Usage and performance declines in a classroom-integrated digital learning software over the course of an academic year

Markus Wolfgang Hermann Spitzer
Martin-Luther University, Halle-Wittenberg, Germany
Halle, Germany
sfs.spitzer@gmail.com

Lisa Bardach
Universiy of Giessen, Germany
Giessen, Germany
Lisa.Bardach@psychol.uni-giessen.de

Younes Strittmatter
Princeton University, Princeton, NJ, US
Princeton, USA
ys5852@princeton.edu

Korbinian Moeller
Loughborough University, Loughborough, UK
Loughborough, United Kingdom
K.Moeller@lboro.ac.uk

Abstract

In increasing numbers of classrooms worldwide, students use digital learning software. However, we know little about the trajectories of usage and the performance within such digital learning software over the academic year. This study analyzed real-world longitudinal data from a mathematics learning software used in classrooms in Germany and the Netherlands (1̃6,000 students who worked on >23 million problems). We evaluated students' usage and performance trajectories across an academic year by examining the percentage of students using the software, worked-through problems, active days and weeks, as well as performance. Our results indicate a decline in both usage and performance over the course of the academic year, with overall lower usage in Germany than in the Netherlands. Our findings highlight the need for further research into the factors maintaining or increasing the usage of and performance in classroom-integrated digital learning software over extended periods.

CCS Concepts

• **Applied computing** → **E-learning**; **Interactive learning environments**.

Keywords

digital learning software, mathematics, academic performance, naturalistic data, expectancy-value theory, expectancy-value-cost theory

ACM Reference Format:
Markus Wolfgang Hermann Spitzer, Lisa Bardach, Younes Strittmatter, and Korbinian Moeller. 2025. Usage and performance declines in a classroom-integrated digital learning software over the course of an academic year. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3706468.3706554

LAK 2025, Dublin, Ireland

ACM ISBN 979-8-4007-0701-8/25/03
https://doi.org/10.1145/3706468.3706554

1 Introduction

Digital learning software for secondary education has been developed for decades [2, 3, 9, 18] and has seen widespread adoption, with millions of students using it globally across different platforms (e.g., [19, 22, 24, 27, 32]). While a substantial body of research has focused on how such software should be constructed to most effectively support students' learning (e.g., [1, 15, 21, 25, 31]), surprisingly little is known about its long-term usage once integrated into classrooms (but see [4, 14]). In particular, it remains to be seen how consistently (i.e., how often and how intense) students use digital learning software in naturalistic settings after gaining access to it. Additionally, there is only limited understanding of the trajectories of students' performance when using digital learning software over longer periods. Nevertheless, a better understanding of usage and performance trajectories within digital learning software is vital for optimizing how it is used after it is adopted in classrooms. Hence, in this study, we sought to investigate trajectories of students' usage and performance in a classroom-integrated digital learning software over the course of an academic year.

1.1 Background

Digital learning software allows researchers to exploit exceptionally rich data, as each interaction with the software is typically logged. This logging makes data-driven learning analytics approaches possible [8]. However, there is also a growing interest in theory-informed learning analytics [17, 33], often drawing on findings from well-controlled laboratory settings, including self-reported data. Research questions and hypotheses may be derived from these theories and can be tested using rich naturalistic data from digital learning software. For instance, a critical factor that has been shown to affect students' learning is their motivation [11, 12].

In fact, there is accumulating evidence suggesting motivational declines after starting a study program [7, 10, 20, 30]. For instance, Sutter and colleagues [30] reported that the motivation of college students attending a statistics course declined over the course of the semester. Students' declining motivation was also associated with reduced performance. These results align with other reports by Benden and Lauermann [7] who reported motivational declines over the course of a semester among first-semester college students attending a math course, particularly after the second week of a semester. They also found that the degree of motivational decline during the second and third week of the semester significantly predicted students' performance in the course's final exam. Looking at which processes drive these changes in motivation, another recent study by Lee et al. [20] evaluated motivational changes across four different college STEM courses (i.e., General Chemistry, Organic Chemistry, Foundations of Programming, and Introduction to Computer Science) and observed that components of students' motivation, such as the course-specific intrinsic value and utility value, declined over time.

Together, these studies suggest that student motivation typically declines after starting a study program and that declining motivation significantly predicts poorer student performance. This growing body of research highlights the importance of longitudinal studies to better understand the dynamic trajectories of students' learning.

These findings have been explained by Expectancy-Value Theory [11, 12, 35] or Expectancy-Value-Cost Theory [5, 13]. Both theories illustrate how motivation influences students' academic choices, persistence, and performance. Expectancy-Value Theory posits that students' motivation is driven by their expectancy beliefs (i.e., confidence in task success) and value beliefs (i.e., reasons for task engagement). Expectancy-Value-Cost Theory further considers costs, such as mental effort, emotional stress, and opportunity costs, to account for negative factors that deter students from completing tasks. Both frameworks address students' motivation by examining how beliefs about competence, value, and costs influence their engagement and performance. As such, both theories assume that motivation is not static but may change due to situational context, like the learning environment, tasks, and social interactions.

1.2 Research gap

Notably, all the above-described findings about motivational declines stem from in-person college courses assessing weekly self-reports of student motivation and final exams as performance measures. Thus, it is unclear whether such results generalize to digital learning scenarios. Moreover, a considerable advantage of considering log data from digital learning software used in naturalistic settings is that students' performance is continuously tracked with each activity, giving a much more fine-grained impression of students' behavior than weekly self-reports. On the other hand, a direct measurement of motivation via self-reported is typically not included. However, to asses motivation, other behavioral indicators can be considered, such as the frequency of using the software, with lower activity (i.e., fewer worked-through problems and fewer active days or weeks per month) indicating lower motivation, whereas higher activity (i.e., more worked-through problems and more active days or weeks per month) indicating higher motivation. In fact, Expectancy-Value Theory and Expectancy-Value-Cost Theory both suggest that motivation should be reflected in students' persistence in engaging with learning opportunities (e.g., [5, 12]).

1.3 Study overview

In this study, we investigated students' usage and performance trajectories within a digital learning software used in the Netherlands and Germany. In particular, we evaluated an extensive dataset from the digital learning software *bettermarks* for learning mathematics, considering students who used this digital learning software between 2016 and 2023. As indicators for students' usage trajectories, we evaluated the number of active days and active weeks per month. As performance indicators, we evaluated students' error rates, problem difficulty, and relative error rates (i.e., error rates taking into account problem difficulty) when working through problems within the digital learning software during each month. Importantly, the studies described above identified the critical issue of motivational declines over time using data from in-person self-reports and testing, serving as a first step toward addressing potential remediation of the issue. Similarly, this study aimed to evaluate whether such motivational declines (reflected by reduced usage) and in turn performance generalise to trajectories of secondary students' long-term usage and performance when using a digital learning software. Accordingly, we pursued the following research questions (RQ):

RQ1: Does students' *usage* of the digital learning software decline over the course of the academic year?

RQ2: Does students' *performance* within the digital learning software decline over the course of the academic year?

2 Methods

2.1 The bettermarks dataset

The dataset of the digital learning software *bettermarks* analyzed in this study included students from grades 4 to 10 who used the digital learning software for learning mathematics either in the Netherlands or in Germany (e.g., [27, 29, 34]). Students who use the software in the Netherlands are typically in so-called digital learning classes, with the software substituting traditional paper workbooks, and thus use the software more systematically than in Germany, where students typically use the software as a supplement and thus less systematically [4, 27, 28]. *bettermarks* includes mathematical topics aligned with the curriculum for both countries and covers over 100 different topics and over 100,000 different problems, covering the domains functions, numbers, mensuration, and geometry. The digital learning software is integrated into classrooms and acts as a digital workbook for students. That is, teachers assign problem sets to students that they work through. The digital learning software provides feedback and hints to students and also provides teachers with feedback on students' performance on assigned problem sets (i.e., students' accuracy on typically nine problems covering the same topic). In addition, the digital learning software indicates so-called knowledge gaps when detecting misconceptions among students and automatically suggests additional problem sets to students in order to close it. The dataset of the digital learning software includes information on all problem sets a student worked on. In addition to information about the problem set, the software logs the error rate, date and time when students worked-through a given problem set. Students may repeat problem sets as often as they wish. However, the parameterization of problem sets change with each attempt, discouraging rote memorization and promoting active problem-solving. Figure 1 illustrates the user interface of the digital learning software currently used in Germany. The anonymous data set was shared by *bettermarks* upon request. The fully anonymous data set comprised no sensitive personal information (e.g., gender, age, or sociodemographic status). *bettermarks* was not involved in this study's design and data analysis.

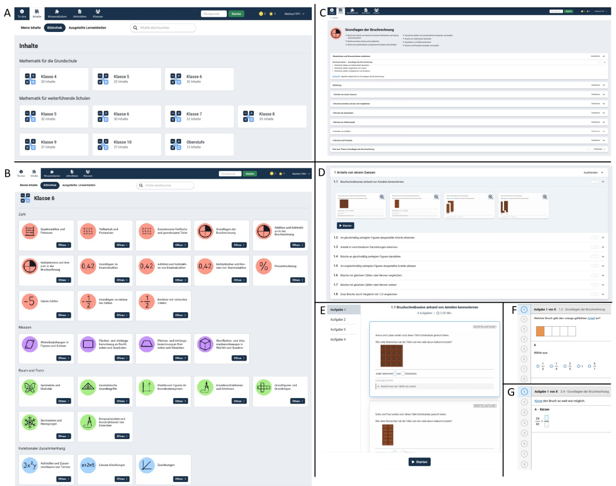

Figure 1: The user interface of the *bettermarks* digital learning software as currently used in Germany. A: Topics structured by grade level. B: Examples of included topics stemming from different domains (indicated as different background colors) suited for sixth-grade students (Klasse 6). C: When selecting a topic (e.g., Basics of fractions; "Grundlagen der Bruchrechnung" in German), users may select between different subtopics (e.g., Shares of a whole; "Anteile von einem Ganzen" in German). D: An exemplary list of problem sets within a subtopic. E-G: Exemplary problems of problem sets included in this subtopic.

2.2 Inclusion criteria for the present study

We considered students and problem sets these students worked through based on the following criteria. First, we included students who registered with *bettermarks* between July 1st, 2016, and August 31st, 2022. Second, we included all students who used the digital learning software in the Netherlands or Germany. Third, to calculate students' error rates, we considered students' first attempt on a completed problem set. However, we also considered students' best attempt and recalculated error rates to assure that our results were similar for the first and the best attempts. Fourth, we only looked at students' usage and performance within their first year of using *bettermarks*. This allowed us to compare the same duration of using the system for different cohorts (i.e., students who registered in 2016, 2017, 2018, 2019, 2020, 2021, and 2022). Fifth, we only considered students who used the digital learning software in each month of the first academic year (i.e., September to May, and not during the summer holidays when considerably fewer students generally use the digital learning software). Finally, we only considered students who registered between July 1st and August 31st during the summer holidays and then looked at their usage and performance throughout the first academic school year they used *bettermarks*. After we applied these inclusion criteria, our dataset comprised 1,888 students from Germany (who worked through 1,201,718 problems) and 14,153 students from the Netherlands (who worked through 22,396,283 problems).

2.3 Data analysis

Our data analysis was conducted using R [23]. To address RQ1 on evaluating students' usage trajectories, we considered the number of worked-through problems, active days, and active weeks per month for each country as dependent variables and indicators of students usage. An active day or an active week is when students worked through one problem set on a day or week. We ran three hierarchical linear regression models (one for each dependent variable) with the two independent variables month (interval scaled) and country (as a categorical variable) and a random intercept for students. We evaluated differences between the two countries as students' usage in the Netherlands should be higher [4, 28].

To address RQ2, we evaluated students' performance, considering their error rates, problem set difficulty, and relative error rates for each month and each country. Error rates were obtained from *bettermarks* and reflect performance independent of problem difficulty. We calculated difficulty as the average error rate of each problem set for all students who worked through this problem set. As a final indicator for students' performance, we examined students' relative error rates. Relative error rates accounted for problem set difficulty and can thus be interpreted as error rates relative to the difficulty of problems. Relative error rates were calculated by subtracting difficulty from individual students' error rates (e.g., an error rate of 10% on a problem set with an average difficulty of 30% resulted in a relative error rate of -20%). Note that lower and more negative relative error rates indicated better performance. We then ran a hierarchical linear regression model for each of these three dependent variables (i.e., error rates, difficulty, and relative error rates), with the two independent variables month and country and a random intercept for students. Importantly, we considered students' first attempts when we evaluated error rates, problem set difficulty, and relative error rates. However, results on students' best attempts were also evaluated (not reported here) and substantiated the findings on first attempts.

3 Results

Students' usage and performance trajectories are illustrated in Figure 2. Statistical details of the hierarchical linear regression models on indicators of students' usage are depicted in Table 1 whereas results for students' performance measures are given in Table 2.

3.1 RQ1: Usage trajectories

The results of the hierarchical linear regression model for worked-through problems suggested a significant main effect for month and for country. The decline of worked-through problems was significantly more pronounced in the Netherlands than in Germany, reflected by the significant interaction between the two independent variables. We observed the same pattern of results for the number of active days per month as well as the number of active weeks per month: students from the Netherlands showed significantly more active days and weeks per month than German students. However,

Table 1: Students' usage with respect to problems, days, and weeks.

	Problems				Days				Weeks			
Coeffcient	estimate	SE	*t*	*p*	estimate	SE	*t*	*p*	estimate	SE	*t*	*p*
Intercept	172.79	0.99	173.85	**<0.001**	6.82	0.03	203.44	**<0.001**	3.29	0.01	310.18	**<0.001**
Month	-8.25	0.12	-67.67	**<0.001**	-0.25	0.01	-59.57	**<0.001**	-0.07	0.01	-47.25	**<0.001**
Country	76.10	0.99	76.57	**<0.001**	2.67	0.03	79.57	**<0.001**	0.61	0.01	57.74	**<0.001**
Month:Country	-3.92	0.12	-32.18	**<0.001**	-0.13	0.01	-30.71	**<0.001**	-0.02	0.01	-16.31	**<0.001**
Marginal R^2 / Conditional R^2	0.197 / 0.425				0.198 / 0.414				0.125 / 0.196			

Table 2: Students' performance with respect to error rates, difficulty, and relative error rates.

	Error rates				Difficulty				Relative error rates			
Coeffcient	estimate	SE	*t*	*p*	estimate	SE	*t*	*p*	estimate	SE	*t*	*p*
Intercept	0.28	0.01	142.31	**<0.001**	0.29	0.01	342.08	**<0.001**	-0.09	0.01	-47.26	**<0.001**
Month	0.01	0.01	58.05	**<0.001**	0.01	0.01	80.25	**<0.001**	0.01	0.01	50.10	**<0.001**
Country	0.03	0.01	16.17	**<0.001**	0.03	0.01	34.25	**<0.001**	0.03	0.01	18.98	**<0.001**
Month:Country	-0.01	0.01	-15.23	**<0.001**	-0.01	0.01	-13.02	**<0.001**	-0.01	0.01	-15.21	**<0.001**
Marginal R^2 / Conditional R^2	0.024 / 0.450				0.089 / 0.287				0.019 / 0.463			

irrespective of country, active days and weeks declined significantly over the course of the academic year. This decline was again more pronounced in the Netherlands than in Germany as indicated by the significant interaction of the variables months and country. In sum, the results of these analyses indicated that students' usage declined continuously each month over the course of the academic year. In addition, usage was generally higher in the Netherlands than in Germany, and the decline in usage was more pronounced in the Netherlands than in Germany. This may be explained by higher initial usage in the Netherlands, leading to a steeper decline.

3.2 RQ2: Performance trajectories

We observed a significant decline in performance reflected by significantly increasing error rates over the course of the academic year. Error rates increased for both countries. However, the increase was more pronounced for Germany than for the Netherlands, as indicated by the significant interaction of the dependent variables month and country. Figure 2D depicts that error rates were significantly higher in the Netherlands than in Germany at the beginning of the academic year, but this difference decreased towards the end of the academic year.

Results for difficulty as the dependent variable indicated that students worked through significantly more difficult problem sets towards the end of the academic year and that students from the Netherlands worked through significantly more difficult problem sets than German students. A significant interaction between month and country further suggested that this difficulty difference between the two countries decreased over the course of the academic year, with a steeper difficulty increase observed in Germany than in the Netherlands.

For relative error rates as the dependent variable, results indicated that these increased over the course of the academic year for both countries, as indicated by a significant main effect for month, replicating observed performance declines stated earlier. In addition, relative error rates were higher in Germany than in the Netherlands suggesting significantly better performance in the Netherlands than in Germany. The interaction between month and country was significant and suggested that this relative error rate difference between the two countries decreased over the course of the academic year, with a steeper increase in relative error rate observed in Germany than in the Netherlands.

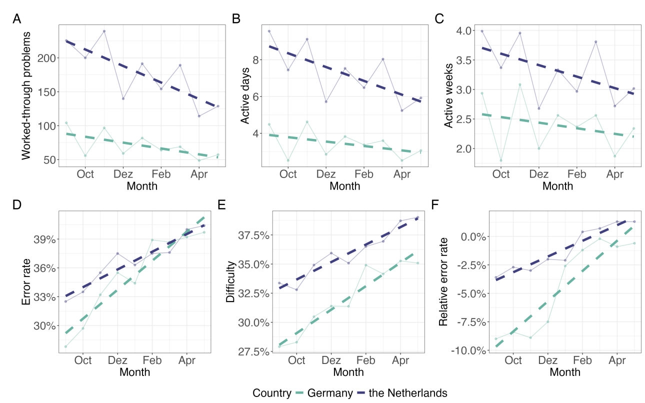

Figure 2: Students' usage and performance of the digital learning software as a function of months and country separated for number of worked-through problems (A), number of active days (B), number of active weeks (C), error rates (D), difficulty (E), relative error rate (F) as dependent variables. Dots illustrate monthly averages for each country and dashed lines indicate the regression line for each country, respectively.

4 Discussion

In this article, we evaluated usage and performance trajectories of students regularly using a digital learning software in the classroom over the course of an academic year. Our findings indicated that students' usage and performance declined with time, reflected by declines in worked-through problems, active days, and weeks per month, as well as declining students' performance. Additionally, indicators for usage and performance were consistently higher in the Netherlands, where students' typically use the software more systematically [4, 28].

To the best of our knowledge, this study is the first to evaluate long-term trajectories of students' usage and performance within a digital learning software for secondary education employed in a naturalistic setting (i.e., in classrooms). Such research is important as enormous efforts have been made to develop and test digital learning software in (quasi) experimental settings (e.g., [1, 15, 21, 25, 31]). However, there is a scarcity of research examining the actual usage and performance trajectories when digital learning software has been implemented in classrooms—without influences of any kind of experimental manipulation that might bias students' motivation to keep using the software or performing well due to the evaluating character of an intervention study, for instance. Nevertheless, our findings are in line with previous research on declines in motivation over the course of a semester in tertiary education [7, 10, 20, 30].

At first sight, these results may be unexpected as they seem to suggest that students' performance decreases the longer students use the learning software over the course of an academic year. However, it is imperative to emphasize that we did not evaluate *learning gains* or how much knowledge students accumulated over time as done in previous studies on students' learning gains or performance gains as a function of students' usage of a digital learning software [14, 19, 27]. Instead, we evaluated their overall trajectories of usage and performance (i.e., error rates and relative error rates on different problem sets) separately at specific time points. As the content of the digital learning software is separated into different topics (e.g., basics of fractions), and students work through different topics within the academic year, we did not evaluate continuous learning trajectory but rather evaluated students' performance on different topics which can be more or less independent from each other. Moreover, the curriculum of problems within each topic is usually designed in a way that difficulty is steadily increasing within a topic (i.e., relatively easy problems with illustrations come first, followed by mere calculation problems of moderate difficulty, and finally text problems on a given topic, that are typically difficult; [26]). Therefore, the observed performance decreases do not indicate that students did not learn over the course of the academic year. Yet, they indicate that with the academic year progressing, students perform poorer within the digital learning software—even when accounting for the difficulty of the worked-through problem sets. This pattern of results can be explained by motivational declines over the course of the academic year as observed by others (e.g., [7, 10, 20, 30]. As such, a decline in performance may depend on the decline in usage which in turn reflects a decline in motivation. These findings have considerable implications as discussed in the following.

4.1 Implications

Our findings have implications for practitioners, software developers, and policymakers who aim to create more supportive and motivation-enhancing digital learning software environments. These findings are crucial for practitioners such as teachers and school administrators, as they highlight the patterns of students' declining usage and performance when using digital learning software over the course of the academic year. This suggests a need for practitioners to actively monitor and adapt how digital learning software is incorporated in classrooms. Tailored interventions, such as more targeted support, motivational incentives, or adaptions of usage frequency, may help maintain students' continuous usage and improve outcomes over time, ensuring that educational technology is effectively integrated into long-term teaching approaches. For software developers, the present findings highlight the need for creating digital learning software in a way that facilitates to maintain students' long-term usage and performance. This may involve additional adaptive features, or gamification elements that respond to declining motivation and usage over time, as experimental research showed that game elements mitigate students' engagement and long-term usage (e.g., [16]). For policymakers, our findings underscore the relevance of systematic integration of digital learning software into curricula, as shown by the overall higher usage in the Netherlands. Policy interventions may focus on training teachers and students to use these tools effectively throughout the academic year.

4.2 Limitations and future research

There are some limitations to consider when interpreting the current findings. First, we reported on grand averages across all grade levels implemented within the digital learning software. Thus, future research may examine potential differential effects on students' usage and performance trajectories. For instance, one might assume that older students who typically have more experience with digital learning software, as well as better self-regulation capacities, show less (or even no) decline in usage and performance over time. Moreover, it is important to note that students using the digital learning software typically get problems assigned by their teachers. One might speculate that the observed decline in software usage (i.e., fewer problems worked through and fewer active days and weeks) reflects a decrease in teacher-assigned tasks or students' completion of assignments as the academic year progresses, potentially creating a feedback loop that influences how many problems teachers assign. However, as the digital learning software is used very systematically in the Netherlands (even replacing typical workbooks) it seems rather unlikely that teachers incorporate the digital learning software so much less in their teaching as suggested by our results. As such, this account remains speculative and future research is needed to evaluate this explanation. Importantly, future research should also consider to more prominently examine factors that maintain or increase students usage and performance over time. For instance, specific motivational features such as gamification elements may lead to flatter usage and performance declines or may even help to increase motivation and usage [6]. A major benefit of our study is that we evaluated data from two countries that used the same software and observed similar usage and performance declines across both countries. This allows to generalize the trending effect despite the specific integration of the software within a country and despite relatively higher usage in the Netherlands. However, as with other findings from one digital learning software, future research should consider to examine whether our findings generalize across different digital learning software and across other countries.

4.3 Conclusion

Digital learning software is currently implemented in thousands of classrooms across the globe, with hardly any research following students' long-term usage and performance trajectories after its implementation. Our findings indicate usage and performance declines within a classroom-integrated digital learning software over the course of an academic year in two countries. This highlights the need to further consider data from naturalistic settings on long-term usage and performance trajectories. It is imperative to find out whether applied software is used, and if so, how much and how well students perform within it. The present findings are a first step to more research on following students long-term usage and performance trajectories across different digital learning software and countries. Finding out how students' long-term usage and performance can be maintained or even increased seems vital to exploit the full potential of digital learning software.

References

[1] Vincent AWMM Aleven and Kenneth R Koedinger. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science* 26, 2 (2002), 147–179.

[2] John R Anderson, C. Franklin Boyle, and Brian J Reiser. 1985. Intelligent tutoring systems. *Science* 228 (1985), 246–255. https://doi.org/10.4324/9781315617572

[3] John R Anderson, Albert T Corbett, Kenneth R Koedinger, and Ray Pelletier. 1995. Cognitive tutors: Lessons learned. *The journal of the learning sciences* 4, 2 (1995), 167–207.

[4] Lisa Bardach, Korbinian Moeller, Miguel Ruiz-Garcia, Younes Strittmatter, Jennifer Meyer, Sebastian Musslick, and Markus Spitzer10. 2024. Intelligent tutoring systems need real teachers. (2024). https://osf.io/preprints/osf/8y6wm

[5] Kenneth E Barron and Chris S Hulleman. 2015. Expectancy-value-cost model of motivation. *Psychology* 84 (2015), 261–271.

[6] Nathalie Barz, Manuela Benick, Laura Dörrenbächer-Ulrich, and Franziska Perels. 2024. The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis. *Review of Educational Research* 94, 2 (2024), 193–227.

[7] Daria K Benden and Fani Lauermann. 2022. Students' motivational trajectories and academic success in math-intensive study programs: Why short-term motivational assessments matter. *Journal of Educational Psychology* 114, 5 (2022), 1062.

[8] Doug Clow. 2013. An overview of learning analytics. *Teaching in Higher Education* 18, 6 (2013), 683–695.

[9] Albert T Corbett, Kenneth Koedinger, and William S Hadley. 2001. Cognitive tutors: From the research classroom to all classrooms. In *Technology enhanced learning*. Routledge, 215–240.

[10] Jennifer Henderlong Corpus, Kristy A Robinson, and Stephanie V Wormington. 2020. Trajectories of motivation and their academic correlates over the first year of college. *Contemporary Educational Psychology* 63 (2020), 101907.

[11] Jacquelynne S Eccles and Allan Wigfield. 2002. Motivational beliefs, values, and goals. *Annual review of psychology* 53, 1 (2002), 109–132.

[12] Jacquelynne S Eccles and Allan Wigfield. 2020. From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary educational psychology* 61 (2020), 101859.

[13] Jessica Kay Flake, Kenneth E Barron, Christopher Hulleman, Betsy D McCoach, and Megan E Welsh. 2015. Measuring cost: The forgotten component of expectancy-value theory. *Contemporary educational psychology* 41 (2015), 232–244.

[14] Anna Hilz, Karin Guill, Janina Roloff, Daniel Sommerhoff, and Karen Aldrup. 2023. How to Continue? New Approaches to Investigating the Effects of Adaptive Math Learning Programs on Students' Performance, Self-Concept, and Anxiety. *Journal of Intelligence* 11, 6 (2023), 108. https://doi.org/10.3390/jintelligence11060108

[15] Yun Huang, Nikki G Lobczowski, J Elizabeth Richey, Elizabeth A McLaughlin, Michael W Asher, Judith M Harackiewicz, Vincent Aleven, and Kenneth R Koedinger. 2021. A general multi-method approach to data-driven redesign of tutoring systems. In *LAK21: 11th International Learning Analytics and Knowledge Conference*. 161–172.

[16] Stefan E Huber, Rodolpho Cortez, Kristian Kiili, Antero Lindstedt, and Manuel Ninaus. 2023. Game elements enhance engagement and mitigate attrition in online learning tasks. *Computers in Human Behavior* 149 (2023), 107948.

[17] Mohammad Khalil, Paul Prinsloo, and Sharon Slade. 2023. The use and application of learning theory in learning analytics: A scoping review. *Journal of Computing in Higher Education* 35, 3 (2023), 573–594.

[18] Kenneth R Koedinger, John R Anderson, William H Hadley, and Mary A Mark. 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education* 8 (1997), 30–43.

[19] Kenneth R Koedinger, Paulo F Carvalho, Ran Liu, and Elizabeth A Mclaughlin. 2023. An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences* 120, 13 (2023), e2221311120. https://doi.org/10.1073/pnas

[20] So Yeon Lee, Ella Christiaans, and Kristy A Robinson. 2024. Development in context: Comparing short-term trajectories of expectancy, task values, and costs in four university STEM courses. *Contemporary Educational Psychology* (2024), 102282.

[21] Manolis Mavrikis, Nikol Rummel, Michael Wiedmann, Katharina Loibl, and Wayne Holmes. 2022. Combining exploratory learning with structured practice educational technologies to foster both conceptual and procedural fractions knowledge. *Educational Technology Research and Development* 70, 3 (2022), 691–712. https://doi.org/10.1007/s11423-022-10104-0

[22] Martijn Meeter. 2021. Primary school mathematics during Covid-19: No evidence of learning gaps in adaptive practicing results. *Trends in Neuroscience and Education* 25, October (2021), 100163. https://doi.org/10.1016/j.tine.2021.100163

[23] R Core Team. 2013. R: A Language and Environment for Statistical Computing. http://www.r-project.org/

[24] Steve Ritter, Michael Yudelson, Stephen E Fancsali, and Susan R Berman. 2016. How mastery learning works at scale. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. 71–79.

[25] Bethany Rittle-Johnson and Kenneth R Koedinger. 2005. Designing knowledge scaffolds to support mathematical problem solving. *Cognition and Instruction* 23, 3 (2005), 313–349.

[26] Markus Spitzer, Miguel Ruiz-Garcia, Younes Strittmatter, and Korbinian Moeller. 2023. On the difficulty of rational number formats. (2023). https://osf.io/7qebn

[27] Markus Wolfgang Hermann Spitzer. 2022. Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education* (2022). https://doi.org/10.1007/s10212-021-00546-0

[28] Markus Wolfgang Hermann Spitzer, Miguel Ruiz-Garcia, and Korbinian Moeller. 2024. Basic mathematical skills and fraction understanding predict percentage understanding: Evidence from an intelligent tutoring system. *British Journal of Educational Technology* (2024).

[29] Martin Stapel, Zhilin Zheng, and Niels Pinkwart. 2016. An Ensemble Method to Predict Student Performance in an Online Math Learning Environment. In *Proceedings of the 9th International Conference on Educational Data Mining*. 231–238.

[30] Claudia C Sutter, Delaram A Totonchi, Jamie DeCoster, Kenneth E Barron, and Chris S Hulleman. 2024. How does expectancy-value-cost motivation vary during a semester? An intensive longitudinal study to explore individual and situational sources of variation in statistics motivation. *Learning and Individual Differences* 113 (2024), 102484.

[31] Danielle R Thomas, Jionghao Lin, Erin Gatz, Ashish Gurung, Shivang Gupta, Kole Norberg, Stephen E Fancsali, Vincent Aleven, Lee Branstetter, Emma Brunskill, et al. 2024. Improving student learning with hybrid human-AI tutoring: A three-study quasi-experimental investigation. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*. 404–415.

[32] Martin J. Tomasik, Laura A. Helbling, and Urs Moser. 2020. Educational gains of in-person vs. distance learning in primary and secondary schools: A natural experiment during the COVID-19 pandemic school closures in Switzerland. *International Journal of Psychology* 56, 4 (2020), 566–576. https://doi.org/10.1002/ijop.12728

[33] Qin Wang, Amin Mousavi, and Chang Lu. 2022. A scoping review of empirical studies on theory-driven learning analytics. *Distance Education* 43, 1 (2022), 6–29.

[34] Katharina Alexandra Whalen, Alexander Renkl, Alexander Eitel, and Inga Glogger-Frey. 2024. Digital re-attributional feedback in high school mathematics education and its effect on motivation and achievement. *Journal of Computer Assisted Learning* 40, 2 (2024), 478–493.

[35] Allan Wigfield and Jacquelynne S Eccles. 2000. Expectancy–value theory of achievement motivation. *Contemporary educational psychology* 25, 1 (2000), 68–81.