the] tutee” (p. 97). But computer-based scaffolding is most often not generated according to an interaction of a theory of the task and a theory of the tutee (Pea, 2004; Puntambekar & Hübscher, 2005). Meta-analyses indicate that such limited customization does not improve learning (Belland et al., 2015; Belland, Walker, Kim, et al., 2017). Poor performance of computer-based scaffolding customization is in part due to the fact that the logic of all existing methods is centered in strict pattern matching. That is, within many existing computer-based scaffolding systems, when students perform a certain sequence of actions or respond to a closed-ended quiz in a particular way, the actions are compared to a set of actions associated with adding or fading scaffolding and then the system triggers the fading or adding of scaffolding based on the degree of whether the actions match exactly or not. Other computer-based scaffolding systems set scaffolding to fade based on fixed time intervals or when students click a button that says that they do not need any more help. To develop a computer-based scaffolding system that goes beyond scaffolding customization on simple pattern matching, it is necessary to set up conditions by which scaffolding systems can produce output by matching patterns in actions of current users with patterns of data accumulated over time (Chen et al., 2016; Sun, 2013; Theodoridis, 2015).

Another possible reason is that scaffolding design is often not driven by systematic synthesis of research results. This may be because, until recently, there were few systematic syntheses of the scaffolding literature. Rather, there were many conceptual frameworks meant to guide scaffolding design, but these often reflected the theoretical positions of the authors. Scaffolding is underpinned by a variety of theories, including cultural historical activity theory (Leont'ev, 1974; Luria, 1976), ACT-R (Anderson et al., 1997), and knowledge integration (Linn et al., 2003). Scaffolding design is often informed by the premises of one of the theories or a particular line of research, rather than driven by systematic empirical evidence gathered from the overall scaffolding literature.

Another reason is that scaffolding form, and the context in which it is used, often varies considerably. For example, scaffolding is used in the context of many different problem-centered instructional models, including problem-based learning, project-based learning, design-based learning, modeling/visualization, and inquiry-based learning.

Cluster Analysis and Scaffolding Customization

While scaffolding is highly effective in general, the impact often varies in magnitude across education levels, disciplines, and types of context in which scaffolding is used. For example, previous meta-analysis results showed that effectiveness was stronger among adult learners than among learners at other education levels and when used in project-based learning than when used in other types of instructional approaches (Belland, Walker, Kim, et al., 2017).

Inconsistency in scaffolding effectiveness results in part from the poor performance of computer-based scaffolding customization. Within current scaffolding systems, dynamic assessment is limited in that students' performance is assessed by simply matching their response pattern with the pre-set standards (Puntambekar & Hübscher, 2005; Reiser, 2004). Scaffolding customization is also restricted because it is conducted by simply adjusting scaffolding intensity or frequency in a linear manner (Koedinger & Aleven, 2007; VanLehn, 2011). For more effective scaffolding customization, there is a need to explore conditions, based on accumulated empirical data, by which the systems can provide optimal scaffolding solutions.

Cluster analysis helps to identify the most effective combinations of scaffolding characteristics for target populations. The clustering algorithm can determine the natural groupings of scaffolding features out of accumulated empirical data that had actual effects on improving learning (Aldenderfer & Blashfield, 1984). This clustering solution can help computer-based scaffolding systems to navigate best scaffolding strategies among different combinations of scaffolding characteristics based on not only learner's simple response patterns but also the tendency of co-occurrence of other attributes such as education levels, context of use, and their associated effect size (Baker & Inventado, 2014).

Research Questions

1. What combinations of scaffolding characteristics for problem-centered learning lead to medium and large effect sizes among college and graduate-level learners in the technology and engineering disciplines?
2. How are the combinations of scaffolding characteristics related to their context of use?

Method

Data Source

The dataset used in this study is from a meta-analysis that synthesized the results of 144 studies on the effects of computer-based scaffolding on students' cognitive learning outcomes in STEM education (Belland, Walker, Kim, et al., 2017). The 333 coded outcomes were expanded to 13,638 cases, with each case representing one participant from each study, and each case being assigned the mean effect size for its study.

The original dataset consisted of four coding categories as moderators including scaffolding characteristics, study characteristics, student characteristics, and assessment characteristics and their sub-categories. For this study, the Hedge's g effect size was transformed into a categorical variable with four levels: "large" (0.8 or greater), "medium" (0.5 to 0.8), "small" (0.2 to 0.5), and "no effect" (less than 0.2) (Cohen, 1969). The initial dataset included studies ranging from elementary-adult levels, as well as in the context of science, technology, engineering, and mathematics education. Since we were interested in the most effective combinations of scaffolding features in college and graduate level education in the technology and engineering disciplines, we downselected data accordingly. Specifically, we selected 'college' and 'graduate' sub-categories data including 7,294 cases. Then, we used data from the studies whose results showed at least medium effect size in the 'technology' and 'engineering' disciplines, resulting in 1,726 cases in the final dataset (see Table 1).

Analysis

We explored scaffolding clusters using two-step cluster analysis available in SPSS 24 to identify distinct groups of scaffolding attributes tailored to teacher learning of computer science at the undergradute and graduate levels. Two-step cluster analysis is a clustering procedure that combines the principles of hierarchical and partitioning methods. The algorithm first carries out a process similar to the k-means algorithm. Based on these results, a modified hierarchical agglomerative clustering procedure is conducted that combines the objects sequentially to form homogeneous clusters (Mooi & Sarstedt, 2011). This method is particularly useful when handling categorical or mixed scale variables and it allowed us to draw meaningful outcomes from our multivariate categorical data.

Table 1. Variables used in cluster analysis

Categories	Attributes	Sub-categories
Scaffolding Characteristics	Scaffolding Intervention	Conceptual, Metacognitive, Strategic, Motivation
	Scaffolding Intended Outcomes	Higher-order Thinking Skills, Knowledge Integration, Enhance Motivation
	Scaffolding Strategy	Specific, Generic
	Scaffolding Change	Fading, Adding, Fading/Adding, No Change
	Scaffolding Schedule	Performance-adapted, Fixed Time Interval, Self-selected, No Schedule
	Education Level	College, Graduate
Study Characteristicss	Context of Use	Project-based Learning, Problem-based Learning, Problem-solving, Modeling/Visualization, Project-based Learning, Learning by Design, Case-based Learning, Inquiry-based Learning
	General Disciplines	Technology, Engineering
Assessment Characteristics	Assessment Level	Principles, Concept, Application
Outcome	Effect Size	Large, Medium

Scaffolding characteristics variables were the main input variables used to find cluster solutions. In addition to that, we used ‘context of use’ variable to see what combinations of scaffolding characteristics are effective under which instructional contexts, including project-based learning, inquiry-based learning, problem-based learning, etc.

Last, we assessed the stability of the cluster solution by checking the replicability of a cluster solution across different samples from the same dataset. For this, we split the sample into two parts and performed the same cluster analysis method on both parts and checked if the solutions are similar in terms of the number and characteristics of the clusters.

Results

Model Fit

Two-step cluster analysis in SPSS offers an overall goodness-of-fit measure of the cluster model called *silhouette measure of cohesion and separation*. It is based on average distances between objects and can vary between -1 and +1, with the values of more than 0.50 indicating a good solution. Our clustering model showed a value of 0.62, which indicates good fit.

Figure 1 shows each input variable's importance in determining clusters. The importance of major scaffolding characteristics variables is evenly distributed, indicating this output is not biased in terms of predicting each cluster membership.

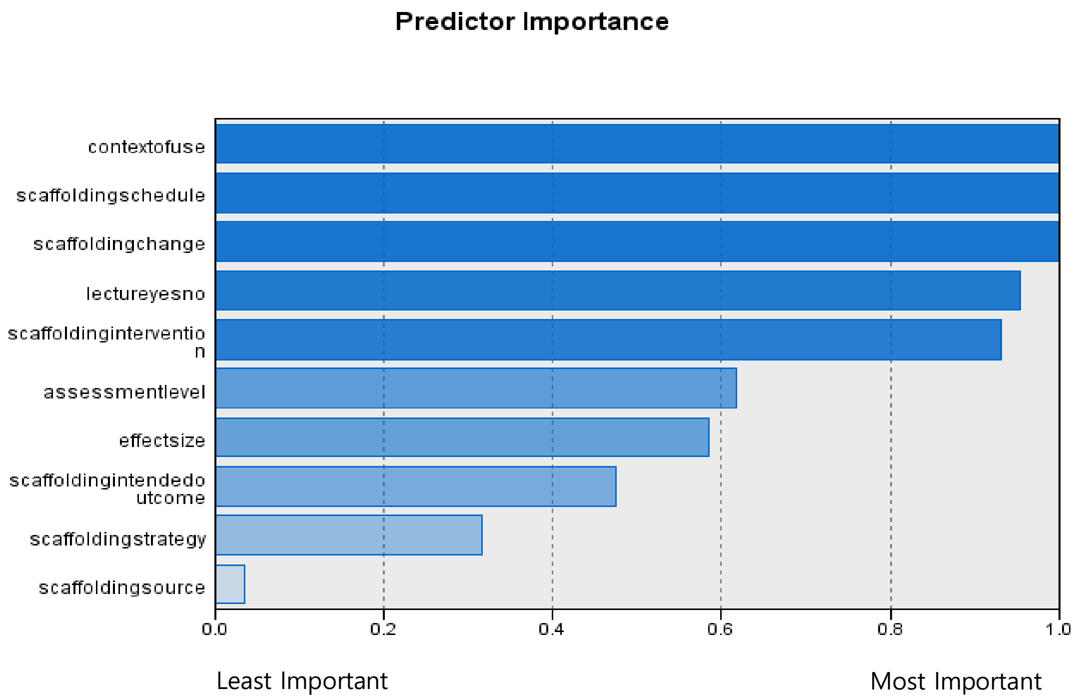
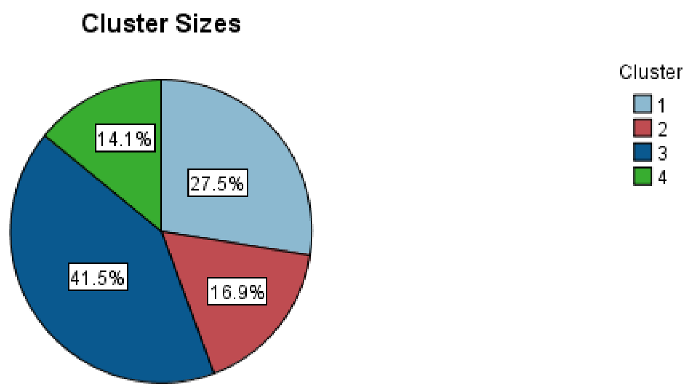


Figure 1. Predictor importance

Number of Clusters

Two-step cluster analysis allows researchers to specify the number of clusters. We determined the optimal number of clusters based on the three criteria. First, outcomes needed to allow us to have enough combinations capturing subtle differences between segments. Next, the ratio of cluster sizes–the size of the largest cluster to the smallest cluster– needed to be less than three so that the smallest cluster can have enough representativeness in comparison with the largest one. Last, the stability of cluster outcomes where the cluster membership of individuals does not change or only changes little when different clustering methods are used (Mooi & Sarstedt, 2011). The optimal number of clusters in our solution turned out to be four (see Figure 2). This number of clusters prevented a few binary variables from acting as swamping variables and hindering drawing meaningful outcomes. Also, each of the four clusters demonstrated a stable distribution with sufficient sample size and the cluster membership of individuals changed little under the repetitive clustering procedure.



Size of Smallest Cluster	244 (14.1%)
Size of Largest Cluster	717 (41.5%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	2.94

Figure 2. Number of clusters and sizes

Additionally, we also used two R packages - 'cluster' and 'klaR' - to validate our decision on the number of clusters and cluster outcomes. The 'cluster' package is used to calculate the dissimilarity matrix (Maechler et al., 2019) and 'klaR' is used to run the k-modes clustering algorithm (Roever et al., 2020). We also used the 'stats' (*R: The R Stats Package*, n.d.) and 'RColorBrewer' (Neuwirth, 2014) packages for visualization via heatmaps and color effect adjustment. The same eight categorical variables used in SPSS analysis (i.e., scaffolding change, scaffolding intended outcome, scaffolding intervention, scaffolding strategy, scaffolding schedule, context of use, assessment level, and effect size) were included in R analysis.

Using the elbow method, we found the optimal number of clusters to be 8. Figure 3 shows the change in total within-cluster variation as the number of clusters increases. Total within-cluster difference score indicates how similar observations are in a designated cluster. A low within-cluster difference indicates that the observations in a cluster share similar characteristics and result in a more compact cluster. When cluster number is equal to eight, the total within-cluster difference reaches the lowest point. K-modes algorithm assign each observation to a cluster based on the same criterion.

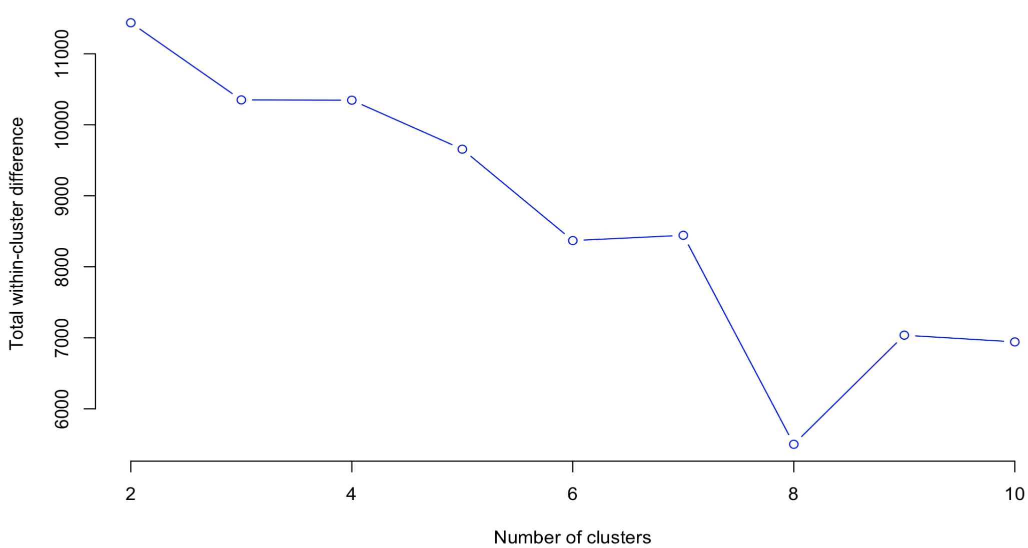


Figure 3. Elbow Plot

Even though eight clusters were able to catch subtle variations within the dataset better, it also made it complicated to interpret clustering outcomes and their relationships. To better understand the relationship between variables among different clusters and to get more meaningful implications in terms of scaffolding use, we tried to narrow down the numbers and identify significant patterns out of eight clusters. By comparing and contrasting each clusters' profile, we were able to identify four distinct cluster patterns in terms of the combination of variable dimensions within each cluster, which were similar results with the ones from SPSS analysis. So, while presenting and discussing the eight clustering outputs in detail in the next section, we were also able to combine clusters that had almost identical patterns but a small difference and compare and contrast subtle differences among them and draw out meaningful insights in terms of scaffolding use.

Profile and Visulization of Cluster Outputs

In this section, we profile and visualize eight cluster outputs from R analysis and compare with four cluster outputs from SPSS analysis to understand the relationships among variables that may affect the effectiveness of scaffolding. As shown in Figure 4, our cluster outputs reflect multi-dimensional characteristics of the clusters, showing relative distributions along the dimension of input variables. Specifically, eight cluster outputs are on the x-axis and eight different variables are presented on the y-axis of each heat map and each variable has different dimensions. Different shades of blue represent the different frequencies of observations, where the darker the color of the grid within the map, the more observations are in that dimension of the variable.

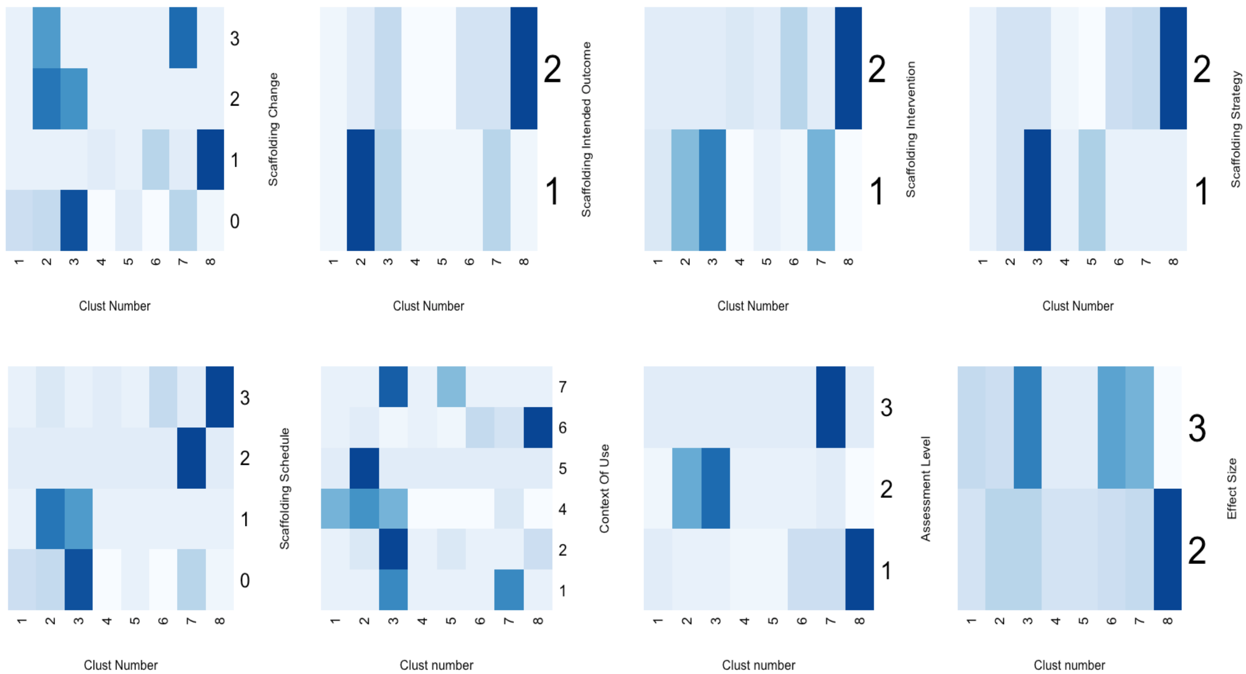


Figure 4. Heat map of Cluster Outputs

Among the eight cluster outputs, clusters 4, 6, and 8 have very similar characteristics with the only difference being the composition of the assessment level variable (See Table 2). This set of clusters has the largest frequency of observation (950 out of 1726 observed cases (55%) and represent a relatively clearer cluster center and more distinct characteristics than other clusters. Adding strategic support based on students' performance is predominant compared to other elements within each variable. Also, higher order thinking skills as a scaffold's intended outcome and scaffold that is tailored to specific content (compared to general support) are the predominant elements of this group of clusters. So, in general, the profile of the clusters indicate that adding support in terms of students' strategy use on the basis of their performance has at least medium size effects on improving their higher order thinking skills in problem-solving contexts. This result is matched with cluster 3 from the SPSS analysis results (Table 3). In terms of assessment level, both concept and principles level of assessment present in this group of clusters, indicating that implementing this type of scaffolding can have a significant effect either

Table 2. Profile of cluster outputs from R analysis

scaffolding change	1 (N=71)	2(N=182)	3(N=277)	4(N=34)	5(N=33)	6(N=179)	7(N=213)	8(N=737)
None	71(100%)	78(43%)	256(92%)	0(0%)	33(100%)	0(0%)	95(45%)	20(3%)
Adding	0(0%)	0(0%)	0(0%)	34(100%)	0(0%)	179(100%)	16(8%)	717(97%)
Fading	0(0%)	26(41%)	21(8%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Fading/adding	0(0%)	78(66%)	0(0%)	0(0%)	0(0%)	0(0%)	102(48%)	0(0%)
intended outcome	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
knowledge integration	0(0%)	64(35%)	21(8%)	0(0%)	0(0%	0(0%)	20(9%)	0(0%)
higherorderskills	71(100%)	118(65%)	256(92%)	34(100%)	33(100%)	179(100%)	193(91%)	737(100%)
intervention	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
conceptual(content)	65(92%)	182(100%)	277(100%)	0(0%)	33(100%)	20(11%)	198(93%)	0(0%)
strategic(process)	6(8%)	0(0%)	0(0%)	34(100%)	0(0%)	159(89%)	15(7%)	737(100%)
strategy	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
generic	0(0%)	13(7%)	110(38%)	0(0%)	33(100%)	0(0%)	0(0%)	0(0%)
specific	71(100%)	169(93%)	176(62%	34(100%)	0(0%)	179(100%)	213(100%)	737(100%)
schedule	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
none	71(100%)	78(43%)	256(92%)	0(0%)	33(100%)	0(0%)	95(45%)	20(3%)
fixed	0(0%)	26(14%)	21(8%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Self-selected	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	82(38%)	0(0%)
Performance-adapted	0(0%)	78(43%)	0(0%)	34(100%)	0(0%)	179(100%)	36(17%)	717(97%)
context of use	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
CBL	0(0%)	0(0%)	77(28%)	0(0%)	0(0%)	0(0%)	80(38%)	0(0%)
IBL	0(0%)	13(7%)	110(40%)	0(0%)	13(39%)	0(0%)	0(0%)	20(3%)
Learning by desiggn	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%	0(0%)	0(0%)
Modeling/visualization	51(72%)	67(37%)	50(18%)	0(0%)	0(0%)	0(0%)	16(8%)	0(0%)
Problem-based learning	0(0%)	58(32%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)
Problem-solving	20(28%)	44(24%)	0(0%)	34(100%)	0(0%)	179(100%	117(55%)	717(97%)
Project-based learning	0(0%)	0(0%)	40(14%)	0(0%)	20(61%)	0(0%)	0(0%)	0(0%)
assessment level	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
concept	51(72%)	13(7%)	21(8%	0(0%)	0(0%)	153(85%)	162(76%)	737(100%)
principles	20(28%)	169(93%)	256(92%)	34(100%)	33(100%)	26(15%)	36(17%)	0(0%)
application	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	0(0%)	15(7%)	0(0%)
effect size	**1 (N=71)**	**2(N=182)**	**3(N=277)**	**4(N=34)**	**5(N=33)**	**6(N=179)**	**7(N=213)**	**8(N=737)**
medium	6(8%)	124(68%)	110(40%)	0(0%)	0(0%)	46(26%)	100(47%)	737(100%)
large	65(92%)	58(32%)	167(60%)	34(100%	33(100%)	133(74%)	113(53%)	0(0%)

when the learning outcome is measured on the basic knowledge level or when it is measured on rule application (transfer) level.

Table 3. Profile of cluster outputs from SPSS analysis

	Cluster 3	Cluster 1	Cluster 2	Cluster 4
Sample Size	717 (41.5%)	474 (27.5%)	291 (16.9%)	244 (14.1%)
Context of Use	Problem-solving (100)	CBL (33.1) IBL (28.7) M&V (25.5) Project-based (12.7)	Problem-solving (94.5) M&V (5.5)	Problem-solving (48.8) PBL (23.8) M&V (19.3) IBL (8.2)
Scaffolding Change	adding (100)	none (100)	adding (71.8) fading adding (28.2)	fading adding (40.2) none (32.4) fading (19.3) adding (8.2)
Scaffolding Schedule	performance adapted (100)	none (100)	performance adapted (71.8) self-selected (28.2)	performance adapted (48.4) none (32.4) fixed (19.3)
Scaffolding Intervention	strategic (100)	conceptual (100)	strategic (66.3) conceptual (33.7)	conceptual (83.2) strategic (16.8)
Assessment Level	Concept (100)	Principles (70.9) Concept (29.1)	Concept (73.9) Principles (26.1)	Principles (66.4) Concept (27.5) Application (6.1)
Effect Size	medium (100)	medium (52.7) large (47.3)	large (91.1) medium (8.9)	medium (53.3) large (46.7)
Scaffolding Intended Outcomes	higher order thinking skills (100)	higher order thinking skills (100)	higher order thinking skills (100)	higher order thinking skills (57.0) knowledge integration (43.0)
Scaffolding Strategy	specific (100)	specific (67.1) generic (32.9)	specific (100)	specific (100)

Note: CBL: Case-Based Learning, IBL: Inquiry-Based Learning, PBL: Problem-Based Learning, M&V: Modeling & Visualization.

As a specific example, Corbett and Anderson (2001) examined scaffolding conditions in relation to problem solving performance and learning under intelligent tutoring systems. Undergraduates participated in a programming course where they learned programming knowledge and process while working at their own pace with cognitive tutor and completed tests which measure their programming related concepts (e.g., code evaluation, code debugging, and code generation). The student group who was given immediate feedback when they typed an incorrect symbol performed best on tests and completed the problems fastest. This result shows that providing immediate feedback based on students' performance can be efficient and effective tutorial support for learning a complex problem solving skill such as programming.

On the contrary, cluster 7 suggests that, for the same intended outcome of scaffolding (i.e., higher order skills), the similar effect size can be achieved through making some variations and giving some flexibility to scaffolding change and schedule. Indeed, cluster 2 from SPSS analysis has a similar pattern and implication (Table 3), which shows that adding or fading/adding support adjustment or even scaffolding without any scheduled change can be done effectively based on the students' performance or by their own request. This type of scaffolding was proposed in Yin et al. (2013) that is intended to promote students' learning conceptual knowledge using participatory simulation on mobile devices. In the study, students were taught the rules of the sorting algorithms and asked to simulate the sorting algorithms and master it at a certain level. The system on the mobile device assisted students if they needed help by providing information about where the mistakes are and how to correct them as well as additional teacher's help based on students' inquiry. As students were gradually able to understand the methods and strategies and have better conceptual knowledge, the system reduced the help gradually, and students were required to solve the problem themselves. This adding/fading support adjustment based on students' performance and their request brought about significantly higher accuracy rates in students performing the complicated sorting algorithms.

Besides, Yeh et al. (2010) showed in their study that similar scaffolding condition demonstrated large effect size not only when the assessment is conducted in the basic knowledge level, but also in more adaptive principles level (e.g., knowledge transfer). In the study, they investigated the effects of different formats of self-explaining prompts as a scaffolding based on learners' knowledge level when learning with dynamic multimedia materials. Then, they measured learning outcomes based on three categories which include declarative knowledge, near-transfer, and far-transfer. In the transfer measure, students were asked to apply rules and algorithms they've learned to similar situations. The results confirmed that adaptive self-explaining prompts can serve as effective scaffolding not only for improving conceptual knowledge but also enhancing knowledge transfer.

Next, clusters 1, 3, and 5 together have distinct characteristics compared to other clusters in that they demonstrate that conceptual scaffolds— providing targeted hints and prompts about the content or helping learners to structure their content knowledge through a concept map— can be effective for higher order thinking skills without any customization during the intervention under certain circumstances. This group of clusters shows a similar pattern with cluster 1 from SPSS analysis results. Specifically, in this case, either the fundamental nature of the scaffold or the frequency of it does not change in response to anything. This intervention contrasts with other interventions where the nature and the frequency of scaffold change based on student's performance, request, or pre-defined (fixed) number. According to the profile of the clusters, this type of scaffolding change and schedule is effectively applicable to both cases where the elements of scaffolding is adapted to specific content (e.g., feedback for engineering principles (Rodriguez et al., 2006)) and where the element is not tailored to specific content, thus more generic (e.g., instruction for information evaluation (Wiley et al., 2009)). This scaffolding characteristic can be utilized with various problem-centered instructions including case-based learning and inquiry-based learning.

For instance, the intervention used in the Demetriadis et al. (2008) study exemplifies this type of scaffold. In this study, elaborative question prompts were used as a scaffolding intervention to activate students' context-generating cognitive processes during case-based learning. Scaffolded student group (i.e., prompted to consistently answer a set of questions meant to engage them in deep information processing) performed significantly better in both domain knowledge acquisition and knowledge transfer tests. The scaffolding was not adapted for customization in a sense that the same set of question prompts were persistently presented to every student each time they navigated to a new case and answered to scenario questions.

Wiley et al. (2009) utilized similar scaffolding characteristics in their study under the inquiry-based learning situation. They used an instructional unit (named SEEK) as a scaffolding intervention to help students to evaluate the reliability of information sources from multiple internet websites during a science inquiry task. The SEEK instructional unit consisted of declarative information about source evaluation, evaluation template, and expert feedback and this pre-set unit was presented consistently to the students over the session without any customization. Students who used SEEK scaffold displayed greater performance in their reliability judgments of information sources than students who did not.

Last, cluster 2 indicates the effectiveness of scaffolding across levels of different dimensions of variables, suggesting that scaffolding intervention can be designed in different ways while remaining highly effective. This observation aligns with previous meta-analysis results that showed no difference in effect sizes between scaffoldings that include fading, adding, adding/fading, or no customization and their adjustment logic by which scaffolding change is implemented (Belland, Walker, Kim, et al., 2017). This observation is also evidenced by several empirical studies.

Butz et al. (2006) demonstrated that a combination of adding and fading scaffolding based on students' performance can have a large effect size in association with developing higher order thinking skills during problem-based learning. They used an interactive multimedia intelligent tutoring system that can modify the instructional sequence and the amount of detail presented to students according to their weakness. This intelligent tutoring system was designed to assist electrical engineering undergraduate students in solving a real-life problem. The students who were scaffolded using this system scored higher on their performance measures than the students who were not.

Also, the combination of adding and fading support based on students' performance showed at least medium effect size when the scaffolding targets to improve knowledge integration (i.e., integration of new knowledge with existing mental models) for a deep understanding of content during simulation based learning. Kumar et al. (2007) used, in their study, conversational agents (i.e., CycleTalk) as a form of dynamic learning support to enhance students' knowledge related to the thermodynamic class. Students were instructed to build their designs using the simulation software and evaluate their design, meant to have them learn related concepts while doing the activity. The CycleTalk agent provided the knowledge construction dialogues, at the same time monitored the conversation pattern and provided hints and prompts during the conversation. As a result, dynamic support implemented with tutorial dialogue agents brought about significantly more learning among students.

General Findings Across Cluster Outputs

When examining the 8-cluster and 4-cluster solutions, one notices several trends. First, scaffolding that leads to medium or large effect sizes almost exclusively is intended to improve higher-order thinking skills, rather than knowledge integration.

In addition, in the 8-cluster solution, all clusters include a dominant percentage of context-specific scaffolds except for cluster 5. Furthermore, all 4 clusters in the 4-cluster solution used predominantly context-specific scaffolding. However, this does not indicate effective scaffolding solutions should opt for context-specific scaffolds over the generic scaffolds. Because the vast majority of data selected (1570 out of 1726 cases) were associated with context-specific scaffolds, this result rather implicates that context-specific scaffold is the most frequently used strategy in technology and engineering disciplines.

Overall, conceptual and strategic scaffolding interventions appear with similar frequency, reflecting equal importance in achieving the intended outcome. That means not only supporting students to acquire content knowledge to be used for problem solving but also providing guidance from the strategy use perspective are critical elements for successful problem-centered instruction.

In the 8-cluster solution, all clusters except for clusters 1 and 5 are associated with a dynamically changing scaffold that includes predominantly adding, or a combination of fading and adding support. This observation implies that, with no difference in effect sizes, designers can customize scaffolding in conjunction with the frequently co-occurring scaffolding adjustment logic in technology and engineering classes, which was students' performance followed by students' self-selection.

Also of interest is that scaffolding appeared to be most effective when it was either not customized, or customized on the basis of performance.

Discussion

Scaffolding is an intervention that leads to stronger learning than lecture across learner populations, and education and assessment levels (Belland, 2017; Belland, Walker, & Kim, 2017; Belland, Walker, Kim, et al., 2017; Ma et al., 2014; Steenbergen-Hu & Cooper, 2013, 2013; VanLehn, 2011). Furthermore, it leads to strong within-subjects gains across learner populations, and education and assessment levels (Belland, Walker, & Kim, 2017). But it is not useful to simply tell a designer to use scaffolding, as scaffolding has so many variations, and can be used in the context of many different problem-centered instructional models and STEM disciplines. Rather, it is important to consider what combination of scaffolding features are most effective for which students in which conditions. Many researchers endeavor to find such an ideal combination of scaffolding features in isolation, varying one feature between two groups and comparing learning outcomes. The results of any individual study along those lines can only provide some evidence that the particular feature may positively or negatively influence learning under similar conditions to the study. Meta-analysis can be a helpful approach to systematically synthesize results across studies (Borenstein et al., 2009; Cooper et al., 2009; Hedges & Vevea, 1998). However, meta-analyses cannot always provide clear design guidance. For example, in our scaffolding meta-analysis, there were no differences in effect size based on whether scaffolding was faded, added, faded and added, or not at all (Belland, Walker, Kim, et al., 2017). But this finding strongly contradicted the majority of scaffolding theory and much empirical research. Therefore, we set out to conduct the current study to see what combinations of scaffolding features tended to lead to the strongest effect sizes within the context of computer science education at the college and graduate school levels. By going beyond the meta-analytic results to use two-step clustering, we were able to see the conditions under which (a) fading and adding is useful (when it is context-specific, conceptual scaffolding designed to enhance knowledge integration, and such fading and adding is based on performance and self-selection), (b) adding is useful, and (c) no scaffolding customization is necessary.

By clustering an expanded version of a comprehensive scaffolding meta-analysis dataset, this study used a systematic approach to determine the optimal combinations of scaffolding features within computer science education at the college and graduate school levels. The results indicate that, within college- and graduate-level technology and engineering education, scaffolding is most effective when it is (a) designed to enhance higher order thinking skills, and (b) is either not customized or is customized on the basis of performance. In this way, we helped to bring clarity to how to designing scaffolding to maximize learning in college and graduate level computer science courses. Contrary to the suggestions of much scaffolding literature (McNeill et al., 2006; Pea, 2004; Puntambekar & Hübscher, 2005), utilizing fading by itself is not associated with medium or large effect sizes in college and graduate-level technology and engioneering education. In contrast, using context-specific scaffolding that is designed to impact higher-order thinking skills with conceptual, context-specific scaffolding leads to a large effect size. In addition, adding scaffolding by itself, and using a combination of fading and adding, were also associated with medium and large effect sizes.

Knowing the combinations of scaffolding features that are associated with medium and large effect sizes can help technology and engineering educators design scaffolding that can help their students learn at an optimal level.

Limitations and Suggestions for Future Research

The dataset used for this clustering study originated with a meta-analysis coding dataset. As is the case with any systematic synthesis effort, meta-analysis collapses across studies to help researchers identify evidence associated with specific instructional features (Cooper et al., 2009). For example, there is a range of fixed customization approaches used in scaffolding, with some preprogramming scaffolding to fade after a certain amount of time has passed, and some after a certain number of iterations of the problem solving process has occurred. By identifying all such scaffolding systems as incorporating fixed customization, one is leaving some detail out of the analysis. But that is simply the nature of meta-analysis (Borenstein et al., 2009; Cooper et al., 2009). That simplification is compounded in clustering, in that a cluster that reflects any particular coding category also reflects the simplification that is inherent in meta-analysis. However, without meta-analyses, one would be subject to the inherent biases reflected in narrative reviews.

This paper essentially employed an approach of dimension reduction (the coding of the included studies in the original meta-analysis) followed by clustering (D'Enza et al., 2014). One potential issue is that the cluster structure may have been distorted by the dimension reduction (D'Enza et al., 2014). However, clustering would not have been possible had the underlying articles not been coded in the first place, and we followed a rigorous process in which all included students were coded by two coders who then came to consensus and inter-rater agreement before coming to consensus was assessed (Belland, Walker, Kim, et al., 2017).

Acknowledgements

This research was partially supported by grants 1906059, 1927595, and 1251782 from the National Science Foundation. Any opinions, findings, or conclusions are those of the authors and do not necessarily represent official positions of NSF.

References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis* (Vol. 2). Sage Publications.

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

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics* (pp. 61–75). Springer New York. https://doi.org/10.1007/978-1-4614-3305-7_4

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

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

Belland, B. R., Walker, A. E., & Kim, N. J. (2017). A Bayesian network meta-analysis to synthesize the influence of contexts of scaffolding use on cognitive outcomes in STEM education. *Review of Educational Research*, *87*(6), 1042–1081. https://doi.org/10.3102/0034654317723009

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. https://doi.org/10.3102/0034654316670999

Belland, B. R., Walker, A. E., Olsen, M. W., & Leary, H. (2015). A pilot meta-analysis of computer-based scaffolding in STEM education. *Educational Technology and Society*, *18*(1), 183–197.

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

Butz, B. P., Duarte, M., & Miller, S. M. (2006). An intelligent tutoring system for circuit analysis. *IEEE Transactions on Education*, *49*(2), 216–223. https://doi.org/10.1109/TE.2006.872407

Chen, Y., Yang, J., Wang, C., & Park, D. (2016). Variational Bayesian extreme learning machine. *Neural Computing & Applications*, *27*(1), 185–196. https://doi.org/10.1007/s00521-014-1710-1

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

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

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 245–252. https://doi.org/10.1145/365024.365111

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Demetriadis, S. N., Papadopoulos, P. M., Stamelos, I. G., & Fischer, F. (2008). The effect of scaffolding students’ context-generating cognitive activity in technology-enhanced case-based learning. *Computers & Education*, *51*, 939–954. https://doi.org/10.1016/j.compedu.2007.09.012

D’Enza, A. I., Velden, M. V. de, & Palumbo, F. (2014). On joint dimension reduction and clustering of categorical data. In *Analysis and Modeling of Complex Data in Behavioral and Social Sciences* (pp. 161–169). Springer, Cham. https://doi.org/10.1007/978-3-319-06692-9_18

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, *52*, 338–348. https://doi.org/10.1016/j.chb.2015.05.041

Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher*, *42*(1), 38–43. https://doi.org/10.3102/0013189X12463051

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). Lawrence Erlbaum Associates.

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

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

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

Kim, N. J., Belland, B. R., & Kim, Y. (2017). Data mining meta-analysis coding to develop smart learning systems that dynamically customize scaffolding. *Proceedings of the 23rd Americas Conference on Information Systems*.

Koedinger, K. R., & Aleven, V. (2007). Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. *Educational Psychology Review*, *19*(3), 239–264. https://doi.org/10.1007/s10648-007-9049-0

Kumar, R., Rosé, C. P., Wang, Y.-C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 383–390.

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

Linn, M., Clark, D., & Slotta, J. (2003). WISE design for knowledge integration. *Science Education*, *87*, 517–538. https://doi.org/10.1002/sce.10086

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

Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, *41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

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. https://doi.org/10.1037/a0037123

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., maintenance(1999-2000)), ctb] (port to R., Studer, M., Roudier, P., Gonzalez, J., Kozlowski, K., pam()), E. S. (fastpam options for, & Murphy (volume.ellipsoid({d >= 3})), K. (2019). *cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.* (Version 2.1.0) [Computer software]. https://CRAN.R-project.org/package=cluster

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

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics*. Springer.

National Academies of Sciences, Engineering, and Medicine. (2017). *Information technology and the U.S. workforce: Where are we and where do we go from here?* National Academies Press. https://doi.org/10.17226/24649

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes* (Version 1.1-2) [Computer software]. https://CRAN.R-project.org/package=RColorBrewer

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

Poitras, E., & Lajoie, S. (2014). Developing an agent-based adaptive system for scaffolding self-regulated inquiry learning in history education. *Educational Technology Research & Development*, *62*(3), 335–366. https://doi.org/10.1007/s11423-014-9338-5

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. https://doi.org/10.1207/s15326985ep4001_1

R: The R Stats Package. (n.d.). Retrieved March 22, 2020, from https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html

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

Rodriguez, D., Sicilia, M. A., Cuadrado-Gallego, J. J., & Pfahl, D. (2006). e-Learning in Project Management Using Simulation Models: A Case Study Based on the Replication of an Experiment. *IEEE Transactions on Education*, *49*(4), 451–463. https://doi.org/10.1109/TE.2006.882367

Roever, C., Raabe, N., Luebke, K., Ligges, U., Szepannek, G., Zentgraf, M., & Meyer, D. (2020). *klaR: Classification and Visualization* (Version 0.6-15) [Computer software]. https://CRAN.R-project.org/package=klaR

Saye, J., & 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. https://doi.org/10.1007/BF02505026

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. https://doi.org/10.1037/a0032447

Sun, S. (2013). A review of deterministic approximate inference techniques for Bayesian machine learning. *Neural Computing & Applications*, *23*(7/8), 2039–2050. https://doi.org/10.1007/s00521-013-1445-4

Theodoridis, S. (2015). *Machine learning*. Academic Press. http://proquest.safaribooksonline.com/9780128015223

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

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, *46*, 1060–1106. https://doi.org/10.3102/0002831209333183

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

Yeh, Y.-F., Chen, M.-C., Hung, P.-H., & Hwang, G.-J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers & Education*, *54*(4), 1089–1100.

Yin, C., Song, Y., Tabata, Y., Ogata, H., & Hwang, G. J. (2013). Developing and implementing a framework of participatory simulation for mobile learning using scaffolding. *Educational Technology and Society*, *16*(2), 137–150.