

Investigating Adaptive Activity Effectiveness Across Domains: Insights into Design Best Practices

Rachel Van Campenhout, Bill Jerome, Jeffrey S. Dittel, and Benny G. Johnson

VitalSource Technologies, Pittsburgh PA, USA
Rachel.Vancampenhout@vitalsource.com

Abstract. Courseware as an adaptive instructional system is a complex environment to develop. The student will encounter lessons of content with integrated formative practice, adaptive activities, and assessments in their learning path. The alignment of all course features, including the scaffolding structure of the adaptive activities, may vary between courses and the teams who created them. In a previous analysis of adaptive activities [1], these activities had net positive effects on student learning estimates and summative assessment scores. In this paper, we will analyze three additional non-STEM courses that had less effective adaptive activities using the same methods as the original study, and further investigate course features that could be influencing their effectiveness, such as alignment, difficulty, and amount of practice. The results of this analysis can provide guidance on how to best create content for adaptive courseware and provide an example of the critical role data analysis has in the evaluation and iterative improvement of student learning environments.

Keywords: Adaptive Activities, Adaptive Courseware, Formative Practice, Learning by Doing, Learning Outcomes.

1 Introduction

The Acrobatiq by VitalSource platform delivers data-driven courseware that provides a comprehensive learning environment for students, from text content and media to formative practice and summative assessments. In particular, the courseware utilizes real-time student data to deliver adaptive practice activities throughout a student's learning path. The adaptive activities are personalized for each course learning objective and provide students scaffolded questions adapted to fit their needs. Previous analysis of a Probability and Statistics course investigated how the adaptive activities impacted student learning by analyzing their impact on two metrics: learning estimates and summative assessment scores [1]. It was found that the adaptive activities helped increase students' learning estimates. It was also found that the adaptive activities helped students increase their summative assessment scores. Students who were able to increase their learning estimates via the adaptive activities also scored higher on summative assessments than their peers who did not increase their learning estimates.

This study aims to build upon this research in two ways. First, we will replicate the analysis done for the previous study's research questions on several other non-STEM

courses (Project Management, Finance, and Macroeconomics) that had less effective adaptive activities. For each of these courses, we will answer the following research questions: “How do the adaptive activities affect learning estimates for students?” and, “How do the adaptive activities affect student scores on summative assessments?”

Second, by analyzing a variety of courses with variation in results, we can investigate a second line of inquiry on the features of the courseware that could influence the degree of effectiveness for these adaptive activities. The research question we will seek to answer is: “What features of courseware design influence the effectiveness of the adaptive activities for increasing student learning estimates and outcomes?” As this second area of investigation concerns course features, the original Probability and Statistics course from previous research will be used to compare results with the additional courses analyzed for the first research questions. To answer this research question, we will investigate course features such as question alignment, difficulty, and the amount of adaptive practice for each course. The question difficulty for each course can be analyzed for the formative questions, adaptive questions, and summative questions separately to examine any potential relationship to the effectiveness of the adaptive activities. For example, if the summative question difficulties are substantially different from the formative or adaptive practice, then it is possible the adaptive practice may not have a measurable effect on summative scores.

By analyzing these course features in relation to the effectiveness of the adaptive activities, we can identify a set of guidelines for authoring and developing course features. While tools such as the adaptive practice can be used to help students learn and improve outcomes, implementation of these tools can vary and impact their effectiveness. By evaluating both the effectiveness of adaptive activities and course features which could impact their effectiveness, we can suggest implementation guidance for adaptive activities across subject domains. Understanding course design through data is needed to continue data-driven iterative improvements that benefit the learner. Courseware should be able to maximize learning benefits for students no matter the domain. It is critical for features—especially adaptivity—to be effective regardless of content area.

2 Methods

The Courseware. The courseware analyzed were all developed on the Acrobatiq platform using the same learning methodology and design. Each course includes units of content, broken down into modules, broken down further into individual lesson pages. Each lesson page is organized around a learning objective, and all content and formative practice aligns—and is tagged to—that learning objective. The formative practice provides students with opportunities to practice what they have learned as they work through the content, receiving immediate targeted feedback. Formative practice is a foundational learning method shown to increase learning for different age groups and different subjects [2] and feedback increases the benefits of formative practice [3]. The formative practice integrated with the content is a type of learning by doing, which has

been shown to have about six times the effect size than reading alone [4]. This learning by doing approach has also been shown to be causal for learning [5–7].

The formative practice questions are not graded, but student responses contribute to their learning estimate—a predictive measure generated for each student on each learning objective. This learning estimate is based on a machine learning model which uses item response theory (IRT) [8, 9] to predict how well the student might perform on a learning objective in a summative assessment. This IRT approach takes into account the psychometric properties of questions when constructing the ability estimate: the difficulty and discrimination of questions is modeled by a two-parameter logistic model. A Bayesian approach [10] estimates the posterior distributions of the IRT question parameters from data in addition to the student ability posterior distribution from the formative and adaptive questions students answered. The learning estimate value between 0 and 1 is derived from the ability posterior, with higher values indicating a higher probability of obtaining a passing score on that learning objective’s summative assessment. The model will assign a category of low, medium, or high as long as there is enough data for the model to have sufficient confidence in its prediction.

The learning estimate drives the personalization in the adaptive activities. Placed after a series of lesson pages in the module, the adaptive activity uses the student’s learning estimate for each learning objective to determine the set of questions the student will receive. Students with high learning estimates will only see the most difficult questions which should mirror those on an assessment. Students with low or medium learning estimates receive those same question in addition to easier questions that will help to scaffold knowledge gaps leading to the hardest questions. The low and medium level questions can provide steps or sub-steps toward the hard questions, reducing cognitive load for those students much like a worked example [12]. The purpose of the scaffolding is to meet students at their ability level and build from there, much like Vygotsky’s zone of proximal development [11]. The adaptive activities are formative in nature. They do not produce a score for the gradebook, but are graded on completion only, regardless of accuracy. Like the formative practice, students receive immediate, targeted feedback and can continue answering the questions until they get them correct. Responses on the adaptive activity contribute to the final learning estimate calculations as well.

The Population. Each of the three courses selected were run at the same traditional four-year institution in 2019. The courseware was used as the primary learning resource in introductory courses led by faculty. No experimental manipulations were used for this population; students engaged in the courseware as a natural part of their credit-bearing course. The platform did not collect demographic information for students.

The Data. The courseware data necessary for this analysis are the formative question attempts, accuracy, learning estimate states, adaptive activity attempts and accuracy, and summative assessment attempts and accuracy. The unit of analysis for this study were the student-learning objective pairs, i.e., a single student’s work on a single learning objective. As in the original study, the data set for each course was reduced to only

include learning objectives with formative, adaptive, and summative question attempts. The final data for each course is listed in Table 1.

Table 1. The data for each course analyzed.

Course	Students	Learning Objectives	Total Data Records
Project Management	72	47	2798
Macroeconomics	84	58	2797
Finance	76	50	3053

3 Results

3.1 RQ 1: How do the adaptive activities affect learning estimates for students?

Overall Learning Estimate Change. The first analysis was to evaluate the overall learning estimate changes from before to after the adaptive activity for all student-learning objective pairs. This same analysis was done on all three courses. For Project Management, there were 1,589 instances in which a learning estimate was available immediately before and after the adaptive practice, and the mean learning estimate change was -0.024, which was overall negative. For Finance, there were 1,623 instances of learning estimate changes, with a mean change of -0.024 as well. For Macroeconomics, there were a total of 1,695 instances of learning estimates with a mean of -0.007. As in the original study, for each course a Shapiro-Wilk test showed the learning estimate differences were not normally distributed. A one-sample Wilcoxon signed rank test showed the median learning estimate change was statistically different from 0 ($p \lll 0.001$) for all courses.

Table 2. The learning estimate increases and decreases for each course.

Statistic	Project Management		Finance		Macroeconomics	
	Learning Estimate Increase	Learning Estimate Decrease	Learning Estimate Increase	Learning Estimate Decrease	Learning Estimate Increase	Learning Estimate Decrease
count	619	970	709	914	778	917
mean	0.045433	-0.068897	0.046014	-0.077883	0.026653	-0.036113
std	0.045714	0.052469	0.060242	0.075473	0.037764	0.036812

We can compare these learning estimate changes to those found in the Probability and Statistics course [1]. In that course, 62% of all learning estimate changes were positive with a mean increase of 0.132, while the learning estimate decreases had a mean of -0.064. The learning estimate increases were twice as large compared to the decreases. For these three business courses, there are more decreases in learning estimates than increases, and the mean decrease is 1 to 3 percentage points larger than the mean increase. This tells us that further investigation is needed to understand why the results are contradictory to the original analysis. Misalignment of course features such as question difficulty or the design of the adaptive activities themselves could be affecting these overall learning estimate changes.

Learning Estimate Change by Category. The second analysis was to evaluate the learning estimate changes according to each category of learning estimate before and after the adaptive activity. A student's learning estimate categories immediately prior to starting the adaptive activity determines the level of scaffolding they receive for each learning objective. The questions in the adaptive activity are formative as well, so contribute to the learning estimate and therefore students can change learning estimate categories after completing the adaptive activity.

Project Management. In Project Management, the majority of all student estimates remained in their category. The largest change in category was the medium learning estimate category—after completing the adaptive activity, approximately 20% of medium learning estimates shifted to high and 20% shifted to low. The high and the low category both only had a small portion of estimates change to the next available category. This does not tell an overwhelmingly clear story. It seems that learning estimate category changes after the adaptive activities nearly balance out.

Table 3. The number of instances of learning estimate changes after completing adaptive questions, grouped by learning estimate category.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	471	77	2
Medium (before adaptive)	112	284	104
Low (before adaptive)	1	41	453

Finance. The learning estimate category groups for Finance show a similar pattern when compared to Project Management. The majority of all student learning estimates remained in the same category. The medium category had the most change, with approximately 40% changing category at a nearly equal split between high and low.

Nearly 17% of high learning estimate instances changed from high to a lower category, while only 7% of low learning estimate instances changed from low to a higher category.

Table 4. The number of instances of learning estimate changes after completing adaptive questions, grouped by learning estimate category.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	541	99	10
Medium (before adaptive)	95	296	101
Low (before adaptive)	1	26	341

Macroeconomics. There was less movement between categories in this course, compared to the others. For both the high and low categories, no students had learning estimate instances move 2 categories (high to low or low to high). Very few students even moved one category. The medium category had the largest move, but more to low than to high.

Table 5. The number of instances of learning estimate changes after completing adaptive questions, grouped by learning estimate category.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	861	33	0
Medium (before adaptive)	50	214	69
Low (before adaptive)	0	30	420

3.2 RQ 2: How do the adaptive activities affect mean summative assessment scores for students?

The learning estimate is a predictive measure generated for each student on each learning objective, but the more traditional method of evaluating student learning is through a summative assessment. Students completed a quiz shortly after the adaptive activity. We compared the mean summative assessment scores for each learning estimate category.

Project management. The first—and most surprising—result was the overall range of mean summative scores. All categories averaged over 0.9 (90%) from low to high, with a range of 0.044 (4.4%). Students performed very well on the summative assessments, though the reason for this is yet unknown. When looking at the upward movement within a category, we find that students who increased their learning estimate after the adaptive practice (students who changed from medium before to high after; low before to medium/high after) scored higher than their peers who did not increase their learning estimate. There are a few anomalies in the trend, but those are mostly where there are small amounts of data (i.e. the high before adaptive to low after adaptive category has a higher mean summative score than any other high category, but that score is also created from only 2 instances).

Table 6. Mean summative scores by learning estimate category before and after the adaptive activity questions.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	0.932	0.921	0.958
Medium (before adaptive)	0.937	0.902	0.908
Low (before adaptive)	1.00	0.922	0.914

Finance. The mean summative assessment scores for each learning estimate category before/after the adaptive activity show the same trends as the Project Management. Students who had a high category after the adaptive activities had higher summative assessment scores than those who had medium or low learning estimates after the adaptive activities. Students with medium learning estimates after the adaptive activities had higher scores than those with low learning estimates. These trends are also consistent within the high, medium, and low category before the adaptive activity: within each of those categories, mean summative scores increased as the learning estimate categories increased after the adaptive activities. The range of mean scores is more consistent with expectations, ranging from 0.733 (73.3%) to 0.863 (86.3%) (the mean score of 1.00 was generated from one data point so was not used as the mean high score).

Table 7. Mean summative scores by learning estimate category before and after the adaptive activity questions.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	0.863	0.817	0.733
Medium (before adaptive)	0.844	0.814	0.706
Low (before adaptive)	1.00	0.767	0.748

Macroeconomics. The mean summative assessment scores for Macroeconomics do not follow the trends of Project Management and Finance. Instead of higher learning estimate categories corresponding to higher summative assessment scores, these scores appear to be a more random pattern. The spread of these scores is also very close: all categories fall between 0.844 and 0.882, less than 4% difference.

Table 8. Mean summative scores by learning estimate category before and after the adaptive activity questions.

Learning Estimate Category	High (after adaptive)	Medium (after adaptive)	Low (after adaptive)
High (before adaptive)	0.882	0.863	NA
Medium (before adaptive)	0.857	0.872	0.874
Low (before adaptive)	NA	0.882	0.844

Overall, the analysis of all three courses for the first two research questions shows general inconsistencies with the original study. All three courses showed the adaptive activities had a net negative impact on the learning estimates. Two of the three courses (Project Management and Finance) confirmed the trend that students who increased their learning estimate after the adaptive practice increased their summative assessments compared to their peers who did not. Two of the three courses (Project Management and Macroeconomics) showed unusual results in the summative assessment scores as well, with all ranges of students performing high and with a very narrow spread of scores. Course features for each course will be investigated to help contextualize these findings.

3.3 RQ 3: What features of courseware influence the effectiveness of the adaptive activities?

Alignment. To gain a bigger-picture understanding of what could be influencing some of these unexpected findings, we will look at the difficulty alignment of question types: formative, adaptive, and summative. Question alignment ensures that the students receive learning benefits from formative practice questions and adaptive questions, while also being fairly evaluated in the summative assessment. A misalignment between question difficulty in different question types could be affecting the benefit of the adaptive activities. For instance, if the adaptive activities were more difficult than the formative practice questions, then on average students would not do as well on them and learning estimates would decrease. If summative assessment questions were much more difficult, then the formative and adaptive questions would not have sufficiently prepared students and they would not do as well on the summative questions.



Fig. 1. Line charts of all four courses showing the mean scores on formative, adaptive, and summative questions by learning estimate category.

For all four courses, the relationship between the formative practice and adaptive practice mean accuracy was close despite differences in direction. In Probability and Statistics, students in all categories did slightly better in the adaptive activities than the formative practice. In the other three business courses, students had slightly lower mean accuracy scores on adaptive activity questions than from the formative questions. This could suggest that the formatives were easier than the adaptive activities (or the adaptive activities were more difficult than the formatives).

The mean accuracy/scores on formatives, adaptive activities, and summative assessments are unaligned for some courses. For instance, Project Management had formative and adaptive mean scores which were relatively parallel for students at all learning estimate levels, but then the summative mean scores for all learning estimate categories were extremely high. Students at the low level had a mean accuracy in the 40% range for formatives and adaptive activities, yet scored in the 90% range for the summative assessments. It is unlikely that this increase in accuracy was caused by the formative practice. It is also unlikely that the summative assessments (which were written by the same subject matter experts who wrote the formative practice) were so easy that all students scored within a few percentage points of each other. This leads us to hypothesize that some external variable is the cause for this misalignment, such as the assessments being treated as practice tests taken collaboratively in class. Different implementation practices of the courseware between courses could easily affect the data seen here, but would not be captured via the platform itself so only inferences can be made.

The alignment of practice across the course is a good indicator of how effective the course design is, as well as when external factors may be influencing outcomes. Probability and Statistics is visibly the most challenging course with about a 20-point mean accuracy/score spread between low and high learning estimate groups, with changes under 8 points between formative questions and summative questions across all groups. However, the business courses show a dramatic misalignment between the formative and adaptive questions and the summative questions, for reasons not yet known.

Amount of Adaptive Practice. We also compared the average amount of adaptive practice available for each learning objective across all courses. The Probability and Statistics course had an average of 14.05 adaptive practice questions per learning objective. Project Management averaged 6.69, Finance averaged 5.17, and Macroeconomics averaged 4.85. Probability and Statistics had approximately double the available adaptive practice and also had a positive net change in learning estimate scores with a mean increase three times higher than any other course. This suggests that the amount of available practice is related to the possible increase of learning estimates. While the amount of practice may not be sufficient to guarantee increased learning estimates, it is a necessary component.

The next feature we investigated was the amount of scaffolded practice in the adaptive activities, i.e. how many questions were tagged for delivery at the low, medium, and high level. The goal of the scaffolding questions is to provide a series of questions

on a learning objective that start easy and increase in difficulty to help students who were struggling on that learning objective make more incremental steps in difficulty. There were guidelines for the development of the activities, but the number and level of scaffolded questions was up to the discretion of the subject matter experts writing the activities. It was reasonable to expect some learning objectives would need more or fewer scaffolded questions depending on how complex the learning objective was. However, the breakdown of adaptive questions by level showed surprising results compared to our expectations. Each course had learning objectives with no scaffolding questions at all (all questions for the learning objective were set to high) effectively meaning they were non-adaptive for that learning objective. The breakdown of adaptive questions per course by quartile revealed that several courses had far fewer low and medium level questions than others. To illustrate the differences between courses, we will look at the amount of adaptive practice available for learning objectives at the 75th percentile.

Table 9. Adaptive practice available at the 75th percentile for each scaffold level.

Course	Low Questions	Medium Questions	High Questions
Project Management	1	1	7
Macroeconomics	1	1	4
Finance	1	1	3.75
Probability and Statistics	5	5	9

Given that all courses had learning objectives without scaffolded questions, none of them meet the expectations for the design of the activities. However, there are meaningful differences in the amount of scaffolded practices for learning objectives in each course. At the 75th percentile, the three business courses each only have a single question per objective at the low and medium level. The ratio of scaffolding questions to high questions favors the high questions in all three courses. At the 75th percentile, the Probability and Statistics course had the highest number of low and medium-level questions (5 and 5, respectively) which puts the number of scaffolding questions higher than the number of high-level questions. Struggling students received more scaffolded support in Probability and Statistics and had more positive learning estimate changes and more instances of increased summative assessment scores.

Learning Estimate Changes by Learning Objective. Given that the mean adaptive scores for all three business courses trended down and there were learning objectives without scaffolding, we approached the data from a different perspective. What were the adaptive characteristics of learning objectives that increased and decreased learning estimates the most? To illustrate this, we will start with examples from Finance, which had a mean learning estimate change of -0.024. The learning objective that had the

largest decrease in learning estimates had a mean learning estimate change of -0.245, about ten times larger than the mean decrease. When we look at the adaptive questions for this objective, we find that there was 1 low question, 1 medium question, and 27 high questions. The ratio of scaffolded questions is extreme. The learning objective that had the next largest decrease had a mean change of -0.114 (four times the mean) had 2 low, 0 medium, and 6 high. This ratio also favored difficult questions. When we look at the learning objectives that increased learning estimates, we find a different trend. The two learning objectives that increased learning estimates the most each had a total of 9 adaptive questions: 2 low, 4 medium, 3 high and 2 low, 3 medium, 4 high respectively. More than half of the questions were at the low and medium level for these learning objectives. Seeing similar trends in all courses, this indicates that the ratio of scaffolded questions matters for how helpful the adaptive set is for students on a given learning objective. This analysis also shows how specific learning objectives that are outliers can shift the aggregated means for all learning objectives.

Alignment of Adaptive Scaffolding. Considering the alignment of formative question accuracy to adaptive question accuracy from previous analysis combined with the lack of scaffolded questions on some learning objectives in all courses, we analyzed the difficulty level of the adaptive questions according to the difficulty level at which they were tagged. We compared how different learning estimate categories performed on different levels of scaffolded questions. For Probability and Statistics as well as Finance, the questions tagged to low, medium, and high did perform according to the expected trend (high-level questions were more difficult to answer and low-level questions were easier to answer, across all groups). However, we found that for Macroeconomics and Project Management that each level of question was almost equally as difficult for students (meaning a student in the low learning estimate category did not find the easy questions any easier than the hard questions). Scaffolded questions not being written at the intended level of difficulty likely contributed to the overall lack of effectiveness of those adaptive activities.

4 Conclusion

There are many factors that can influence the success of adaptive instructional systems. In the comparison of courseware in this paper, we saw varying results for the effectiveness of the adaptive activities. The features that seemed to have the largest impact on the effectiveness of the adaptive activities were the design of the adaptive activities themselves—the number of adaptive questions as well as item analysis of the questions at each level. Adaptive activities with more questions per learning objective and high ratios of scaffolded questions to hard questions were most successful in increasing learning estimates. This isn't to say that other course features such as alignment of formative, adaptive, and summative questions are not important. It is likely that alignment will always impact the effectiveness of adaptivity and assessment more generally, but the effects of alignment were obscured by the findings of the adaptive activity design itself. There are also complicating external factors—such as how the courseware

features are used in the classroom environment or individual student engagement behaviors—which cannot be accounted for in this paper.

Courseware is a complex learning environment to design. The initial design of adaptivity within courseware was research-based yet imperfect, and iterative improvement is vital to the development process of courseware in general, and adaptivity specifically [13]. The analysis in this study identified that the effect of adaptive activities will vary depending on their design and content. By identifying the variables which could influence the success of adaptivity for students and their learning, best practices can be identified for future development and current courseware can be improved for the betterment of the students. The analysis of these courses suggests that the adaptive activities:

- Should have more than eight questions per learning objective, generally
- Should have a minimum ratio of 50% scaffolding questions to high questions for every learning objective
- Should ensure low and medium questions are written to their respective level

This paper also shows how critical the use of data is for the improvement of student learning. Consider the earlier design note that the number and level of scaffolding questions was up to the discretion of the subject matter expert. It's likely that the subject matter expert evaluated some learning objectives and considered them simpler, therefore not in need of as much scaffolding. Yet the data showed that some students struggled on all learning objectives and that less scaffolding was less helpful. This can be a symptom of “expert blindness;” it can be challenging for experts to imagine the first-time learning experience of students, especially those who are struggling. The data remind us to never make assumptions on whether or not students may need additional support. Data can also help to adjust the difficulty level of the adaptive questions over time, as subject matter experts may intend on writing to a specific difficulty level and yet it may be more or less difficult for students in practice when completed in the end environment. This analysis has provided valuable insights and guidance for how to scaffold practice for students to optimize the benefits of adaptive practice. These insights can benefit many courses and many more students in future iterations of design.

Acknowledgements. The authors sincerely thank Leandro Ucha for his assistance in preparing the data for the analysis of the adaptive activities.

References

1. Van Campenhout, R., Jerome, B., & Johnson, B. G. The impact of adaptive activities in Acrobatiq courseware: Investigating the efficacy of formative adaptive activities on learning estimates and summative assessment scores. In: Sottolare R., Schwarz J. (eds) Adaptive Instructional Systems. HCII 2020. LNCS, vol 12214. Springer. pp 543–554. (2020). https://doi.org/10.1007/978-3-030-50788-6_40
2. Black, P., William, D.: Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* 92(1), 81–90 (2010). <https://doi.org/10.1177/003172171009200119>

3. Dunlosky, J., Rawson, k., Marsh, E., Nathan, M., Willingham, D.: Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest* 14(1) 4–58 (2013). <http://dx.doi.org/10.1177/1529100612453266>
4. Koedinger, K., Kim, J., Jia, J., McLaughlin, E., Bier, N.: Learning is not a spectator sport: doing is better than watching for learning from a MOOC. *Learning at Scale*, pp. 111–120. Vancouver, Canada (2015). <http://dx.doi.org/10.1145/2724660.2724681>
5. Koedinger, K., McLaughlin, E., Jia, J., Bier, N.: Is the Doer Effect a causal relationship? How can we tell and why it's important. *Learning Analytics and Knowledge*. Edinburgh, United Kingdom (2016). <http://dx.doi.org/10.1145/2883851.2883957>
6. Koedinger, K. R., Scheines, R., & Schaldenbrand, P. Is the doer effect robust across multiple data sets? *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018*, 369–375. (2018).
7. Olsen, J., Johnson, B. G.: Deeper collaborations: A finding that may have gone unnoticed. Paper presented at the IMS Global Learning Impact Leadership Institute, San Diego, CA (2019).
8. Baker, F. The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation. College Park, MD (2001). Retrieved from <http://echo.edres.org:8080/irt/baker/>
9. Embretson, S., Reise, S.: *Item response theory for psychologists*. Mahwah, NJ: Erlbaum (2000).
10. Fox, J.: *Bayesian Item Response Modeling: Theory and Applications*. Springer (2010).
11. Sanders, D., Welk, D.: Strategies to scaffold student learning: applying Vygotsky's zone of proximal development. *Nurse Educator* 30(5), pp. 203-204 (2005).
12. Sweller, J.: The worked example effect and human cognition. *Learning and Instruction*, 16(2), pp. 165–169 (2006). <https://doi.org/10.1016/j.learninstruc.2006.02.005>
13. Goodell, J., Thai, K. P., A Learning Engineering Model for Learner-Centered Adaptive Systems. In: Stephanidis et al. (eds) *Adaptive Instructional Systems. HCII 2020. LNCS*, vol 12425. Springer. pp 557–573. (2020). <https://doi.org/10.1007/978-3-030-60128-7>