Journal of Computer Assisted Learning

WILEY

Journal of Computer Assisted Learning

ORIGINAL ARTICLE OPEN ACCESS

Intelligent Tutoring Systems Need Teachers

Lisa Bardach[1,2] | Korbinian Moeller[3,4] | Miguel Ruiz-Garcia[5,6] | Younes Strittmatter[7] | Jennifer Meyer[8] | Sebastian Musslick[9,10] | Markus Spitzer[11]

[1]Department of Psychology, University of Giessen, Giessen, Germany | [2]Kaunas University of Technology, Kaunas, Lithuania | [3]Department of Mathematics Education, School of Science, Loughborough University, Loughborough, UK | [4]LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany | [5]Departamento de Estructura de la Materia, Física Térmica y Electrónica, Universidad Complutense de Madrid, Madrid, Spain | [6]Grupo Interdisciplinar de Sistemas Complejos, Universidad Complutense de Madrid, Madrid, Spain | [7]Department of Psychology, Princeton University, Princeton, New Jersey, USA | [8]Department of Teacher Education, University of Vienna, Vienna, Austria | [9]Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany | [10]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island, USA | [11]Department of Cognition and Digital Learning, Martin-Luther University Halle-Wittenberg, Halle, Germany

Correspondence: Lisa Bardach (lisa.bardach@psychol.uni-giessen.de)

Received: 27 June 2025 | **Revised:** 7 October 2025 | **Accepted:** 20 November 2025

Keywords: digital education | digital learning software | dropout | intelligent tutoring system | student engagement

ABSTRACT

Background: The development and distribution of digital learning software, such as intelligent tutoring systems (ITSs), has evolved into a billion-dollar industry, impacting a vast number of students worldwide. A large number of studies on ITSs have focused on their effects on learning outcomes. However, less is known about student engagement and dropout when ITSs are used in real classroom settings over extended periods, particularly with regard to how these patterns may be linked to different assignment scenarios within ITSs.

Objective: The present study aimed to explore whether student engagement and dropout varied based on whether mathematics problems within the ITS were assigned by teachers or self-assigned by students.

Methods: We evaluated rich data from an ITS for learning mathematics used in Germany and the Netherlands (~139,000,000 problems; $n \sim 194{,}000$ students) between 2016 and 2023. To examine whether students' engagement and dropout, both within and between students, varied based on the two assignment scenarios (teacher assigned vs. self-assigned problems), we employed regression and survival analyses.

Results and Conclusions: Our results revealed that, in both Germany and the Netherlands, students with teacher-assigned problems consistently (i) dropped out later, (ii) were active for significantly more weeks and (iii) worked through more mathematics problems each week than those who self-assigned problems. The results were robust across academic school years in the examined period. Overall, our study raises questions about the use of digital learning software as a stand-alone solution and suggests embedding such software in real-life learning scenarios involving teachers.

1 | Introduction

The acquisition of skills in core domains such as mathematics is a fundamental aspect of education and is key to the ability to fully participate in modern societies (Handel 2016; Meehan et al. 2023). Digital learning software, such as intelligent tutoring systems (ITSs), has been developed to facilitate domain-specific learning processes, particularly in mathematics (Anderson et al. 1985; Koedinger et al. 1997; Kulik and Fletcher 2016; Mavrikis et al. 2022; vanLehn 2011; Wang et al. 2023). For example, ITSs provide students with automated feedback, learning aids and additional learning material tailored to individual students' learning progress. By allowing some tasks to be offloaded to the ITS, these systems have

Lay Summary

- What is currently known about this topic?
  - Digital tools like intelligent tutoring systems (ITSs) have been developed to support student learning.
  - Despite large investments, real-world use of ITSs is not yet fully understood.
  - Student dropout and disengagement limit the benefits of digital tools.
  - There is a need to better understand factors linked to dropout and engagement.
- What does this paper add?
  - We examine links between assignment scenarios in ITSs and student engagement/dropout using data from real classroom settings.
  - Students drop out earlier and engage less when they self-assign problems in the ITS than when teachers assign problems to them.
  - We show that these effects appear in both Germany and the Netherlands.
- Implications for practice and/or policy
  - Teachers may need to stay involved when students use ITSs over extended periods.
  - ITSs may be better integrated into classroom practices rather than used in isolation.
  - Involving educators in the design of ITSs may help align these tools with classroom realities.

great potential to assist teachers (Dumont and Ready 2023; Koedinger et al. 1997; Mavrikis et al. 2019; Molenaar 2022; Spitzer and Musslick 2021; Spitzer et al. 2025). To date, there is a large number of studies on the effectiveness of ITSs in promoting students' learning and performance (see e.g., Ma et al. 2014; Xu et al. 2019, for meta-analyses), although it should be acknowledged that most assess immediate in-platform performance outcomes, with relatively few examining distal or transfer outcomes.

Nonetheless, in any case, the effectiveness of digital learning platforms is determined by students' actual engagement with them (i.e., whether students use them). Thus, students' engagement with these systems, as well as their dropout, becomes important additional outcomes in research on digital learning. In line with this, the prediction of student dropout within ITSs and other digital learning applications has consistently been discussed as a critical issue (for reviews see e.g., Ahmad et al. 2023; Chu et al. 2022; Shi et al. 2025). In empirical studies, survival analysis, a statistical method for modelling time to event data, has been employed to examine student dropout over time; for example, by defining dropout as the cessation of interaction with the intelligent tutor before completing all required tasks (Eagle and Barnes 2014). Furthermore, it has been shown that students are clearly distinguishable into categories that denote their patterns of engagement within ITSs (e.g., Quigley et al. 2020). In addition, scholars have developed and tested disengagement tracking systems designed to identify fine-grained disengaged learner behaviours, such as mind-wandering and hasty responding, within ITSs (Chen et al. 2021). Personalised follow-ups have also been implemented in ITSs with the aim of reducing student dropout rates by providing tailored support (Person et al. 2016).

Nevertheless, evidence on the links between different assignment scenarios within ITSs and indicators of students' engagement and dropout is, to the best of our knowledge, currently lacking. Specifically, some ITSs allow teachers to assign tasks to students (e.g., mathematics problems), while also enabling students to self-assign tasks to work through independently. While teacher-assigned problems are likely more strongly embedded in classroom learning and benefit from external scaffolding, self-assigning problems and working through them likely place greater demands on students' self-regulation and motivation (e.g., Fredricks et al. 2004; Reeve 2006; Roll et al. 2025). It seems reasonable to assume that students who self-assign tasks may be more vulnerable to disengagement and dropout; however, this issue has yet to be examined in research on digital learning software integrated in real-life classroom settings.

The present study addressed this gap. We leveraged a big data set from an ITS for mathematics learning (*bettermarks*), including ~194,000 secondary school students who worked on more than ~139 million mathematical problems between 2016 and 2023 in Germany and the Netherlands. Importantly, we took advantage of the circumstance that, in this ITS, students could either self-assign problems for independent learning (assignment scenario: self-assignment) or have teachers assign the problems for them to work through (assignment scenario: teacher-assignment). In addition to contributing to the existing body of knowledge on dropout and engagement in the context of ITSs, research in this area may help generate recommendations for or against the use of ITSs as stand-alone tools. It may also inform frameworks for integrating technology into classrooms and inspire design solutions that support more sustainable use of educational technology.

2 | The Present Study

The present study aimed to add to research on dropout and engagement within ITSs by shedding light on the role of different assignment scenarios within digital learning software. We investigated the following research questions. First, do students' *dropout rates* vary on the basis of the assignment scenario (teacher assignment vs. self-assignment)? We considered students' probabilities of dropping out within a school year in which they used the ITS. This dropout measure provided information about whether the number of days between students' first and last activity in the ITS depended on whether teachers were involved (i.e., assigned problem sets) or not.

To add information on students' *continuous engagement* while using the ITS, we next focused on how many times the ITS was used between the first and last day of use: Does students' continuous engagement in terms of the number of weeks in which the students actively used the ITS depend on the assignment scenario (teacher-assignment vs. self-assignment)? The number of active weeks captured the number of weeks students used the ITS (i.e., worked through at least one problem set[1] in a given week) within the first year after completing the first problem set.

Nevertheless, both the dropout measure and the measure of active weeks fell short of capturing the intensity with which students worked with the ITS. Thus, to add information on how thoroughly students engaged with the ITS, we asked whether students' *engagement intensity* in terms of the average number of mathematics problem sets they worked through per active week depended on the assignment scenario (teacher-assignment vs. self-assignment).

We conducted both between-student analyses (comparing students with teacher-assigned problems to those who self-assigned problems) and within-student analyses (comparing periods when students self-assigned problems to periods when the same students received teacher-assigned problems). Thus, although the study is not experimental in design, its use of naturalistic data enables novel insights into engagement and dropout patterns as they occur in authentic classroom settings. In addition, our data came from Germany and the Netherlands, two countries that differ in their approach to digital education. For example, digital learning software is more widely used in the Netherlands than in Germany (Engzell et al. 2021; van de Werfhorst et al. 2022). This contextual variation enabled us to consider our findings against the backdrop of potential context-specific patterns.

3 | Materials and Method

3.1 | General Information About the Intelligent Tutoring System

The ITS *bettermarks* was first launched in 2008 to help students learn mathematics by providing adaptive feedback, hints and personalised instructions (Spitzer 2022; Spitzer and Musslick 2021). As of fall 2023, *bettermarks* is used primarily by students in Germany and the Netherlands. The content of the ITS covers the curriculum of grades 4–12 (age range 9–18) in Germany and grades 6–9 (age range 11–15) in the Netherlands. Students who use the ITS in the Netherlands typically attend digital learning classes that use digital learning materials. Consequently, these students typically use the ITS more systematically than students in Germany where the ITS is predominantly used to complement traditional learning materials (i.e., textbooks).

3.2 | Learning Content

The learning content is structured hierarchically for each grade level. In particular, the ITS contains books that cover different mathematical topics (e.g., 'Adding and subtracting fractions'; for grade 6 in Germany) for each grade level. Each book comprises several chapters that cover subtopics from the respective book topics (e.g., 'Adding and subtracting fractions with common dominators'). Moreover, each chapter contains several problem sets, each comprising eight to nine problems on average. Answers can be given as free text answers, as multiple-choice answers, or by moving a slider to the correct position. Problems require at least one and up to a maximum of three solution steps. The ITS *bettermarks* covers over 100 different books comprising more than 2000 different problem sets in total.

3.3 | Registration and Use Scenarios

Students who use *bettermarks* are registered through their teachers and may use the ITS in two different ways. Either students can self-assign problem sets for independent learning, or teachers can assign problem sets for their students to work on.

3.4 | Adaptive Features Implemented in the ITS

bettermarks implements the following adaptive features to aid students' learning process. When students work through problems, they are given immediate feedback on whether their response was correct or incorrect. Moreover, teachers also receive feedback on each student's average level of accuracy on each assigned problem set. Furthermore, each problem is accompanied by a hint button that students may press to get additional instructional support via the ITS. Additionally, approximately 20% of problem sets come with content-specific feedback. That is, when students commit a pre-specified error, they receive tailored feedback on the underlying misconception. For instance, when students incorrectly add two fractions with different denominators by separately adding the numerators and denominators, they receive automatic feedback encouraging them to first expand the fractions to a common denominator before adding the numerators only. Moreover, students are incentivised to collect coins and stars by achieving a certain accuracy level on problem sets (i.e., 100% = 1 star, 85%–99% = 3 coins; 75%–84% = 2 coins; 60%–74% = 1 coin). Finally, students can repeat problem sets. However, the parameterization of each problem set changes with each new attempt to prevent students from memorising answers and to motivate them to work through each problem contained in a problem set.

3.5 | Log Data and Data Privacy

bettermarks logs data while it is used. Specifically, *bettermarks* logs the specific problem, as well as the chapter and the book the problem came from. It also logs the accuracy of the students' answers and how the problem sets were assigned (teacher-assigned or self-assigned). Finally, it records the dates that correspond with when the student worked on a specific problem set and when the students and teachers registered to use the software. This information allowed us to analyse when students registered with *bettermarks* and for how long they actively used it.

All users consented to *bettermarks'* terms and conditions with respect to the storing of their data. All the stored data was fully anonymous. Thus, no data can be traced back to any individual student. This anonymizing of the data also meant that we had no demographic information (e.g., age or gender) about the students.

bettermarks shares their data for independent scientific research. *bettermarks* was not involved in the current study's design, and the results do not necessarily represent the opinion of *bettermarks*.

3.6 | Inclusion Criteria and Sample

The data set considered for the present study was selected on the basis of the following inclusion criteria. First, we included all students who used *bettermarks* in the Netherlands and Germany. Second, users had to have registered between 1 July 2016 and 1 September 2022 and had to have worked on problem sets between 1 July 2016 and 31 August 2023. Thus, we considered students who had been registered on *bettermarks* for at least 1 year. Finally, we were interested only in students' activity that took place within the first year after they started to work through the first problem set following their registration. Therefore, we considered only the first 364 days of students' activities. After these inclusion criteria were applied, our final sample analysis comprised 194,911 students who collectively worked through 18,453,537 problem sets (i.e., 139,171,942 single problems). We conducted both within- and between-student analyses. The within-student analysis comprised 18,086 students from Germany who worked on 1,745,022 problem sets (self-assignment: 301,587; teacher-assignment: 1,443,435) and 3193 students from the Netherlands who worked on 1,132,849 problem sets (self-assignment: 121,869; teacher-assignment: 1,010,980). The between-student analysis included 132,470 students (self-assignment: 822; teacher-assignment: 131,648) from Germany who worked on 6,517,072 problem sets (self-assignment: 21,735; teacher-assignment: 6,495,337) and 41,162 students (self-assignment: 386; teacher-assignment: 40,776) from the Netherlands who worked on 9,058,594 problem sets (self-assignment: 8864; teacher-assignment: 9,049,730).

3.7 | Ethics Statement

In this study, we report a retrospective study of archived data provided by *bettermarks*. No sensitive user data were included; it is not possible to track the data back to any software user, and the data are fully anonymous. Institutional ethical approval was not necessary to analyse the anonymous data. All users provided consent for their data to be stored and analysed when they signed up to use the software.

3.8 | Independent and Dependent Variables

We considered two independent variables across all analyses: students' *assignment scenario* (i.e., teacher-assignment vs. self-assignment) and the *country* in which students used *bettermarks* (i.e., Germany vs. the Netherlands). Both were treated as categorical variables.

We considered the following three dependent variables as indicators of students' engagement with the ITS, each of which captures students' use of *bettermarks* from a different point of view to represent students' use of the ITS more comprehensively. First, we considered students' *dropout*, beginning with the first day on which a student worked through a problem set until the last day on which the student worked through a problem set (i.e., the dropout day) as long as the students used the ITS within the first 364 days after registering. This measure was subjected to a survival analysis to determine students' risk of dropping out each day after the first day on which they worked through a problem set.

Second, we considered the number of *active weeks* for each student, reflecting the number of weeks during which students used the ITS (i.e., worked through at least one problem set in a particular week) within the first 364 days after they worked through the first problem set. Third, we focused on the number of *problem sets per week* (i.e., the number of problem sets students worked through in an active week) to reflect the average number of problem sets students worked through per active week.

3.9 | Data Analysis

We used the R environment for statistical data analysis (R Core Team 2013). We ran separate analyses for each of the four dependent variables. Some students engaged in both assignment scenarios and received assignments from their teacher but also self-assigned problem sets. This allowed us to conduct within-student comparisons. However, some students only self-assigned problem sets, whereas other students only received teacher assignments. Thus, we also considered between-student comparisons for these students. As the data set was very large, p values were considered statistically significant only when $p < 0.001$ in all our analyses.

To estimate students' risk of dropping out after first using the ITS and whether differences between *assignment scenario* influenced students' risk of dropping out, we ran a survival analysis that evaluated students' survival probability within the first year as a function of days since the day on which students worked through their first problem set. Importantly, when applying this convention, all students dropped out at some point within the first 364 days. We looked at students' survival probability within only the first 250 days, as we were interested only in students' use and risk of dropping out during the academic school year. We ran the survival analysis for the within-student comparisons and the between-student comparisons with the two main effects of *assignment scenario* and *country*.

We also computed linear regressions to examine the effect of *assignment scenario* and *country* on the number of *active weeks* and *problem sets per week*. We included the main effects and the interaction between the *assignment scenario* and *country* in each of the regression models. The two main effects allowed us to evaluate differences between *assignment scenarios* and *countries*, whereas the interaction term allowed us to evaluate whether differences in *assignment scenario*s were more pronounced in one of the two countries.

Finally, we computed additional survival analyses and regression analyses to examine whether the observed effects were robust across years (see Figures S1–S3 and Tables S1–S3). The data analysed in our study and all analysis codes can be found on the Open Science Framework (OSF, anonymised link for peer review: https://osf.io/8cunk/?view_only=eb6931d494484cbeb99fcc0085c6e924).

4 | Results

For all our engagement measures (dropout, number of active weeks, average number of problem sets worked through per active week), we conducted both a within-student analysis and a between-student analysis. Some students received assignments from their teacher, but also self-assigned problem sets during the first 364 days of use. This allowed us to conduct within-student comparisons. However, some students only self-assigned problem sets, whereas other students only received teacher assignments, which allowed us to conduct between-student comparisons for these students. It was possible to conduct these analyses, as the ITS stores information about the type of assignment (i.e., teacher-assignment or self-assignment) for each problem set. All the students in the between-student analysis were free to use both assignment scenarios, but either their teachers did not assign problems to them in the ITS, or the students chose never to use the ITS for their own learning purposes.

Figure 1 depicts the results of survival analyses for capturing dropout, reflecting students' probabilities of surviving during the first 250 days of using the ITS. The period of 250 days was chosen to approximate the number of actual school days for which students should still be engaged (because the summer holidays had not started yet; see also Figures S1 and S2 for the robustness analyses). Figure 2 illustrates the results of the linear regression analyses with respect to the number of active weeks and the number of problems per week that students worked through (see Figure S3 for the robustness analyses). Table 1 presents the statistical details of the linear regression analyses (see Tables S1–S3 for the robustness analyses). Figure 3 also depicts the combined information on the number of active weeks and the number of problem sets students worked through in a 2-D illustration. Details on the results of each set of analyses are reported below.

4.1 | Dropout

The survival analysis indicated that students' probabilities of dropping out were considerably lower when problem sets were assigned by teachers than when students self-assigned the problem sets. This finding was consistent across Germany and the Netherlands (see Figure 1). We observed this effect for both the within-student and between-student analyses when applying a Cox proportional hazard model. For the within-student comparison, the coefficient of *assignment scenario* was −0.592 ($p < 0.001$), indicating a hazard ratio of 0.55, which suggests that the risk of dropping out for students who self-assigned problem sets was more than twice as high per day of actively using the ITS compared with students whose problem sets were assigned by teachers ($p < 0.001$). The coefficient for *country* was −0.08 ($p < 0.001$), indicating a hazard ratio of 0.92, which suggests that students had an 8% higher risk of dropping out in Germany than in the Netherlands.

For the between-student comparison, the coefficient for the assignment scenario was −0.757, indicating a hazard ratio of 0.47 ($p < 0.001$). This value suggests that the risk of dropping out for students who self-assigned problem sets was approximately twice as high per day compared with students whose problem sets were assigned by teachers. The coefficient for *country* was −0.07, indicating a hazard ratio of 0.93 ($p < 0.001$). This value

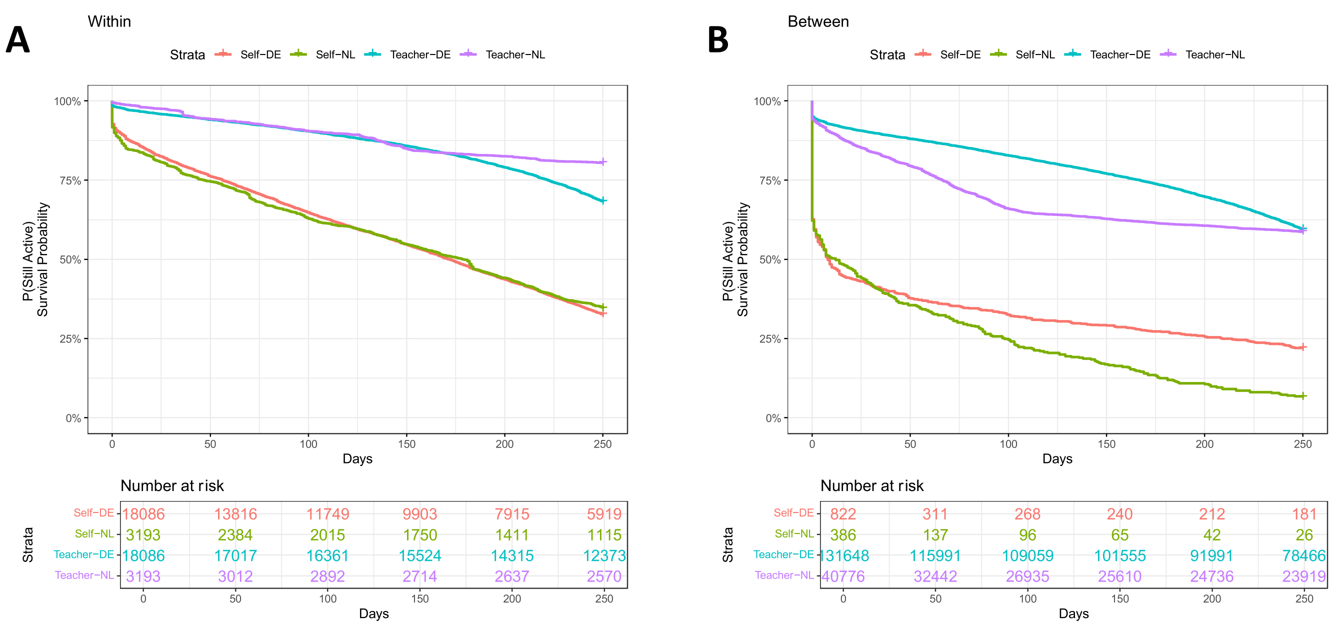

FIGURE 1 | Survival functions for the two assignment scenarios for (A) the within-student comparison and (B) the between-student comparison. Solid lines indicate the survival curves for students who self-assigned the problem sets (labelled as Self-DE for Germany in red and Self-NL for the Netherlands in green) and students whose problem sets were assigned by teachers (labelled as Teacher-DE in turquoise and Teacher-NL in purple). Shaded areas (if visible) indicate the 95% confidence interval. The strata below each subplot reflect the absolute number of active students for every 50-day interval since students' first activity on *bettermarks*.

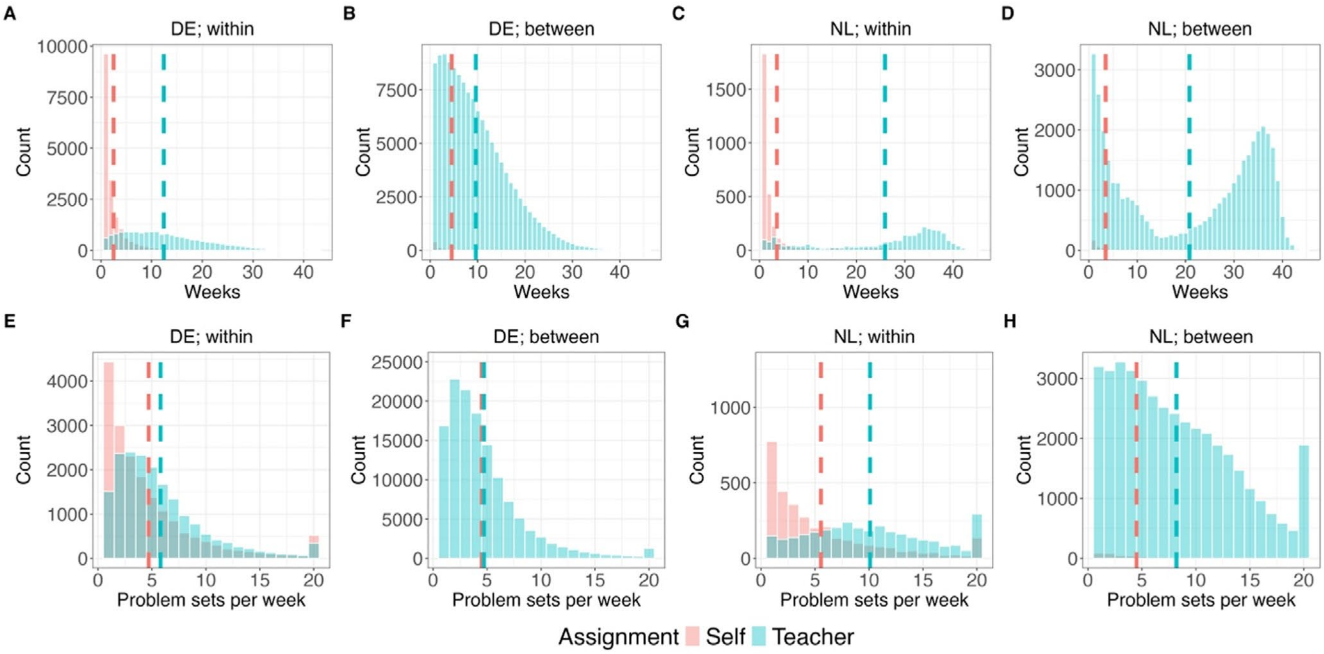

FIGURE 2 | Histograms depicting weeks, problem sets and problem sets per week as a function of assignment scenario and country. Results of all analyses indicate that students actively used the ITS for longer (more weeks) and with greater intensity (more problem sets per week; except for the German between-student analysis) when teachers assigned the problem sets than when students self-assigned the problem sets. Vertical dashed lines illustrate the means for each assignment scenario.

means that students had a 7% higher risk of dropping out in Germany than in the Netherlands.

4.2 | Weeks of Active Use of the ITS

The within-student analysis revealed that students in Germany were active for 12 weeks when the problems were assigned by their teachers (between-student analysis: 9 weeks) but only for 2 weeks when they self-assigned the problems (between-student analysis: 4 weeks). These differences indicate that, when focusing on within-student comparisons, students used the ITS for 10 more weeks (between-student analysis: 5 more weeks) when teachers assigned the problem sets. A similar result was observed in the Netherlands. The within-student analysis showed that students in the Netherlands used the ITS for 25 weeks (between-student analysis: 20 weeks) when the teachers assigned the problem sets but only 3 weeks (between-student analysis: 3 weeks) for self-assigned problems, indicating an even larger 22-week difference for the within-student analysis in using the ITS (between-student analysis: 17 more weeks).

These differences in average weeks were substantiated by the linear regression results (see Figure 2), with a significant main effect of *assignment scenario* (see Table 1 for all regression results), indicating more weeks of active use of the ITS for when teachers assigned problem sets compared with students' self-assignments. The significant main effect of *country* indicated that students used the ITS for more weeks in the Netherlands than in Germany. The significant interaction between *assignment scenario* and *country* indicated that the effect of assignment scenario was larger in the Netherlands than in Germany.

4.3 | Number of Problem Sets Completed Per Week

In Germany, the within-student analysis showed that students worked through four problem sets per week (between-student analysis: four per week) when the problem sets were self-assigned and five problem sets per week when teachers assigned the problem sets (between-student analysis: four per week). This difference shows that students in Germany worked through similar numbers of problem sets per week in each assignment scenario (a difference of one more problem set per week was observed in the within-student comparison, whereas no difference was observed in the between-student comparison).

The results from the within-student analysis indicated that students worked through more problem sets per week in the Netherlands than in Germany, with students who self-assigned the problem sets working through 5 problem sets per week (between-student analysis: 4 per week), whereas students whose problem sets were assigned by their teachers worked through 10 problem sets per week (between-student analysis: 8 per week) in the Netherlands.

The linear regression results indicate a significant main effect of *assignment scenario*, with students working through more problem sets per week when teachers assigned the problem sets than when students self-assigned the problem sets (see Table 1). The main effect of *country* was also significant, indicating that students worked through more problem sets per week in the Netherlands than in Germany. The interaction between *assignment scenario* and *country* was significant as well, and it revealed that the effect of students working through more problem sets per week when teachers assigned the problems was larger in the Netherlands than in Germany.

TABLE 1 | Linear regression results for assignment scenario (teacher-assignment vs. self-assignment), country (Netherlands vs. Germany) and their interaction predicting the number of average weeks and the number of problem sets students worked through per week.

Predictors	Weeks (within)				Weeks (between)				Problem sets/week (within)				Problem sets/week (between)			
	b	SE	*t*	*p*	*b*	SE	*t*	*p*	*b*	SE	*t*	*p*	*b*	SE	*t*	*p*
Intercept	11.12	0.05	246.50	**<0.001**	9.62	0.14	68.38	**<0.001**	6.53	0.03	198.87	**<0.001**	5.49	0.06	88.78	**<0.001**
Scenario	8.04	0.05	178.14	**<0.001**	5.59	0.14	39.74	**<0.001**	1.42	0.03	52.41	**<0.001**	0.95	0.06	15.35	**<0.001**
Country	3.66	0.05	81.04	**<0.001**	2.57	0.14	18.23	**<0.001**	1.29	0.03	39.33	**<0.001**	0.88	0.06	14.28	**<0.001**
Scenario × Country	3.07	0.05	68.03	**<0.001**	3.06	0.14	21.76	**<0.001**	0.89	0.03	32.98	**<0.001**	0.88	0.06	14.25	**<0.001**
R^2			0.512				0.219				0.088				0.123	

Note: p-values are < .001 (in bold).

5 | Discussion

National, regional and local education systems and schools worldwide allocate substantial resources to digital learning software with the goals of improving the learning outcomes of millions of students, reforming the education system and enhancing global competitiveness and economic growth (Burns 2023). Nevertheless, despite their significant cost and widespread adoption, substantial gaps in current knowledge about digital learning software, such as ITSs, remain. Specifically, although numerous studies on ITSs have explored their effects on learning outcomes, much less is known about student engagement and dropout in real-world classroom settings over extended periods. In particular, the relationship between these patterns and different assignment scenarios (teacher assignments of tasks within the ITSs vs. self-assignment) is insufficiently understood.

The present study addressed this gap. We considered three indicators that captured different aspects of students' dropout and (lack of) engagement with the ITS: Dropout, students' continuous engagement in terms of the number of active weeks, and students' engagement intensity in terms of the number of problem sets they completed per active week.

Our results provide robust and consistent evidence on (i) dropout, (ii) active weeks and (iii) intensity of ITS use, highlighting large differences depending on the assignment scenario. First, without teacher assignments, the risk of dropping out was twice as high as it was when teachers assigned the problem sets (within-student results and between-student results). Second, converging results were observed for students' continuous engagement in terms of the number of weeks they actively used the ITS, such that students were actively engaged with the ITS for longer when teachers assigned problem sets compared to when students self-assigned them. Third, the results for engagement intensity showed that students worked through a comparable or larger number of problem sets per week when the teacher assigned the problem sets to them, compared with when the students self-assigned the problem sets.

Potential reasons for the findings may be linked to a lack of relevant self-regulation skills and students' motivation to remain engaged with the ITS without teacher support (Azevedo et al. 2012; Bardach et al. 2023; Cai et al. 2025; Järvelä et al. 2021; Melnikoff et al. 2022). Prior research has emphasised that self-regulated learning skills are essential for navigating digital environments effectively (Azevedo et al. 2012; Järvelä et al. 2021). Without teacher support, many students may lack the metacognitive strategies and motivational resources needed to persist, particularly over extended periods of use. This interpretation seems consistent with our data: students were more likely to stay engaged longer and complete more problem sets when teachers assigned the content. At the same time, even academically strong and self-regulated learners may experience disengagement in the absence of teacher presence (Jaggars and Xu 2016). As we did not have information on students' self-regulated learning and other relevant characteristics, we cannot directly test these assumptions in our data. Moreover, motivational theories suggest that the absence of social accountability and feedback in self-assigned

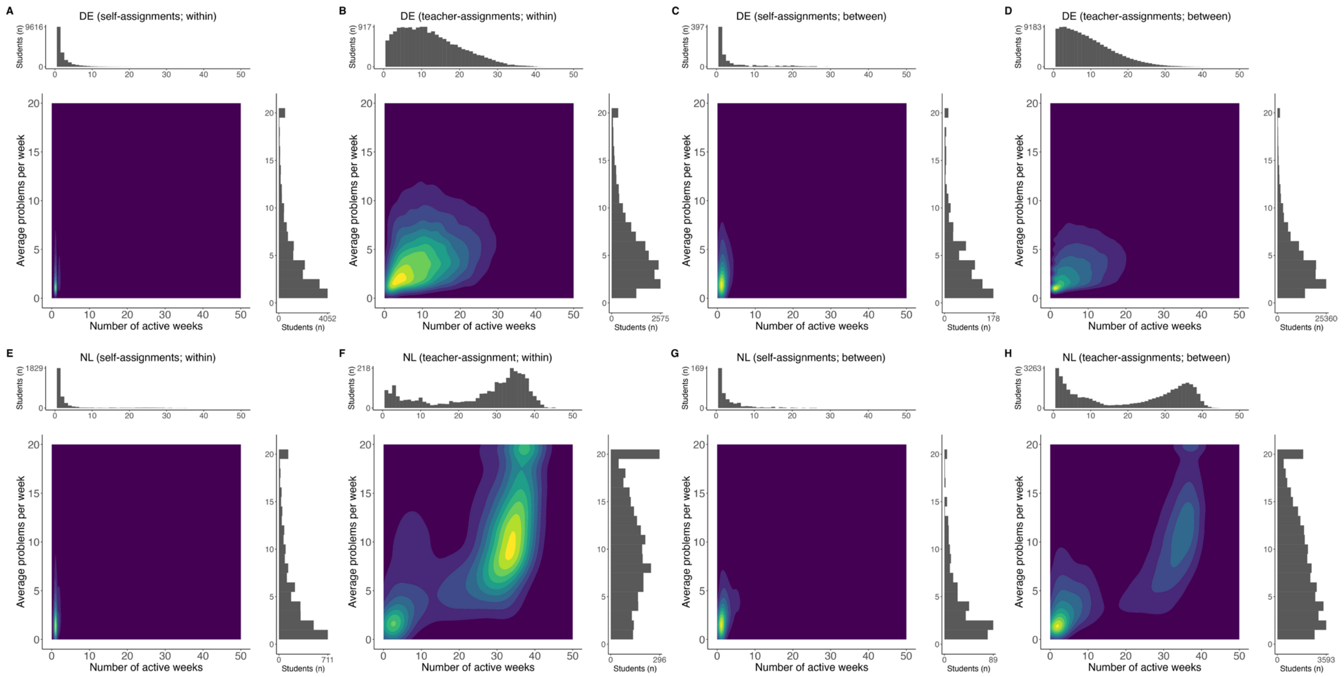

FIGURE 3 | The average number of problem sets students worked through per active week as a function of active weeks for Germany (DE) and the Netherlands (NL) and self-assigned versus teacher-assigned problem sets (for both within-student and between-student comparisons). Amount of use by German students when self-assigning problem sets (within-student comparison) (A)/(between-student comparison) (E); when problem sets were assigned by teachers (within-student comparison) (B)/(between-student comparison) (F); Amount of use by students from the Netherlands when self-assigning problem sets (within-student comparison) (C)/(between-student comparison) (G); when problem sets were assigned by teachers (within-student comparison) (D)/(between-student comparison) (H); High use is depicted in yellow. Low use is depicted in blue. Each subplot has its individual density scale for better visibility. When more than 20 problem sets were worked through per active week, they were still represented as 20 for illustrative purposes.

scenarios may lower students' perceived value and relevance of the tasks, contributing to early disengagement (Eccles and Wigfield 2020; Ryan and Deci 2000). When students are left to self-direct their learning without adequate support structures, those with weaker self-regulatory capacities may struggle disproportionately, leading to increased dropout rates (Efklides 2011; Zimmerman 2000). In contrast, teacher assignments likely serve both as an external motivator and a source of guidance, helping students sustain effort and engagement over time. Our findings reinforce the importance of integrating human support into technology enhanced learning environments, a factor that we suspect may be particularly important for students who do not yet have the skills or motivation to stay engaged with ITSs independently.

We also noted several interesting differences between the two countries we investigated. For instance, the risk of dropout in Germany was higher than in the Netherlands for both analyses, which may be linked to the better integration of digital learning software in classrooms in the Netherlands (Engzell et al. 2021; van de Werfhorst et al. 2022). Moreover, the effect for continuous engagement seemed to be more pronounced in the Netherlands where students had 22 more active weeks with teacher support in the within-student comparisons and 17 more in the between-student comparisons. In Germany, students were active for nearly 10 more weeks when the problems were assigned by their teachers than when they were self-assigned (within-student comparisons), and they were active for nearly five more weeks when the teachers assigned the problems (between-student comparisons). It may be that because ITSs are implemented more systematically in the Netherlands across different subjects (Engzell et al. 2021; Meeter 2021; van de Werfhorst et al. 2022), students are able to compare the ITS-related teacher support they receive in different subjects. Therefore, students' continuous engagement when using the ITS investigated herein for mathematics learning may suffer more from a lack of teacher support when using this specific ITS. Another reason for the differences between countries may be linked to general differences in the patterns of use between Germany and the Netherlands, such that German students were much less active, even when supported by their teachers (discussed in more detail for engagement intensity below; see also Figure 3). Finally, comparing the effects on engagement intensity between Germany and the Netherlands indicated that, even when teacher assignment was in place, the number of problem sets students worked through was low (see Figure 2). Compared with the Netherlands, Germany is characterised by a lower degree of technological preparedness and digitalization in education (Engzell et al. 2021; van de Werfhorst et al. 2022). Relatedly, German teachers might not be as prepared to deliver digital education and may be less proficient in doing so than their Dutch colleagues. Furthermore, there is evidence from international comparisons that the digital competencies of German students are comparatively weaker than those of students from other countries (Fraillon et al. 2014). All these issues may limit the extent to which German students actively engaged with the ITS even when they were scaffolded by the teacher. Overall, the results may have been influenced by country-specific differences, such as the integration of technology and broader characteristics of the school systems. Thus, it might not yet be feasible to generalise the findings from one context to another.

Our findings have several potential implications. For educational policy and practice, the results indicate that teachers may need to be actively involved, especially when students are working with ITSs for an extended period of time. Rather than replacing teachers and serving as isolated stand-alone solutions, ITSs thus appear to require thoughtful integration into classroom practices by the teacher. For the development of digital learning software, it is possible that in order to make ITS use sustainable, the role of teachers should be considered as part of the development process from the very beginning, and that ultimately, the design of ITSs should take teachers' needs and preferences into account (Coburn and Penuel 2016; Howard et al. 2021). While teacher presence is often limited in ITSs, several features could help address this gap and could feasibly be implemented in the ITS we studied. For example, real-time teacher dashboards, assignment customization tools and in-system teacher–student feedback loops (e.g., Holstein et al. 2018; Roschelle et al. 2016) offer promising avenues for enhancing teacher and social presence and could be systematically tested through teacher-involved interventions.

In general, in order to guarantee that ITSs are proficiently incorporated into teaching, teachers need to embody the appropriate skills and motivations (Rubach and Lazarides 2021), which should ideally be systematically fostered from teachers' early career stages onward (European Commission 2020; OECD 2015). For research, our findings suggest that close examinations of students' dropout and engagement with digital technology offer a promising approach. As most research on ITSs has focused on achievement outcomes (Kulik and Fletcher 2016), we propose that dropout and engagement with ITSs should be considered as an additional, influential outcome that deserves researchers' full attention (Bartelet et al. 2016; Chevalère et al. 2023).

6 | Limitations and Future Research Directions

Our study has several limitations, which should be considered when interpreting the findings and should be addressed in future research. Even though our analyses were based on a rich data set that included millions of worked-through problem sets, no sociodemographic or other information about the students was available. For example, it is possible that socioeconomically disadvantaged students are in even greater need of teacher support and may thereby be disproportionately prone to dropping out, thus further increasing inequities in education (Dumont and Ready 2023). Students' abilities may also be a decisive factor influencing their ITS use: students with higher abilities may be more likely to engage with the ITS and self-assign more problems than students with lower abilities, who may need more teacher support to remain engaged with digital learning software. Also, we lacked contextual information (e.g., the socioeconomic composition of schools and regions). We therefore envision future research that combines student background and contextual data with ITS data. Also, we discussed motivational problems and self-regulation failures as potential causes of a lack of engagement; however, we could not test these claims empirically. Thus, there is a need for future studies to triangulate behavioural data from ITSs with self-reports on students' motivational and self-regulatory states, among other data sources (Dijkstra et al. 2023; Klose et al. 2024). Such research that combines different sources of data is necessary to illuminate the psychological processes that underlie our findings. Another limitation stems from the fact that, in our between-student analyses, the group of students who used the ITS with the support of their teachers was much larger than the group of students who used the ITS without their teachers. At the same time, given that all students could use the ITS on their own, it is an interesting additional finding that the vast majority chose not to do so. Finally, we compared only two countries, and future studies on teacher support and ITS engagement should explore cross-country and cross-cultural variations in more depth.

7 | Conclusions

In conclusion, a wide range of studies have established that ITSs facilitate students' learning and can support teachers (Koedinger et al. 1997, 2023; Kulik and Fletcher 2016; Spitzer and Musslick 2021; Wang et al. 2023). However, a dearth of studies using data from real-world learning contexts has explored the role of different assignment scenarios (teacher-assigned problems vs. self-assignment of problems within the ITS) and their role for dropout and engagement. Our study, which leveraged extensive data from real classrooms to provide an ecological snapshot of actual ITS usage, was designed to address this gap. Our findings strongly indicate that dropout and engagement vary systematically depending on the two different assignment scenarios. If we want students to remain engaged with ITSs, and, thus, to be able to reap the benefits in terms of learning gains, the support of teachers seems to be key.

Author Contributions

The Lisa Bardach, Korbinian Moeller, and Markus Spitzer conceptualized the study. Markus Spitzer performed the statistical analyses. The Lisa Bardach and Markus Spitzer wrote the paper. All other authors contributed to reviewing and editing the paper.

Acknowledgments

Open Access funding enabled and organized by Projekt DEAL.

Funding

Lisa Bardach is supported by a Jacobs Foundation Research Fellowship, a Fellowship from the Baden-Württemberg Foundation (Eliteprogram for Post-docs), and and research funding No P-EDU-23-28, which is co-funded by the European Union (the project “Breakthrough in Educational Research” No 10-044-P-0001) under the 1st April 2025 Agreement with the Research Council of Lithuania (RCOL, LMTLT) and the 14th April 2025 Joint Activity Agreement with Kaunas University of Technology. This work was supported by a Research Grant from HFSP (Ref.-No: RGEC33/2024; https://doi.org/10.52044/HFSP.RGEC332024.pc.gr.194170).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data analysed in our study and all analysis codes can be found on the Open Science Framework (OSF, https://osf.io/8cunk/?view_only=eb6931d494484cbeb99fcc0085c6e924).

Endnotes

[1] Problem sets contain an average of eight to nine single mathematical problems (see Method).

References

Ahmad, K., W. Iqbal, A. El-Hassan, et al. 2023. "Data-Driven Artificial Intelligence in Education: A Comprehensive Review." *IEEE Transactions on Learning Technologies* 17: 12–31.

Anderson, J. R., C. F. Boyle, and B. J. Reiser. 1985. "Intelligent Tutoring Systems." *Science* 228, no. 4698: 456–462. https://doi.org/10.1126/science.228.4698.456.

Azevedo, R., R. F. Behnagh, J. M. Harley, and G. Trevors. 2012. "Metacognition and Self-Regulated Learning in Student-Centered Learning Environments." In *Theoretical Foundations of Learning Environments*, 171–197. Routledge.

Bardach, L., T. Yanagida, T. Goetz, H. Jach, and R. Pekrun. 2023. "Self- and Externally Regulated Learning in Adolescence Self- and Externally Regulated Learning in Adolescence. Developmental Trajectories and Relations With Teacher Behavior, Parent Behavior, and Adacemic Achievement." *Developmental Psychology* 59, no. 7: 1327–1345. https://doi.org/10.1037/dev0001537.

Bartelet, D., J. Ghysels, W. Groot, C. Haelermans, and H. Maassen van den Brink. 2016. "The Differential Effect of Basic Mathematics Skills Homework via a Web-Based Intelligent Tutoring System Across Achievement Subgroups and Mathematics Domains: A Randomized Field Experiment." *Journal of Educational Psychology* 108, no. 1: 1–20. https://doi.org/10.1037/edu0000051.

Burns, M. 2023. "Barriers and Supports for Technology Integration: Views From Teachers." https://doi.org/10.54676/LCEB5308.

Cai, Z., C. Liu, J. Zhan, et al. 2025. "Effect of Collaboration in Digital Game-Based Learning: The Roles of Flow Experience and Intrinsic Motivation." *Journal of Computer Assisted Learning* 41, no. 4: e70065.

Chen, S., Y. Fang, G. Shi, et al. 2021. "Automated Disengagement Tracking Within an Intelligent Tutoring System." *Frontiers in Artificial Intelligence* 3: 595627.

Chevalère, J., H. S. Yun, A. Henke, N. Pinkwart, V. V. Hafner, and R. Lazarides. 2023. "A Sequence of Learning Processes in an Intelligent Tutoring System From Topic-Related Appraisals to Learning Gains." *Learning and Instruction* 87, no. 2: 101799. https://doi.org/10.1016/j.learninstruc.2023.101799.

Chu, H. C., G. H. Hwang, Y. F. Tu, and K. H. Yang. 2022. "Roles and Research Trends of Artificial Intelligence in Higher Education: A Systematic Review of the Top 50 Most-Cited Articles." *Australasian Journal of Educational Technology* 38, no. 3: 22–42.

Coburn, C. E., and W. R. Penuel. 2016. "Research–Practice Partnerships in Education: Outcomes, Dynamics, and Open Questions." *Educational Researcher* 45, no. 1: 48–54. https://doi.org/10.3102/0013189X16631750.

Dijkstra, S. H. E., M. Hinne, E. Segers, and I. Molenaar. 2023. "Clustering Children's Learning Behaviour to Identify Self-Regulated Learning Support Needs." *Computers in Human Behavior* 145: 107754. https://doi.org/10.1016/j.chb.2023.107754.

Dumont, H., and D. D. Ready. 2023. "On the Promise of Personalized Learning for Educational Equity." *npj Science of Learning* 8, no. 1: 26. https://doi.org/10.1038/s41539-023-00174-x.

Eagle, M., and T. Barnes. 2014. "Modeling Student Dropout in Tutoring Systems." In *Intelligent Tutoring Systems: 12th International Conference, ITS 2014, Honolulu, HI, USA, June 5–9, 2014*, 676–678. Springer International Publishing.

Eccles, J. S., and A. Wigfield. 2020. "From Expectancy-Value Theory to Situated Expectancy-Value Theory: A Developmental, Social Cognitive, and Sociocultural Perspective on Motivation." *Contemporary Educational Psychology* 61: 101859.

Efklides, A. 2011. "Interactions of Metacognition With Motivation and Affect in Self-Regulated Learning: The MASRL Model." *Educational Psychologist* 46, no. 1: 6–25.

Engzell, P., A. Frey, and M. D. Verhagen. 2021. "Learning Loss due to School Closures During the COVID-19 Pandemic." *Proceedings of the National Academy of Sciences of the United States of America* 118, no. 17: e2022376118. https://doi.org/10.1073/PNAS.2022376118.

European Commission. 2020. "Education and Training Monitor 2020—Teaching and Learning in a Digital Age." Publications Office of the European Union. https://data.europa.eu/doi/10.2766/917974.

Fraillon, J., J. Ainley, W. Schulz, T. Friedman, and E. Gebhardt. 2014. "Preparing for Life in a Digital Age: The IEA International Computer and Information Literacy Study International Report." In *Preparing for Life in a Digital Age*. Springer International Publishing. https://doi.org/10.1007/978-3-319-14222-7.

Fredricks, J. A., P. C. Blumenfeld, and A. H. Paris. 2004. "School Engagement: Potential of the Concept, State of the Evidence." *Review of Educational Research* 74, no. 1: 59–109. https://doi.org/10.3102/00346543074001059.

Handel, M. J. 2016. "What Do People Do at Work? A Profile of U.S. Jobs From the Survey of Workplace Skills, Technology, and Management." *Journal of Labor Market Research* 49, no. 2: 177–197. https://doi.org/10.1007/s12651-016-0213-1.

Holstein, K., B. M. McLaren, and V. Aleven. 2018. "Student Learning Benefits of a Mixed-Reality Teacher Awareness Tool in AI-Enhanced Classrooms." In *International Conference on Artificial Intelligence in Education*, 154–168. Springer International Publishing.

Howard, S. K., J. Tondeur, J. Ma, and J. Yang. 2021. "What to Teach? Strategies for Developing Digital Competency in Preservice Teacher Training." *Computers in Education* 165: 104149. https://doi.org/10.1016/j.compedu.2021.104149.

Jaggars, S. S., and D. Xu. 2016. "How Do Online Course Design Features Influence Student Performance?" *Computers & Education* 95: 270–284. https://doi.org/10.1016/j.compedu.2016.01.014.

Järvelä, S., J. Malmberg, E. Haataja, M. Sobocinski, and P. A. Kirschner. 2021. "What Multimodal Data Can Tell Us About the Students' Regulation of Their Learning Process?" *Learning and Instruction* 72: 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004.

Klose, M., P. Handschuh, D. Steger, and C. Artelt. 2024. "Not Easy to Get Off Track: Motivational Tajectories of Learners Completing a Non-Formal Online Course." *Computers in Human Behavior* 159: 108322. https://doi.org/10.1016/j.chb.2024.108322.

Koedinger, K. R., J. R. Anderson, W. H. Hadley, and M. A. Mark. 1997. "Intelligent Tutoring Goes to School in the Big City." *International Journal of Artificial Intelligence in Education* 1, no. 1: 30–43.

Koedinger, K. R., P. F. Carvalho, R. Liu, and E. A. Mclaughlin. 2023. "An Astonishing Regularity in Student Learning Rate." *Proceedings of the National Academy of Sciences* 120, no. 13: e2221311120. https://doi.org/10.1073/pnas.2221311120.

Kulik, J. A., and J. D. Fletcher. 2016. "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review." *Review of Educational Research* 86, no. 1: 42–78. https://doi.org/10.3102/0034654315581420.

Ma, W., O. O. Adesope, J. C. Nesbit, and Q. Liu. 2014. "Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis." *Journal of Educational Psychology* 106, no. 4: 901–918.

Mavrikis, M., E. Geraniou, S. Gutierrez Santos, and A. Poulovassilis. 2019. "Intelligent Analysis and Data Visualisation for Teacher Assistance Tools: The Case of Exploratory Learning." *British Journal of Educational Technology* 50, no. 6: 2920–2942. https://doi.org/10.1111/bjet.12876.

Mavrikis, M., N. Rummel, M. Wiedmann, K. Loibl, and W. Holmes. 2022. "Combining Exploratory Learning With Structured Practice Educational Technologies to Foster Both Conceptual and Procedural Fractions Knowledge." *Educational Technology Research and Development* 70, no. 3: 691–712. https://doi.org/10.1007/s11423-022-10104-0.

Meehan, L., G. Pacheco, and T. Schober. 2023. "Basic Reading and Mathematics Skills and the Labour Market Outcomes of Young People: Evidence From PISA and Linked Administrative Data." *Economic Record* 99, no. 327: 473–491. https://doi.org/10.1111/1475-4932.12755.

Meeter, M. 2021. "Primary School Mathematics During Covid-19: No Evidence of Learning Gaps in Adaptive Practicing Results." *Trends in Neuroscience and Education* 25: 100163. https://doi.org/10.1016/j.tine.2021.100163.

Melnikoff, D. E., R. W. Carlson, and P. E. Stillman. 2022. "A Computational Theory of the Subjective Experience of Flow." *Nature Communications* 13, no. 1: 2252. https://doi.org/10.1038/s41467-022-29742-2.

Molenaar, I. 2022. "The Concept of Hybrid Human-AI Regulation: Exemplifying How to Support Young Learners' Self-Regulated Learning." *Computers and Education: Artificial Intelligence* 3: 100070. https://doi.org/10.1016/j.caeai.2022.100070.

OECD. 2015. *Students, Computers and Learning: Making the Connection*. OECD Publishing. https://doi.org/10.1787/factbook-2015-68-en.

Person, P., T. Galinho, H. Boukachour, F. Lecroq, and J. Grieu. 2016. "Dynamic Representation and Interpretation in a Multiagent 3D Tutoring System." In *Recent Contributions in Intelligent Systems*, edited by V. Sgurev, R. R. Yager, J. Kacprzyk, and K. T. Atanassov, 205–227. Springer International Publishing.

Quigley, D., D. Caccamise, J. Weatherley, and P. Foltz. 2020. "Exploring Video Engagement in an Intelligent Tutoring System." In *Adaptive Instructional Systems: Second International Conference, AIS 2020, Held as Part of the 22nd HCI International Conference, HCII 2020*, 519–530. Springer International Publishing.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing [Computer Software]." R Foundation for Statistical Computing. https://www.R-project.org/.

Reeve, J. 2006. "Teachers as Facilitators: What Autonomy-Supportive Teachers Do and Why Their Students Benefit." *Elementary School Journal* 106, no. 3: 225–236. https://doi.org/10.1086/501484.

Roll, I., E. S. Wiese, Y. Long, V. Aleven, and K. R. Koedinger. 2025. "Tutoring Self- and Co-Regulation With Intelligent Tutoring Systems to Help Students Acquire Better Learning Skills." In *Advances in Intelligent Tutoring Systems*, edited by R. Nkambou, R. Azevedo, and J. Vassileva. Springer.

Roschelle, J., M. Feng, R. F. Murphy, and C. A. Mason, 2016. "Online Mathematics Homework Increases Student Achievement." *AERA Open* 2, no. 4: 2332858416673968.

Rubach, C., and R. Lazarides. 2021. "Addressing 21st-Century Digital Skills in Schools–Development and Validation of an Instrument to Measure Teachers' Basic ICT Competence Beliefs." *Computers in Human Behavior* 118: 106636. https://doi.org/10.1016/j.chb.2020.106636.

Ryan, R. M., and E. L. Deci. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." *American Psychologist* 55, no. 1: 68–78.

Shi, H., N. Zhang, S. Caskurlu, and H. Na. 2025. "Applications of Machine Learning for At-Risk Student Prediction in Online Education: A 10-Year Systematic Review of Literature." *Journal of Computer Assisted Learning* 41, no. 4: e70058.

Spitzer, M. W. H., and S. Musslick. 2021. "Academic Performance of K-12 Students in an Online-Learning Environment for Mathematics Increased During the Shutdown of Schools in Wake of the Covid-19." *PLoS One* 16, no. 8: 1–16. https://doi.org/10.1371/journal.pone.0255629.

Spitzer, M. W. H. 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* 37, no. 1: 39–53. https://doi.org/10.1007/s10212-021-00546-0.

Spitzer, M. W. H., K. Moeller, E. Richter, L. Frede, and L. Bardach. 2025. "A Red Button Does It All: A Red Button Nudges Students to Close Their Knowledge Gaps in an Intelligent Tutoring System." *Learning and Instruction* 99: 102181.

van de Werfhorst, H. G., E. Kessenich, and S. Geven. 2022. "The Digital Divide in Online Education: Inequality in Digital Readiness of Students and Schools." *Computers and Education Open* 3, no. 3: 100100. https://doi.org/10.1016/j.caeo.2022.100100.

vanLehn, K. 2011. "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems." *Educational Psychologist* 46, no. 4: 197–221. https://doi.org/10.1080/00461520.2011.611369.

Wang, H., A. Tlili, R. Huang, et al. 2023. "Examining the Applications of Intelligent Tutoring Systems in Real Educational Contexts: A Systematic Literature Review From the Social Experiment Perspective." *Education and Information Technologies* 28, no. 7: 9113–9148. https://doi.org/10.1007/s10639-022-11555-x.

Xu, Z., K. (K.) Wijekumar, G. Ramirez, X. Hu, and R. Irey. 2019. "The Effectiveness of Intelligent Tutoring Systems on K-12 Students' Reading Comprehension: A Meta-Analysis." *British Journal of Educational Technology* 50, no. 6: 3119–3137.

Zimmerman, B. J. 2000. "Attaining Self-Regulation: A Social Cognitive Perspective." In *Handbook of Self-Regulation*, edited by M. Boekaerts, P. R. Pintrich, and M. Zeidner, 13–39. Academic.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** jcal70159-sup-0001-Supinfo.docx.