Learning and Instruction 99 (2025) 102141

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Adaptive teaching with technology enhances lasting learning

Leonie Sibley[a,b,*,1], Armin Fabian[a,2], Christine Plicht[a,c,d,3], Lisa Pagano[a,4], Niklas Ehrhardt[a,e,5], Luisa Wellert[a,6], Thorsten Bohl[a,7], Andreas Lachner[a,8]

[a] University of Tübingen, Germany
[b] Zurich University of Teacher Education, Switzerland
[c] Hans-Küng-Gemeinschaftsschule Tübingen, Germany
[d] Minna-Specht-Gemeinschaftsschule, Germany
[e] Schickhardt-Gemeinschaftsschule, Germany

ARTICLE INFO

Keywords:
Adaptive teaching
Educational technology
Co-design
Lasting learning

ABSTRACT

Background: A central goal in education is to productively handle heterogeneity among students. Adaptive teaching, which integrates formative assessment, differentiation, and individual support, is a promising model for addressing student diverse prerequisites such as prior knowledge. While educational technology is increasingly employed to implement adaptive teaching, there is limited knowledge on whether its effectiveness generalizes across different subjects and educational contexts.
Aims: This study aims to investigate the immediate and long-term effects of technology-enhanced adaptive teaching on students' cognitive, metacognitive, and motivational outcomes, and crucially, whether the effectiveness can be generalized across various educational settings. Additionally, we examined whether these effects depend on specific boundary conditions (prior knowledge, domain, degree of adaptivity).
Sample: The study involved $N = 656$ students from grades 7 to 12 ($M = 14.91$ years, 45 % female) across various subjects ($k = 12$ teaching units) from six different schools in Germany.
Methods: Using the Localize-Generalize-Transfer (LoGeT) framework, we adopted a comprehensive implementation approach combining co-design and ManyClasses approaches. We compared the effects of technology-enhanced adaptive teaching with a business-as-usual control condition, focusing on learning, monitoring accuracy, interest, and self-efficacy.
Results: We found that students in the technology-enhanced adaptive teaching condition outperformed students in the business-as-usual condition regarding delayed but not immediate learning outcomes. None of the other effects were significant. Interestingly, subject domain and the level of adaptive elements moderated the effect.
Conclusions: This research addresses a gap in understanding the generalizability of technology-enhanced adaptive teaching, demonstrating its broader applicability and contribution to lasting learning outcomes.

This article is part of a special issue entitled: Adaptivity published in Learning and Instruction.

* Corresponding author. University of Tübingen, Germany.
E-mail addresses: leonie.sibley@phzh.ch (L. Sibley), armin.fabian@uni-tuebingen.de (A. Fabian), christine.plicht@uni-tuebingen.de (C. Plicht), lisa.pagano@outlook.de (L. Pagano), niklas.ehrhardt@uni-tuebingen.de (N. Ehrhardt), luisa.wellert@uni-tuebingen.de (L. Wellert), thorsten.bohl@uni-tuebingen.de (T. Bohl), andreas.lachner@uni-tuebingen.de (A. Lachner).
[1] Leonie Sibley (technology-enhanced adaptive teaching, generative learning): Lagerstrasse 2, 8090 Zurich, Switzerland.
[2] Armin Fabian (technology-enhanced adaptive teaching, TPACK): Europastrasse 6, 72072 Tübingen, Germany.
[3] Christine Plicht (technology-enhanced adaptive teaching): Carl-Diem-Strasse 112, 72760 Reutlingen, Germany.
[4] Lisa Pagano (technology-enhanced adaptive teaching): Wilhelmstrasse 31, 72074 Tübingen, Germany.
[5] Niklas Ehrhardt (technology-enhanced adaptive teaching): Schickhardtstrasse 30, 70199 Stuttgart, Germany.
[6] Luisa Wellert (technology-enhanced adaptive teaching, ITS): Wilhelmstrasse 31, 72074 Tübingen, Germany.
[7] Thorsten Bohl (adaptive teaching): Wilhelmstrasse 31, 72074 Tübingen, Germany.
[8] Andreas Lachner (teaching and learning with technologies): Wilhelmstrasse 31, 72074 Tübingen, Germany.

https://doi.org/10.1016/j.learninstruc.2025.102141
Received 18 July 2024; Received in revised form 13 March 2025; Accepted 14 April 2025
Available online 13 June 2025

1. Introduction

Recent large-scale attainment studies have demonstrated that schools often struggle to handle the heterogeneity among students effectively (Enchikova et al., 2024; OECD, 2023). Therefore, supporting students in their individual learning processes based on their needs regarding their prerequisites (e.g., prior knowledge) is key to fostering their educational achievement. A prominent instructional model that explicitly addresses students' heterogeneity regarding their prerequisites is adaptive teaching (Corno, 2008). Adaptive teaching systematically combines formative assessment, differentiation (i.e., macro adaptation), and individual support (i.e., micro adaptation) to enhance students' meaningful and lasting learning. In this context, educational technologies can be promising tools to help implement adaptive teaching.

Previous research has primarily focused on advanced and fully automated technologies, such as intelligent tutoring systems (Keuning & van Geel, 2021; Kulik & Fletcher, 2016; Molenaar & Knoop-van Campen, 2019; van Leeuwen et al., 2019). These sophisticated approaches generally involved single interventions covering a limited set of topics in specific subjects, which makes it difficult to generalize the findings across different subjects and contexts in which such technology is not available and teachers have to use available technologies to realize adaptive teaching (Sibley et al., 2024).

In the current study, we therefore aimed to address this gap by investigating a) the immediate and long-term effects of technology-enhanced adaptive teaching (compared to a business-as-usual control condition) on students' cognitive (i.e., knowledge outcome), metacognitive (i.e., monitoring accuracy), and motivational (i.e. interest, self-efficacy) outcomes and b) whether the obtained effects depend on specific boundary conditions such as students' prerequisites (i.e., prior knowledge) and class-level conditions (i.e., domain, degree of adaptivity). We applied a comprehensive implementation approach, the Localize-Generalize-Transfer (LoGeT) framework (Lachner et al., 2024), combining co-design approaches with ManyClasses approaches to examine the generalizability of adaptive teaching across subject-specific classroom settings ($k = 12$ teaching units). The generalizability of adaptive teaching was tested in different educational settings across various subjects (e.g., Mathematics, ethics) and grades (7–12).

2. Adaptive teaching

Adaptive teaching has been a topic of discussion in educational practice for several decades and is still highly relevant today (Corno, 2008; Tetzlaff et al., 2021). Unlike traditional one-size-fits-all approaches, the concept of adaptive teaching involves tailoring teaching methods to meet the specific needs of students to improve their learning outcomes (Brod, 2024; Corno, 2008; Karst et al., 2022; Schmid et al., 2022; Tetzlaff et al., 2021). Following adaptive approaches (cf. Corno, 2008; Tetzlaff et al., 2021), we particularly drew on the framework proposed by Sibley et al. (2024), in which adaptive teaching constitutes three iterative constructs: formative assessment, macro-level adaptations, and micro-level adaptations.

Formative assessment is regarded as the backbone of adaptive teaching (Sibley et al., 2024). Formative assessments constitute informal and continuous diagnoses with the aim of monitoring students' current learning progress. These assessments commonly include instructional activities such as quizzes, tests, or observations that provide teachers with insights into students' learning progress. Simultaneously, these assessments give students feedback on their own learning. This regular reflection of their performance can help students develop a more accurate assessment of their own knowledge and understanding, contributing to improved monitoring accuracy (i.e., judging their own knowledge accurately). For instance, formative assessments, such as digital student response systems, can offer real-time feedback and prompts that encourage self-reflection, making students more aware of their learning processes and helping them develop better monitoring accuracy (Molin et al., 2022; Shin et al., 2022). Formative assessments provide continuous data on students' learning progress, which teachers can use to provide subsequent macro- and micro adaptations that are tailored to the students' current learning needs. While our earlier examples focused on cognitive aspects, such as prior achievement and performance, adaptive teaching can address multiple facets of student heterogeneity, including cognitive, motivational, and metacognitive dimensions (Sibley et al., 2024).

These adaptations do not follow a strict sequential order but can occur simultaneously or in any sequence as needed, depending on the immediate instructional context and student needs. Adaptive teaching is an iterative process, where formative assessments continuously inform instructional decisions at both the macro and micro levels. The use of educational technologies (such as testing environments or learning analytics) has the potential to provide instant assessment for subsequent macro- or micro adaptations (Brod, 2024; Sibley et al., 2024). Mostly, formative assessments are based on cognitive learning progresses (e.g., learning strategies, current level of knowledge or skills), but may also involve motivational or affective characteristics (Hammer et al., 2021; Jonassen, 2005).

Based on the results of the formative assessments, the teachers can apply *adaptations on a macro-level* (Corno, 2008). Macro-level adaptations typically involve pre-planned strategies that are applied to entire groups of students, ensuring that all members of a group receive similar instructional approaches (Sibley et al., 2024). Most commonly, these groupings are based on their learning levels, such as assigning lower-achieving students to groups that receive more foundational tasks and instructional support, while higher-achieving students work on more advanced tasks. Macro-level adaptations typically involve not only differentiated learning materials but also adjustments in the overall lesson structure and group compositions to address diverse learning needs. Educational technologies, such as learning paths within learning management systems, can offer technical opportunities to provide structured and personalized sequences of instructional activities that are differentiated to meet students' needs. These macro adaptations align well with principles like the expertise-reversal effect (Kalyuga, 2007, Renkl & Atkinson, 2003) and aptitude-treatment interactions, which suggest that the effectiveness of a particular instructional task (e.g., using worked examples versus problem-solving) depends on the level of students' prior knowledge (see Fyfe & Rittle-Johnson, 2016; Nievelstein et al., 2013; Richter et al., 2016, for empirical evidence).

However, these principles are equally relevant for micro-level adaptations, which teachers can implement in real time. Unlike macro-level adaptations, which focus on structuring lessons and organizing groups, micro-level adaptations provide moment-to-moment instructional support that is tailored to individual students' needs during the lesson. For instance, if a teacher notices that a student is struggling with a task, they might offer immediate, tailored feedback or additional materials to help the student overcome specific difficulties. This individualized approach, often based on informal, formative assessments, adjusts instruction to address specific situations as they arise. Technology can help provide students with such individualized support by giving automated computer-based feedback or prompts, for instance by an instructional agent (Mertens et al., 2022; Sibley et al., 2024; Tetzlaff et al., 2021; Wagner et al., 2024).

3. Effectiveness of adaptive teaching

From a theoretical perspective, adaptive teaching is often compared to the large learning gains in one-to-one tutoring (Bloom et al., 1984; Chi et al., 2001; VanLehn, 2011; Wittwer et al., 2010, see also Brod, 2024). Adaptive teaching is considered a strategy of providing individualized learning opportunities to many students, helping them reach their zone of proximal development (Vygotsky, 1978). However, there is limited empirical evidence on the effectiveness of adaptive teaching.

Additionally, the benefits of available educational technology are not well-documented. An exception is the study by Brühwiler and Blatchford (2011). In their correlational study, the authors asked teachers ($N = 49$) to answer an adaptive teaching competency test and subsequently assessed students' learning gains ($N = 898$). The authors found that adaptive teaching competency was significantly related to students' learning outcomes ($\beta = 1.22$), while controlling for students' pre-requisites or class-composition. However, whether these findings may be attributed to adaptive teaching or the use of technology, is an open question and requires further controlled designs in which the use of technology and adaptive teaching is systematically combined.

With a specific focus on educational technologies, there are controlled experimental studies which examined the effect of intelligent tutoring systems (ITS) and personal learning technologies (PLT) on learning. Intelligent tutoring systems are self-contained technologies that formatively diagnose current learners' progress (i.e., formative assessment) and provide adaptive instruction (i.e., micro adaptation), often without the need for specific teacher intervention (Aleven & Koedinger, 2002; Anderson et al., 1995; Brusilovsky et al., 1996; Graesser et al., 2004; Meurers et al., 2019). Several meta-analyses have shown positive effects of intelligent tutoring systems, with effect sizes from small to large (Hillmayr et al., 2020; Kulik & Fletcher, 2016; Ma et al., 2014; Steenbergen-Hu & Cooper, 2014). For instance, Ma et al. (2014) found small to medium effects of ITS in their meta-analysis ($0.36 < g < 0.41$). Interestingly, the authors reported a large heterogeneity among studies, indicating that ITS are not effective in general but might depend on further boundary conditions. The authors therefore conducted moderator analyses and explored study and student characteristics, such as students' prior knowledge and subject domain. Students' prior knowledge did not moderate the ITS effect. The authors argued that this non-effect needs to be interpreted with caution, since most of the studies included in their meta-analysis did not report students' prior knowledge or only contained students with low but not with higher prior knowledge limiting the potential of the moderator analysis. Interestingly, results demonstrated that the effect of ITS indeed differed among subject domain, indicating large effect sizes in learning humanities, moderate effects for most of the STEM subjects (i.e., biology, computer science, physics, mathematics), and small to moderate effect sizes for literacy and language learning and chemistry. These findings highlight that it is crucial to consider study and student characteristics as moderator variables in research on adaptive learning.

Besides Ma et al. (2014), also more recent studies demonstrated the effectiveness of ITS approaches. Alrawashdeh et al. (2024) demonstrated a small effect size ($g = 0.29$), and Hooshyar et al. (2024) reported a medium effect size ($g = 0.42$). In contrast, Zheng et al. (2021) showed a large effect size (ES = 0.812). Notably, the effects of ITS were of similar size to those of one-on-one human tutoring or small group discussions (VanLehn, 2011), indicating that ITS might be as beneficial as traditional tutoring methods. Furthermore, it is important to note that ITS are primarily utilized in STEM domains (Bernacki et al., 2021), but there is also considerable diversity in the conceptualizations of these technologies (van Schoors et al., 2021). Given that ITS are most exclusively implemented in individualized higher education settings, where a student only interacts with the ITS system, it remains an open question whether the findings can generalize to authentic classroom teaching that involve complex interactions and structures among subjects, students, teachers, and the educational technology implemented within the classroom.

To generalize the effects of educational technology for authentic adaptive teaching, Sibley et al. (2024) conducted a mixed-methods study. Following the *Localize-Generalize-Transfer (LoGeT) framework* by Lachner et al. (2024), the authors systematically combined co-design methodologies (localization strategies) and ManyClasses approaches (generalization strategies) with co-constructive transfer activities to generate empirical evidence that could be applicable in educational practice. The LoGeT model integrates three essential stages: localization, generalization, and transfer. In the *localization stage*, co-design approaches bring together educators and researchers to collaboratively design interventions, ensuring that the educational strategies are grounded in real-world contexts and tailored to specific classroom needs. This participatory process helps make the interventions more relevant and effective in diverse instructional settings by actively involving teachers as key stakeholders. The *generalization stage* utilizes principles from the ManyClasses approach (Fyfe et al., 2021), which involves conducting the same experimental setup across multiple educational settings, subjects, and student groups. Instead of focusing on one context, ManyClasses studies test educational principles across varied real-world classrooms, allowing researchers to examine how interventions perform in different environments. For example, Fyfe et al. (2021) used this approach to investigate feedback timing across 38 diverse classes, thereby enhancing the ecological validity and generalizability of their findings. Finally, the *transfer stage* focuses on ensuring that the findings and interventions are accessible and applicable beyond the original study settings. This involves creating practical resources, such as instructional guides or multimedia content, to help educators implement the strategies effectively in their own classrooms. By synthesizing these stages, the LoGeT model aims to bridge the gap between localized empirical studies and widespread educational practice, promoting the scalability of evidence-based educational interventions.

Following the LoGeT model, Sibley et al. (2024) realized a four-year co-design project in which the authors and teachers of different subjects (i.e., mathematics, physics, chemistry, German, English, Spanish, ethics) collaboratively developed adaptive teaching units based on a joint adaptive design framework (Corno, 2008). The authors used a sequential-explanatory mixed-methods approach and combined a quantitative study ($N = 183$), with a qualitative study using semi-structured teacher interviews ($N = 3$) to test the feasibility of adaptive teaching with technology. Overall, the authors found small learning gains from pre- to posttests ($\beta = 0.22$). Additionally, the authors explored potential boundary conditions of adaptive teaching. Boundary conditions refer to contextual or individual factors that influence the effectiveness of adaptive teaching. These conditions determine the extent to which adaptive methods can optimize learning for diverse students in varied settings. For instance, factors such as students' prior knowledge, subject domain, or the degree of adaptivity can shape the impact of adaptive teaching. The results of these analyses showed that students' prior knowledge initially moderated their learning gains, with lower-performing students benefiting the most from adaptive teaching units. However, when controlling for the number of adaptive teaching elements (i.e., degree of adaptivity: number of formative assessments, macro and micro adaptations), prior knowledge was no longer a significant moderator. This suggests that high fidelity in the implementation of adaptive teaching can reduce the impact of individual learning differences, such as prior knowledge. Additional analyses of the realized teaching units showed that the fidelity of the realized adaptive teaching units was substantially correlated with students' learning gains ($\beta = 0.11$). These correlational findings can be regarded as a first step toward the effectiveness of technology-enhanced adaptive teaching. However, due to the absence of a business-as-usual control group, it remains unclear whether the observed learning gains can be attributed to the intervention or to general maturation effects (see Karst et al., 2022, for related approaches in differentiated instruction). Additionally, since the authors only applied an immediate posttest, little is known about potential lasting learning effects of technology-enhanced adaptive teaching. Nevertheless, effects on lasting learning can be assumed as adaptive teaching should allow to establish a rich mental representation of the learning content, which should contribute to a better long-term retrieval (Roelle et al., 2022, 2023).

4. The present study

Previous research indicates a lack of empirical studies that systematically examined the combination of adaptive teaching and the use of technologies on both immediate and lasting student learning in real classroom settings. Furthermore, as previous settings were restricted to specific student populations and subjects, there is a lack of evidence that allows to generalize the obtained findings across subjects and student populations. Sibley et al. (2024) represents an exception, however, the authors did not apply a control group and no delayed posttest, leaving it unclear, whether technology-enhanced adaptive teaching is better than regular teaching regarding students' immediate but also long-term learning.

To close these research gaps, we followed the LoGeT framework (Lachner et al., 2024) and systematically combined a *co-design* (Roschelle et al., 2006) and the *ManyClasses approach* (Fyfe et al., 2021) to test the effectiveness of technology-enhanced adaptive teaching as compared to business-as-usual teaching. Comparing adaptive teaching to a business-as-usual condition ensures that a potential benefit of adaptive teaching is meaningful and practical for real-world educational environments.

In the intervention condition, teachers participated in a professional development program on adaptive teaching with educational technology, where they designed adaptive teaching units across different subjects and grade levels. The development of these teaching units took approximately two to three months. During this period, teachers worked in tandems, exchanging feedback and collaboratively refining their instructional design. At its core, the intervention focused on the implementation of adaptive elements across the four levels: formative assessments, macro-level and micro-level adaptations, and consolidation phases. Teachers were specifically trained to use digital tools as a means to facilitate these adaptive elements, ensuring that technology supported, but did not dictate, the adaptivity of the instruction.

Following the ManyClasses approach, these units were intentionally diverse, varying in subjects, learning goals, and grade levels, and therefore not directly comparable. This diversity was a deliberate choice, as the goal was to investigate whether the concept of technology-enhanced adaptive teaching could be effectively transferred across different educational contexts. To ensure implementation fidelity, the research team played an active role in the co-design process. Teachers received regular support meetings, in which the research team provided evidence-based recommendations for integrating adaptivity into their lesson plans. For example, teachers were introduced to research findings on the aptitude-treatment interactions, which suggest that the effectiveness of an instructional task depends on students' prior knowledge (see Fyfe & Rittle-Johnson, 2016; Nievelstein et al., 2013; Richter et al., 2016, for empirical evidence). In more detail, teachers were informed, for instance, that worked examples are more beneficial for lower-performing students and problem-solving tasks for higher-performing students. Additionally, the research team assisted teachers in identifying and integrating digital tools that could facilitate adaptive elements within their units. These tools included formative assessment tools, platforms for organizing group work, and automated feedback systems in digital learning environments. Given the differences in subjects and learning objectives, we did not prescribe a fixed number of adaptive elements. Instead, we collaborated with the teachers to determine where and how adaptive elements, such as formative assessments or macro- and micro-level adaptations, would be most beneficial in their specific units. This natural variability in the design led us to analyze the degree of adaptivity as a moderator variable. This structured yet flexible co-design approach ensured that the adaptive teaching units were both theoretically grounded and practically feasible.

We used a pre-post-delayed-test design with a control group (parallel class of the respective technology-enhanced adaptive teaching class) to examine the effectiveness of technology-enhanced adaptive teaching regarding students' immediate and long-term (meta-)cognitive and motivational outcomes. We stated the following pre-registered hypotheses (https://aspredicted.org/xk33-327w.pdf):

4.1. Research question 1: Does technology-enhanced adaptive teaching contribute to students' learning?

We hypothesized that students who attended the technology-enhanced adaptive teaching units outperform students who attended the business-as-usual unit regarding their immediate and long-term a) cognitive outcomes (i.e., knowledge outcome), b) meta-cognitive outcomes (i.e., monitoring accuracy), c) and motivational outcomes (i.e., interest, self-efficacy).

4.2. Research question 2: Do boundary conditions moderate the technology-enhanced adaptive teaching effect?

We additionally explored whether student-level characteristics (i.e., prior knowledge) and class-level factors (i.e., domain, degree of adaptivity) moderated the effect of technology-enhanced adaptive teaching to test its generalizability.

5. Method

5.1. Participants and design

We contacted all community schools with upper secondary levels (Gemeinschaftsschulen) in South-Western Germany ($N = 8$) to invite them to collaborate and participate in the study. Of these, four schools agreed to participate. Additionally, two secondary schools (Realschulen) from the local area expressed interest in the study and also took part, bringing the total to six schools that we collaborated with. Over the course of two years, we worked with a total of 10 teachers from these schools. With these teachers we co-designed the teaching units, and the teachers then taught the unit in their own classrooms. For each intervention class, a parallel class from the same school was used as the control group. Thus, we applied a quasi-experimental approach, as we implemented the adaptive teaching units in authentic classroom setting, which did not allow random assignment of classes to our treatment. We used a pre-post-delayed-test design to examine students' immediate and lasting learning (four-week delay) outcomes.

In total, 656 school students participated in the study. As determined by an a-priori-power analysis, we needed 351 students at least to detect small effects with a power of 80 %. We planned to develop and evaluate 14 teaching units. For each teaching unit, a parallel class will serve as a natural control group. We calculate with 15 students per class, so that about 420 students will participate in the study. To detect a small ($f = 0.15$) to medium effect ($f = 0.25$) with a power of 80 % by controlling for class and prior knowledge, a sample size of 128–351 students is needed. Aiming at collecting an additional 20 percent of participants to control for potential clustering effects, we planned to collect data of 421 students. Thus, our sample of 656 students exceeded the required sample size.

The mean age of the students was 14.91 years ($SD = 2.03$), 45 % were female, 53 % were male, and 2 % were diverse. On average, they stated medium to high interest in the specific subject ($M = 2.61$, $SD = 0.63$, on a scale from 1 to 4) and high self-efficacy of using technology ($M = 3.25$, $SD = 0.67$, on a scale from 1 to 4).

5.2. Teaching units

In total, the teachers developed 12 technology-enhanced adaptive teaching units (see Table 1). To support the development of the teaching units, we realized a co-design approach and implemented a long-term professional development program, consisting of regular workshops and feedback sessions, in which the implementation of adaptive teaching was discussed from a research and educational practice perspective.

Table 1
Overview of the teaching units.

Teaching Unit	Subject	Topic	Knowledge Type	Answer Format	Domain	*N*	Grade Level	DA CG | EG	Age	Female	Omega
1	Mathematics	Functions	D + P	C	STEM	24	12	19 | 20	17.69 (0.75)	31 %	0.80
2	Mathematics	Euclidean plane	D + P	C + O	STEM	43	12	21 | 22	17.61 (0.55)	42 %	0.79
3	Mathematics	Integral calculus	D + P	C + O	STEM	39	12	16 | 16	17.65 (0.60)	53 %	0.65
4	Geography	Desert	D	C + O	STEM	62	7	16 | 19	12.41 (0.54)	33 %	0.69
5	Geography	Tropical rainforest	D	C + O	STEM	54	7	15 | 23	12.33 (0.48)	15 %	0.79
6	Social science	Human rights	D	C + O	Humanities	62	11	26 | 14	17.00 (0.90)	38 %	0.80
7	Ethics	Justice	D	C + O	Humanities	27	12	20 | 15	17.71 (0.85)	67 %	0.78
8	English	Cover letter	D + P	C + O	Language	126	9	14 | 20	14.56 (0.70)	52 %	0.71
9	German	Ballads	D + P	C + O	Language	60	7	19 | 24	13.54 (1.07)	54 %	0.83
10	German	Newspaper	D	C	Language	62	8	9 | 13	14.18 (0.78)	42 %	0.61
11	German	Drama	D + P	C	Language	49	10	14 | 23	15.74 (0.61)	49 %	0.83
12	German	Argumentation	D	C + O	Language	48	8	19 | 24	13.83 (1.31)	60 %	0.89

Note. In Unit 8, two teachers collaboratively developed a unit, and each implemented it with their respective classes, resulting in a larger sample size. D = Declarative knowledge. P = Procedural knowledge. C = Closed answer format. O = Open answer format. DA = Degree of Adaptivity.

As a basis, we used the adaptive framework by Sibley and colleagues (2024). We systematically placed adaptive elements (i.e., formative assessment, macro adaptation, micro adaptation, consolidation phases) within all the units, and the teachers used various technologies (e.g., online quizzes, LearningApps, etc.) to support them in conducting assessments, organizing group work, and providing automated feedback. The teaching units were implemented over three to four weeks, depending on whether the subject was a core subject with more weekly hours (duration: three weeks) or a regular subject with fewer hours (duration: four weeks). This approach ensured a relatively consistent total number of instructional hours across units. While the teaching units were pre-planned during the co-design process, the teachers were still able to respond flexibly to situations in the classroom. This allowed them to make adaptive adjustments as needed, even beyond what was initially planned, ensuring that they could address specific student needs in real time.

In each teaching unit, we began with a formative assessment to activate students cognitively and assess their current knowledge level. Based on this assessment, the teacher could group students according to their learning levels and provide differentiated instructions and tasks. For instance, lower-achieving students received basic tasks with more instructions and support, while higher-achieving students worked on more complex transfer tasks with less guidance. After some practice and consolidation, another formative assessment was conducted, allowing the teacher to make further adaptations based on students' progress.

This can be illustrated with the mathematics unit on the "Euclidean plane" (see Appendix B for more details). In the first lesson, the teacher used an online quiz on the topic "The Cross Product – Transforming Plane Equations" to assess students' understanding (formative assessment). Based on the quiz results, students were grouped by different proficiency levels (macro-adaptation) and worked collaboratively on tasks at varying difficulty levels. They received computer-based feedback on their answers (micro-adaptation). In a follow-up lesson, students learned to visualize planes. In this self-regulated learning phase, they were tasked with creating short explanatory videos to explain the contents. These videos received peer feedback and served as another formative assessment to gauge students' understanding and address any misconceptions through group discussions.

In the third lesson, students focused on understanding the relative positions of planes and lines. Again, the lesson started with a formative assessment using an online quiz to check what students had retained about visualizing planes. Based on the results, students worked in homogeneous pairs on tasks using GeoGebra, with lower-achieving students receiving more support while higher-achieving students worked more independently (macro-adaptation). In the following lesson, students practiced further, received individual support and feedback (micro-adaptation). The lesson concluded with another formative assessment, and the results were used to adjust the number of practice tasks in the final lesson (macro-adaptation). In this concluding lesson, the learning objectives were reviewed once more using a final formative assessment.

5.3. Measures

5.3.1. Learning outcome

We tested students' knowledge in a pretest, posttest, and delayed test with a four-week delay. We used parallel versions of the tests to avoid retention effects (Gross et al., 2013). Together with the teachers, we designed the tests for each unit to align with the curricular requirements and learning objectives, ensuring assessments that targeted similar cognitive skills and knowledge types (e.g., declarative versus procedural knowledge). Unlike traditional approaches that standardize difficulty across units, ManyClasses studies test whether observed effects remain consistent across different conditions, supporting broader applicability (Fyfe et al., 2021). This diversity provides valuable insights into the robustness of educational interventions, even with varying difficulty levels and contexts. The tests included closed and open answer formats (see Table 1). Open answers were coded by two independent raters. The interrater reliability was excellent among the teaching units ($ICC_{2,1}$: 0.91–1.00). Students' knowledge was calculated in percentages. Thus, students could achieve zero to 100 percent as a final score. An example of a knowledge test can be found in Appendix A. Overall, the reliabilities of the tests were satisfying or better (see Table 1).

5.3.2. Monitoring accuracy

We assessed students' monitoring accuracy as the second dependent variable. We asked students to rate their own performance in the pre-, post-, and delayed test with the following item: "You will now solve six tasks of the topic Functions. For each correct answer, you will get two points. Thus, you can get a total score of twelve points. How many points do you think you will get?" based on Baars et al. (2017) and Jacob et al. (2022) on a scale from zero points to twelve points. We then subtracted their judgement in percent by their real score (also in percent) which they got in the comprehension tests (bias index, see Schraw, 2009). Thus, negative values (maximum: −100) indicate an underestimate, while positive values (maximum: 100) indicate an overestimation. Zero represents an accurate judgement.

5.3.3. Subject-specific interest

We measured students' interest in the specific subject as the third dependent variable with three items (e.g., "Mathematics captures my attention") adapted from Knogler et al. (2015) on a scale from 1 "not at all" to 4 "completely" (ω = 0.85).

5.3.4. Self-efficacy of using technology

Self-efficacy of using technology in the specific subject was measured with three items (e.g., "I think that I can use educational technology properly to learn Mathematics") adapted from Backfisch et al. (2024) and Rigotti et al. (2008) on a scale from 1 "not at all" to 4 "completely" (ω = 0.83).

5.3.5. Degree of adaptivity

We investigated the number of adaptive elements within each unit as an indicator for the degree of adaptivity and included it as a moderator in our analyses (see also 2.3.5, Kast et al., 2022; Sibley et al., 2024). We asked the teachers to state how many times they implemented formative assessments, adaptations on a macro and micro level, and consolidation phases with a self-generated item for each category (e.g., "How often did you implement a formative assessment during the teaching unit?"). We calculated a sum score as an indicator for the degree of adaptivity (ω = 0.76).

5.3.6. Teaching quality and quantity

We assessed students' perspective on the teaching quality (i.e., cognitive activation, classroom interruption, classroom monitoring, instructional support) to analyze whether the quality was comparable among conditions.

5.3.6.1. Cognitive activation. We assessed students' cognitive activation with the scale of Fauth et al. (2014) which included six items (e.g., "In the unit Functions, my teacher asked me questions that I had to think about very carefully") on a scale from 1 "not at all" to 4 "completely" (ω = 0.73).

5.3.6.2. Classroom interruption. We assessed students' perception of the classroom interruption with the scale by Hammer et al. (2021) which includes three items (e.g., "In the unit Functions, the students often interrupted lessons") on a scale from 1 "not at all" to 4 "completely" (ω = 0.88).

5.3.6.3. Classroom monitoring. We assessed students' perception of the classroom monitoring with the scale by Hammer et al. (2021) which includes four items (e.g., "In the unit Functions, our teacher was always aware of everything happening in the classroom") on a scale from 1 "not at all" to 4 "completely" (ω = 0.84).

5.3.6.4. Instructional support. We assessed students' perception of the instructional support with the scale by Hammer et al. (2021) which includes five items (e.g., "In the unit Functions, our teacher gave us additional support whenever we needed it") on a scale from 1 "not at all" to 4 "completely" (ω = 0.82).

5.4. Procedure

Prior to data collection, received permission from the ethics committee of our university, from the regional Ministry of Education, and the school principals to conduct the study. Additionally, we informed all school students and their legal guardians (e.g., parents) about the aim and procedure of the study and collected their written consent to participate in the study. The participation was voluntary. Students who refused to participate could attend the teaching unit as usual without taking part in data collection. Thus, these participants had no disadvantages. Data was collected and stored in line with the GDPR-guidelines on German servers.

To test the effectiveness of the technology-enhanced adaptive teaching units, we additionally collected data from the parallel class as control condition. The classes were selected based on similar student characteristics and the same curriculum level. The participating teachers agreed with the teachers of the control group on the same learning objectives to ensure that both classes were taught the same contents. Besides this information, the teachers did not share any details of the teaching unit to avoid spillover-effects. Thus, the parallel classes represented the same prerequisites and were taught the same contents. Both lessons were comparable regarding contents and length, however, in the technology-enhanced adaptive teaching unit, we particularly integrated technology to support adaptive teaching while the control class was taught as usual. Nevertheless, the control condition had also access to digital technologies and had the same equipment and conditions.

Before the teaching units started, students answered the pretest. We asked them to state their age and gender and to rate their interest in the specific subject and their self-efficacy in using technology. Then, the students judged their own knowledge and answered the knowledge test. The pretest lasted about 45 min. After the pretest, the teaching units started and took place for three to four weeks. At the end of the teaching units, the students answered the posttest in which we again assessed students' interest, self-efficacy, judgement of learning, and their comprehension. Additionally, students rated the quality of the teaching unit (i.e., cognitive activation, classroom interruption, classroom monitoring, instructional support). Four weeks after the posttest, students answered the delayed posttest with the same variables to investigate potential lasting learning effects.

5.5. Analyses

We used the statistical software R (version 4.3.3) for our analyses and two-tailed tests with an alpha level of 0.05. To deal with missing values, we applied multiple imputations with 50 imputed datasets and 50 iterations using the mice package, where all variables not intended to serve as predictors (e.g., condition, class, cohort, subject) were set to zero. We then performed cluster-robust estimation of fixed effect models in which we accounted for the correlated error terms within a cluster (students nested within classes nested within cohorts) but independent error terms across clusters (see Cameron & Miller, 2015). The manifest variables were z-standardized, allowing the β estimates to be interpreted similarly to Cohen's d in terms of effect size. To investigate potential differences among conditions regarding students' a) cognitive learning, b) meta-cognitive learning, c) subject-specific interest, and d) perceived self-efficacy in using technology, we used the variable of interest as dependent variable, condition as the independent variable, and students' prior outcomes (e.g., prior knowledge) as control variable.

6. Results

6.1. Preliminary results

There were no significant differences among conditions regarding students' prior knowledge ($\beta = -0.08$, $p = .325$) and their monitoring accuracy before the teaching unit ($\beta = -0.04$, $p = .674$). Students also did not differ among conditions regarding their subject-specific interest ($\beta = -0.03$, $p = .720$) nor their perceived self-efficacy of using technology ($\beta = 0.07$, $p = .406$). Regarding the teaching quality, results demonstrated that the teaching units were comparable across conditions as there were no differences regarding cognitive activation ($\beta = 0.07$, $p = .442$), classroom interruption ($\beta = -0.17$, $p = .058$), classroom monitoring ($\beta = 0.11$, $p = .225$), and instructional support ($\beta = 0.06$, $p = .501$).

As a manipulation check, we additionally examined whether the technology-enhanced adaptive teaching units indeed included more adaptive elements than the control teaching units (i.e., degree of adaptivity). Results revealed that the technology-enhanced teaching units did include more adaptive elements ($M = 19.30$, $SD = 3.79$) than the control units ($M = 16.69$, $SD = 4.56$; $\beta = 0.60$, $p < .001$, medium effect). The descriptive values indicate that all teaching included adaptive elements, but the adaptive teaching unit contained more adaptive elements than the control condition. Thus, the manipulation was successful.

6.2. Did technology-enhanced adaptive teaching contribute to students' learning?

We assumed higher outcomes in the technology-enhanced adaptive teaching units than in the control units regarding students' cognitive, meta-cognitive, and motivational (i.e., interest, self-efficacy) outcomes. Regarding students' cognitive outcomes, students in the technology-enhanced adaptive teaching unit descriptively outperformed students in the control condition in the immediate posttest; this effect, however, was not significant ($\beta = 0.16$, $p = .069$). Interestingly, the effect was significant in the four-week delayed test ($\beta = 0.27$, $p = .003$, small effect). See Table 2 for descriptive values and Appendix C for a correlation table. These effects were robust, even when additionally controlling for knowledge type of the teaching unit (see Table 1). Interestingly, when controlling for answer type (closed versus closed and open answer format), not only students' long-term learning outcomes ($\beta = 0.28$, $p = .003$, small effect) were significantly better in the technology-enhanced adaptive teaching unit but also their immediate learning outcome ($\beta = 0.17$, $p = .042$, small effect).

We performed similar analyses regarding students' monitoring accuracy, their subject-specific interest, and their perceived self-efficacy in using technology (see Table 2 for descriptive values and Appendix C for correlation table). However, results revealed no differences among conditions neither regarding the immediate posttest (monitoring accuracy: $\beta = 0.00$, $p = .958$; interest: $\beta = 0.08$, $p = .257$; self-efficacy: $\beta = 0.09$, $p = .202$) nor the delayed posttest (monitoring accuracy: $\beta = -0.13$, $p = .162$; interest: $\beta = 0.04$, $p = .614$; self- efficacy: $\beta = 0.07$, $p = .436$). Apparently, the technology-enhanced adaptive teaching unit positively affected students' cognitive lasting learning, but not their meta-cognition or their motivational orientations.

6.3. Did boundary conditions moderate the adaptive teaching effect?

We additionally explored potential moderation effects on the individual student-level (i.e., prior knowledge) and on the class-level (i.e., domain, degree of adaptivity). Results indicated no significant moderating effects of students' prior knowledge (see Appendix D for detailed results). Regarding domain as a moderator, results revealed a significant interaction between condition and domain in both the immediate posttest ($\beta = 0.29$, $p = .002$, small effect) and the delayed posttest ($\beta = 0.30$, $p = .001$, small effect). Additional simple comparisons revealed that the adaptive teaching units were more effective than the control units in languages (immediate posttest: $\beta = 0.41$, $p = .001$, small effect). This effect was even more pronounced in the delayed posttest ($\beta = 0.58$, $p < .001$, medium effect). However, there were no differences among conditions in STEM subjects and humanities ($p \geq .301$).

Similarly, results revealed a significant interaction effect between condition and the degree of adaptivity regarding the immediate posttest ($\beta = 0.24$, $p = .009$, small effect) and the delayed posttest ($\beta = 0.42$, $p < .001$, small effect). To explore this effect in more detail, we calculated the mean score of the degree of adaptivity of both conditions ($M = 18.06$) as cut-off criteria. Thus, teaching units that contained less than or equal 18 adaptive elements were regarded as having a low degree of adaptivity, whereas teaching units with more than 18 adaptive elements were regarded as having a high degree of adaptivity. Additional analyses with these subsamples revealed that teaching units with low degree of adaptivity resulted in no differences among conditions regarding students learning (immediate posttest: $\beta = -0.21$, $p = .120$; delayed posttest: $\beta = -0.15$, $p = .286$), but teaching units with high degree of adaptivity demonstrated a significant benefit regarding students' learning compared to the control condition (immediate posttest: $\beta = 0.37$, $p = .004$, small effect; delayed posttest: $\beta = 0.44$, $p = .001$, small effect). Interestingly, high degree of adaptivity not only boosted students' learning but also their self-efficacy in using technology: Students in highly adaptive technology-enhanced teaching units showed higher self-efficacy compared to students in the control condition in the immediate posttest ($\beta = 0.34$, $p = .014$, small effect). There were no other significant interaction effects (see Appendix D-F for detailed results).

7. Discussion

The present study aimed to investigate the effectiveness of technology-enhanced adaptive teaching units across diverse subjects compared to traditional teaching methods. Our results provide nuanced insights into the effectiveness of technology-enhanced adaptive teaching

Table 2
Descriptive values of the main variables.

Variable	Pretest		Immediate Posttest		Delayed Posttest	
	CG	EG	CG	EG	CG	EG
Learning Outcome (0–100 %)	33.25 (17.08)	31.75 (18.61)	47.06 (22.28)	50.64 (23.88)	45.06 (19.42)	50.66 (22.82)
Monitoring Accuracy (-100-100 %)	24.81 (25.61)	23.87 (25.96)	5.03 (28.06)	4.88 (27.09)	7.47 (35.20)	3.28 (25.24)
Subject-Specific Interest (1–4)	2.61 (0.65)	2.59 (0.59)	2.57 (0.76)	2.65 (0.77)	2.53 (0.72)	2.55 (0.74)
Self-Efficacy in Using Technology (1–4)	3.22 (0.68)	3.26 (0.66)	3.25 (0.77)	3.35 (0.70)	3.15 (0.70)	3.23 (0.65)

Note. The values are based on row data. CG = Control Group (regular teaching unit). EG = Experimental Group (technology-enhanced adaptive teaching units).

with a focus on cognitive, meta-cognitive, and motivational outcomes, and the moderating effects of prior knowledge, subject domain, and the degree of adaptivity.

Consistent with previous literature (Corno, 2008; Sibley et al., 2024), our results indicate that technology-enhanced adaptive teaching positively influenced students' cognitive learning outcomes. Even though the immediate posttest results showed a descriptive but non-significant advantage for the adaptive teaching condition, the delayed posttest revealed a significant benefit for the adaptive teaching condition. This suggests that adaptive teaching contributes to better long-term knowledge, which aligns with the theoretical framework of adaptive teaching by promoting meaningful and lasting learning through formative assessment, differentiation, and individual support (Corno, 2008; Roelle et al., 2022, 2023; Sibley et al., 2024). Notably, our study also contributes to the existing literature by demonstrating lasting learning effects even four weeks after the intervention, indicating the potential for adaptive teaching to support durable knowledge retention among students. In conclusion, our findings partially support the hypothesis that technology-enhanced adaptive teaching would lead to superior cognitive outcomes and underscores the potential of adaptive teaching strategies to cater to individual student needs effectively across various educational domains.

To test the robustness of our findings, we additionally controlled for knowledge type and answer format. The results remained robust when controlling for knowledge type. Interestingly, when controlling for answer format, students did not only benefit more from the technology-enhanced adaptive teaching units regarding their long-term learning but also regarding their immediate learning. This suggests that answer format significantly influences the effect of technology-enhanced adaptive teaching. Tests with both closed and open answer formats showed stronger results, as students performed better compared to tests with only closed answer formats. A possible explanation is that open formats require active retrieval (Obergassel et al., 2025; Roelle et al., 2023), where students must independently recall and organize information, rather than simply selecting an option. Future studies should investigate these effects further, specifically by examining outcomes with and without open recall items, to better understand the role of adaptive teaching in supporting higher-order cognitive skills.

Contrary to our expectations, the study did not find significant differences between the experimental and control groups in meta-cognitive outcomes (monitoring accuracy) and motivational outcomes (interest and self-efficacy in using technology) in both the immediate and delayed posttests. These findings suggest that while technology-enhanced adaptive teaching enhances cognitive learning over time, it does not necessarily impact on students' meta-cognitive monitoring accuracy or their motivational perceptions. This contradicts prior research that demonstrated that, for instance, digital response systems encourage self-reflection, making students more aware of their learning processes and helping them develop better monitoring accuracy (Molin et al., 2022; Shin et al., 2022). One possible explanation for the non-significant results may relate to measurement of students' interest (e.g., "Mathematics captures my attention") which may have been interpreted by students as an indicator of general, enduring interest in mathematics rather than situational interest. This potential ambiguity could mean that we were inadvertently measuring a more stable trait-like interest, which might not be as easily influenced by adaptive teaching within the study's timeframe. This highlights the need for further research with items explicitly designed to capture situational interest, to better understand how adaptive teaching may affect this dimension of motivation. This outcome highlights the complexity of adaptive teaching's impact and indicates that further research is needed to explore the mechanisms affecting different dimensions of learning.

The examination of boundary conditions provided important insights into the factors that influence the effectiveness of technology-enhanced adaptive teaching. Prior knowledge did not significantly moderate the effects of adaptive teaching, indicating the robustness of adaptive teaching benefits across varying levels of initial student knowledge. Interestingly, we found that the level of adaptive elements implemented within teaching units significantly moderated learning outcomes. Units that incorporated higher levels of adaptivity, including frequent formative assessments and tailored instructional supports, yielded greater learning gains and higher subject-related self-efficacy in using technology. This aligns with previous research highlighting the importance of fidelity in implementing adaptive teaching strategies (Sibley et al., 2024). Moreover, while adaptive teaching was effective across different subjects, our findings suggest variability in its impact depending on the educational domain. Specifically, adaptive teaching demonstrated stronger effects in language subjects compared to STEM or humanities, underscoring the need for subject-specific adaptations in educational practices. With our current data, we are unable to determine why this effect occurred. However, it highlights an important aspect that warrants further investigation in future research to better understand the underlying factors contributing to these differences.

7.1. Educational implications

The findings from our study have several implications for educational practice. A major contribution of our study is the demonstration of the generalizability of technology-enhanced adaptive teaching across different subjects and contexts. While previous research has often been limited to subject-specific applications, we followed the LoGeT framework to generate localized, generalizable, and transferable evidence (see Lachner et al., 2024; Sibley et al., 2024). Our findings demonstrated that adaptive teaching with technology can be effectively implemented in a wide range of subjects still yielding generalizable long-term benefits. This broad applicability is crucial for educational practice, as it suggests that adaptive teaching is not confined to specific disciplines but can be a universal approach to enhancing student learning.

Integrating educational technologies into adaptive teaching frameworks offers educators a powerful tool to personalize learning experiences and address the diverse needs of students effectively. By leveraging formative assessments and tailored instructional supports, educators can foster a more supportive and engaging learning environment, potentially increasing students' cognitive learning and their self-efficacy in using technology. In this vein, teachers need to carefully apply adaptive elements constantly, since our results demonstrated that adaptive teaching units were only more effective than regular teaching when the degree of adaptivity was high. Additionally, adaptive teaching also seems to be very effective for teaching languages. In this light, new adaptive concepts are needed that also support adaptive teaching in STEM and humanities.

Moreover, our study highlights the importance of professional development programs for educators to effectively implement adaptive teaching practices. The close collaboration between the research team and teachers was essential for ensuring that digital tools were effectively used to support adaptive instruction across four levels: formative assessments, macro- and micro-level adaptations, and consolidation phases. However, this intensive support was time-consuming for both researchers and teachers, raising questions about the long-term feasibility of such an approach in everyday teaching practice. To scale this approach sustainably, the competencies required for using digital tools as a means to implement adaptive teaching strategies should be systematically integrated into teacher education and professional development programs. Future training initiatives should focus on equipping teachers with the skills to independently design and implement adaptive teaching elements with digital media, reducing the need for external guidance while maintaining the effectiveness of the approach. By embedding adaptive teaching and technological competencies into teacher preparation and ongoing training, educators can be better supported in leveraging digital technologies to create adaptive learning environments at scale.

7.2. Limitations

Despite the strengths of our study, several limitations warrant consideration. First, the quasi-experimental design limited our ability to establish causal relationships definitively. However, applying research in real classroom settings is challenging but essential to conduct educational relevant research. Future research employing randomized controlled trials could provide stronger evidence of the causal impact of adaptive teaching on learning outcomes. Second, while our study included diverse subjects and student populations, future research could further explore the generalizability of adaptive teaching across different educational contexts and student demographics. Investigating how adaptive teaching impacts students with varying learning profiles and educational needs could provide insights into tailoring adaptive strategies more effectively. Third, our findings showed that the degree of adaptivity was a significant moderator, with only teaching units that had high degree of adaptivity (i.e., a higher number of adaptive elements) leading to better student performance compared to traditional teaching. Thus, the frequency of adaptive elements was critical for the success of the teaching units. However, it is important to note that we measured the degree of adaptivity by counting the number of adaptive elements within a unit, rather than considering dosage (e.g., the duration of each adaptive element) which could have influenced the results differently. Future research should investigate how both the number and the duration of adaptive elements impact learning outcomes.

Fourth, we did not specifically measure how frequently or for how long digital technologies were used in the control and adaptive classes. Therefore, we cannot definitively determine whether the observed effects were due to the adaptive teaching approach, the use of technology, or a combination of both. Future research should explore this interplay further, to distinguish between the effects of adaptive strategies and technology use.

Fifth, a significant amount of missing data was present at different stages of the study, with 24 % missing at the pretest, 35 % at the posttest, and 57 % at the delayed test. While missing data is common in longitudinal field studies, especially in real-world classroom settings, the high proportion of missing data could have affected the robustness of our results. Although we addressed the missing data using multiple imputations, caution is still advised in interpreting the findings. Future studies should seek to reduce missing data, possibly by employing strategies that improve student retention over time. We also did not apply corrections for multiple testing, which should be taken into account when interpreting the results. Future research may consider using statistical adjustments to control the increased risk of Type I errors associated with conducting multiple comparisons.

Sixth, one unexpected finding in our study was the weak correlation between pretest and posttest cognitive performance measures. Typically, stronger associations between measurements at different time points are expected. Several potential explanations may account for this pattern. Methodologically, we used parallel test versions at different measurement points, ensuring that test items remained consistent in format (open-ended or closed questions). These tests were developed by subject matter experts to ensure content validity. However, despite these efforts, it is possible that certain parallel versions (e.g., posttest) were not perfectly equivalent in difficulty or construct coverage, which may have influenced the observed correlations. Small variations in item complexity or response patterns between versions could have introduced inconsistencies in how cognitive performance was captured across time points. Additionally, cognitive learning processes during the intervention might have led to rapid changes, making prior knowledge less predictive of posttest performance. Motivational aspects or differences in test-taking engagement between measurement points could also have played a role. Future studies should examine these factors in more detail, for example, by systematically validating the equivalence of parallel test versions through pretesting, analyzing item difficulties in more depth, or investigating test motivation effects.

An additional limitation of our study is the confounding of subject and grade level within the teaching units. Because certain subjects are typically taught at specific grade levels, it was challenging to disentangle the effects of subject from those of grade level in our analysis. This confounding could impact the interpretation of our results, as differences observed may be influenced by either subject-specific characteristics or developmental factors related to age. Future studies could aim to control these variables more systematically, either by selecting a narrower grade range or by including more diverse subjects within each grade level. Finally, examining the long-term effects of adaptive teaching beyond four weeks could offer deeper insights into its sustainability and impact on students' educational trajectories. Longitudinal studies tracking students' progress over extended periods would contribute to understanding the enduring benefits of adaptive teaching approaches.

7.3. Conclusion

To our knowledge, we were the first who systematically examined technology-enhanced adaptive teaching in real classroom settings across domains to generalize effects of adaptive teaching with technology. Our results present valuable insights into the effectiveness of technology-enhanced adaptive teaching to enhance students' cognitive learning outcomes across diverse educational settings. By demonstrating significant improvements in lasting learning, particularly in language subjects, our findings highlight the transformative potential of adaptive teaching strategies in contemporary educational contexts. Moving forward, continued research and innovation in adaptive teaching practices are crucial to advancing educational equity across subjects and fostering meaningful learning experiences for all students.

Declaration of interest

We authors have no declaration of interest.

CRediT authorship contribution statement

Leonie Sibley: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Armin Fabian:** Writing – review & editing, Resources, Methodology. **Christine Plicht:** Resources, Project administration, Methodology, Data curation. **Lisa Pagano:** Resources, Investigation, Data curation. **Niklas Ehrhardt:** Resources, Project administration, Data curation. **Luisa Wellert:** Resources. **Thorsten Bohl:** Funding acquisition. **Andreas Lachner:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Acknowledgements & funding sources

We would like to thank all participating schools, teachers, students, and the advisory board of the project for a successful collaboration. We additionally would like to thank Lisa Pagano and Jule Walter for their assistance in collecting data and rating qualitative data during the entire duration of the project. The research project was supported by the Robert Bosch Stiftung under the contract number 00906529-001 and the Vector Stiftung under the contract number P2018-0089 from 2022 to 2024.

Appendix A. Test and coding for the teaching unit mathematics 1 – Functions (Test A)

1) Determine the first derivatives of the following functions:
   a) $f(x) = 3 \cdot e^x$
   b) $f(x) = x \cdot e^{-3x}$
2) Which of the following properties is characteristic of the function $f(x) = e^x$, where e is Euler's constant $e = 2718 \ldots$? (multiple answers possible)
   a) It has no zeros.
   b) For the derivative function, the following holds: $f'(x) = f(x)$.
   c) It has two intercepts with the Y-axis.
   d) It is strictly monotonically decreasing for all real numbers.
   e) I do not know.
3) Solve the following equations:
   a) $2^x = 16$
   b) $e^{2x} - 5 = 10$
4) The following applies to the natural logarithm *ln*. Which of the following statements about the natural logarithm are true? (multiple answers possible)
   a) The number $x = \ln(b)$ is the solution of the exponential equation $e^x = b$.
   b) The following applies: $\ln(e) = 1$.
   c) The following applies: $\ln(0) = e$.
   d) The following applies: $\ln(1) = e$.
   e) I do not know.
5) What is true for the function $f(x) = x^2 \cdot e^{-x}$? (Multiple answers possible).
   a) $f(x) \to 0$ for $x \to +\infty$
   b) $f(x) \to +\infty$ for $x \to -\infty$
   c) $f(x) \to -\infty$ for $x \to +\infty$
   d) $f(x) \to 0$ for $x \to -\infty$
   e) I do not know.
6) Describe the effect of the parameter a on the graph of the function.
   a) $f(x) = e^x + a$
   b) $f(x) = e^{x-a}$

Note. Two points can be scored for each task, so that the maximum score is 12 points.

Appendix B. Overview of teaching unit mathematics 2 – Euclidean plane

Lesson	Topic	Adaptive Methods	Technology	Duration
1	The Cross Product – Transforming Plane Equations	• FA: Diagnostic assessment at the beginning to determine students' knowledge levels • Ma-A: Cooperative work based on differentiated performance levels • Mi-A: Practice with direct feedback	Kahoot, TaskCards, Learning Path (on school platform)	135 min
2	Visualizing Planes	• Self-paced learning • Peer assessment for feedback • FA: Formative assessment	Explanatory video	90–135 min
3	Relative Position of Planes and Lines	• FA: Diagnostic assessment at the beginning • Ma-A: Pair work based on differentiated performance levels • Mi-A: Self-study with dynamic geometry representation and direct feedback; possible diagnostics using Classroom	Kahoot, GeoGebra	90–135 min
4	Relative Position of Planes	• FA: Diagnostic assessment at the beginning • Self-paced learning • Mi-A: Individual support and direct feedback during exercises • FA: Diagnostic assessment at the end	Quizzes, StudySmarter	135–180 min
5	Final Exercise	• Ma-A: Randomly generated practice problems with individually adjusted number of practice problems • FA: aDiagnostic assessment at the end		45–90 min

Note. FA = formative assessment, Ma-A: macro adaptation; Mi-A: micro adaptation.

Appendix C. Correlations between dependent variables

		Pretest				Immediate Posttest				Delayed Posttest			
		CL	MA	Int	SE	CL	MA	Int	SE	CL	MA	Int	SE
Pretest	CL	–	**−0.53**	−0.01	−0.05	0.06	0.05	0.09	−0.06	**0.18**	−0.02	0.00	−0.04
	MA	**−0.53**	–	**0.14**	**0.12**	0.01	**0.26**	0.03	0.07	−0.02	**0.14**	**0.15**	**0.18**
	Int	−0.01	**0.14**	–	**0.10**	**0.22**	0.02	**0.60**	0.04	0.10	0.01	**0.54**	**0.17**
	SE	−0.05	**0.12**	**0.10**	–	**0.13**	0.04	**0.15**	**0.58**	**0.26**	0.03	**0.14**	**0.49**
Immediate Posttest	CL	0.06	0.01	**0.22**	**0.13**	–	**−0.58**	**0.25**	**0.12**	**0.64**	−0.12	**0.29**	**0.19**
	MA	0.05	**0.26**	0.02	0.04	**−0.58**	–	−0.03	0.01	**−0.29**	0.27	0.00	0.05
	Int	0.09	0.03	**0.60**	**0.15**	**0.25**	−0.03	–	**0.16**	**0.16**	0.05	**0.52**	**0.24**
	SE	−0.06	0.07	0.04	**0.58**	**0.12**	0.01	**0.16**	–	**0.15**	−0.06	**0.15**	**0.66**
Delayed Posttest	CL	**0.18**	−0.02	0.10	**0.26**	**0.64**	**−0.29**	**0.16**	**0.15**	–	**−0.28**	**0.21**	**0.24**
	MA	−0.02	**0.14**	0.01	0.03	**−0.12**	**0.27**	0.05	−0.06	**−0.28**	–	0.06	−0.08
	Int	0.00	**0.15**	**0.54**	**0.14**	**0.29**	0.00	**0.52**	**0.15**	**0.21**	0.06	–	**0.33**
	SE	−0.04	**0.18**	**0.17**	**0.49**	**0.19**	0.05	**0.24**	**0.66**	**0.24**	−0.08	**0.33**	–

Note. Significant correlations are highlighted in bold. Correlations are based on row data. CL = Cognitive Learning Outcome. MA = Monitoring Accuracy. Int = Subject-Specific Interest. SE = Self-Efficacy in Using Technology.

Appendix D. Moderation analyses with prior knowledge as moderator

	Immediate Posttest			Delayed Posttest		
	β	*SD*	*p*	β	*SD*	*p*
Comprehension						
Condition	0.16	0.09	0.068	**0.27**	**0.09**	**0.003**
Prior knowledge	0.10	0.07	0.181	0.23	0.07	0.001
Condition × Prior knowledge	−0.06	0.10	0.557	−0.08	0.09	0.401
Monitoring accuracy						
Condition	0.03	0.08	0.706	−0.12	0.09	0.185
Prior knowledge	**0.27**	**0.08**	**0.001**	0.04	0.08	0.589
Prior monitoring accuracy	**0.40**	**0.06**	**< 0.001**	**0.18**	**0.07**	**0.015**
Condition × Prior knowledge	0.00	0.09	0.993	0.05	0.09	0.574
Subject-specific interest						
Condition	0.09	0.07	0.218	0.04	0.09	0.610
Prior knowledge	0.08	0.05	0.171	0.03	0.07	0.610
Prior subject-specific interest	**0.60**	**0.04**	**< 0.001**	**0.54**	**0.05**	**< 0.001**
Condition × Prior knowledge	0.03	0.07	0.681	−0.05	0.08	0.525
Self-efficacy						
Condition	0.09	0.07	0.216	0.07	0.08	0.445
Prior knowledge	−0.04	0.06	0.532	−0.03	0.06	0.641
Prior self-efficacy	**0.58**	**0.05**	**< 0.001**	**0.49**	**0.06**	**< 0.001**
Condition × Prior knowledge	0.01	0.07	0.886	0.03	0.08	0.729

Note. Condition is coded as CG = 1 and EG = 1.

Appendix E. Moderation analyses with domain as moderator

	Immediate Posttest			Delayed Posttest		
	β	*SD*	*p*	β	*SD*	*p*
Comprehension						
Condition	0.16	0.09	0.067	**0.29**	**0.09**	**0.001**
Domain	−0.13	0.07	0.057	0.02	0.08	0.795
Prior knowledge	0.07	0.05	0.147	**0.18**	**0.05**	**< 0.001**
Condition × Domain	**0.29**	**0.09**	**0.002**	**0.30**	**0.09**	**0.001**
Monitoring accuracy						
Condition	0.01	0.09	0.918	−0.14	0.10	0.133
Domain	0.08	0.07	0.235	−0.12	0.10	0.210
Prior monitoring accuracy	**0.26**	**0.05**	**< 0.001**	**0.16**	**0.06**	**0.011**
Condition × Domain	−0.08	0.09	0.387	0.03	0.10	0.782
Subject-specific interest						
Condition	0.09	0.07	0.210	0.05	0.08	0.556
Domain	0.05	0.06	0.411	0.04	0.08	0.629
Prior subject-specific interest	**0.60**	**0.04**	**< 0.001**	**0.54**	**0.05**	**< 0.001**
Condition × Domain	0.06	0.08	0.407	0.05	0.08	0.561
Self-efficacy						
Condition	0.09	0.07	0.208	0.07	0.09	0.407
Domain	−0.03	0.05	0.556	0.02	0.08	0.809
Prior self-efficacy	**0.58**	**0.05**	**< 0.001**	**0.50**	**0.06**	**< 0.001**
Condition × Domain	0.03	0.07	0.624	0.02	0.09	0.784

Note. Condition is coded as CG = 1 and EG = 1.

Appendix F. Moderation analyses with the degree of adaptivity as moderator

	Immediate Posttest			Delayed Posttest		
	β	*SD*	*p*	β	*SD*	*p*
Comprehension						
Condition	0.01	0.09	0.870	0.12	0.09	0.171
Degree of adaptivity	**0.13**	**0.06**	**0.025**	0.05	0.07	0.445
Prior knowledge	0.07	0.05	0.125	**0.19**	**0.05**	**< 0.001**
Condition × Degree of adaptivity	**0.24**	**0.09**	**0.009**	**0.42**	**0.09**	**< 0.001**
Monitoring accuracy						
Condition	0.03	0.09	0.718	−0.13	0.09	0.175
Degree of adaptivity	−0.06	0.07	0.350	0.02	0.08	0.791
Prior monitoring accuracy	**0.26**	**0.05**	**< 0.001**	**0.14**	**0.06**	**0.020**
Condition × Degree of adaptivity	0.04	0.09	0.687	−0.05	0.10	0.602
Subject-specific interest						
Condition	0.06	0.08	0.463	−0.03	0.08	0.741
Degree of adaptivity	0.01	0.06	0.895	0.06	0.07	0.354
Prior subject-specific interest	**0.60**	**0.04**	**< 0.001**	**0.53**	**0.05**	**< 0.001**
Condition × Degree of adaptivity	0.08	0.09	0.378	0.12	0.09	0.178
Self-efficacy						
Condition	0.09	0.08	0.270	0.05	0.09	0.612
Degree of adaptivity	−0.13	0.07	0.058	−0.04	0.07	0.593
Prior self-efficacy	**0.56**	**0.05**	**< 0.001**	**0.48**	**0.06**	**< 0.001**
Condition × Degree of adaptivity	**0.29**	**0.10**	**0.002**	0.16	0.11	0.141

Note. Condition is coded as CG = 1 and EG = 1.

Data availability

All data used in the analyses can be accessed at OSF: https://osf.io/vxcru/?view_only=e93b65d2a3064980b0830e68d6932c82.

References

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26* (2), 147–179. https://doi.org/10.1207/s15516709cog2602_1

Alrawashdeh, G. S., Fyffe, S., Azevedo, R. F. L., & Castillo, N. M. (2024). Exploring the impact of personalized and adaptive learning technologies on reading literacy: A global meta-analysis. *Educational Research Review, 42*, Article 100587. https://doi.org/10.1016/j.edurev.2023.100587.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167–207. https://doi.org/10.1207/s15327809jls0402_2

Baars, M., van Gog, T., de Bruin, A., & Paas, F. (2017). Effects of problem solving after worked example study on secondary school children's monitoring accuracy. *Educational Psychology, 37*(7). https://doi.org/10.1080/01443410.2016.1150419

Backfisch, I., Sibley, L., Lachner, A., Kirchner, K. T., Hische, C., & Scheiter, K. (2024). Enhancing pre-service teachers' technological pedagogical content knowledge (TPACK): Utility-value interventions support knowledge integration. *Teaching and Teacher Education, 142*, Article 104532. https://doi.org/10.1016/j.tate.2024.104532

Bernacki, M., Greene, M. J., & Lobczowski, N. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review, 33*, 1675–1715. https://doi.org/10.1007/s10648-021-09615-8.

Bloom, B., Krathwohl, D. R., & Masia, B. B. (1984). *Taxonomy of educational objectives: The classification of educational goals*. Longman.

Brod, G. (2024). There are multiple paths to personalized education, and they should Be combined. *Current Directions in Psychological Science, 33*(3). https://doi.org/10.1177/09637214241242459

Brühwiler, C., & Blatchford, P. (2011). Effects of class size and adaptive teaching competency on classroom processes and academic outcome. *Learning and Instruction, 21*(1), 95–108. https://doi.org/10.1016/j.learninstruc.2009.11.004

Brusilovsky, P., Schwarz, E., & Weber, G. (1996). ELM-ART: An intelligent tutoring system on world wide web. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *Intelligent tutoring systems: Third international conference* (Vol. 1086, pp. 261–269). https://doi.org/10.1007/3-540-61327-7_123

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources, 50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471–533. https://doi.org/10.1016/S0364-0213(01)00044-1

Corno, L. Y. N. (2008). On teaching adaptively. *Educational Psychologist, 43*(3), 161–173. https://doi.org/10.1080/00461520802178466

Enchikova, E., Neves, T., Toleda, G., & Nata, G. (2024). Change in socioeconomic education equity after 20 years of PISA: A systematic literature review. *International Journal of Educational Research Open, 7*(100359), 1–14. https://doi.org/10.1016/j.ijedro.2024.100359

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. https://doi.org/10.1016/j.learninstruc.2013.07.001

Fyfe, E. R., Leeuw, J. R. de, Carvalho, P. F., Goldstone, R. L., Sherman, J., Admiraal, D., Alford, L. K., Bonner, A., Brassil, C. E., Brooks, C. A., Carbonetto, T., Chang, S. H., Cruz, L., Czymoniewicz-Klippel, M., Daniel, F., Driessen, M., Habashy, N., Hanson-Bradley, C. L., Hirt, E. R., … Motz, B. A. (2021). ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science, 4*(3). https://doi.org/10.1177/25152459211027575

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology, 108*(1), 82–92. https://doi.org/10.1037/edu0000053

Graesser, A. C., Shulan, L., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 180–192.

Gross, A. L., Inouye, S. K., Rebok, G. W., Brandt, J., Crane, P. K., Parisi, J. M., Tommet, D., Bandeen-Roche, K., Carlson, M. C., & Jones, R. N. (2013). Parallel but not equivalent: Challenges and solutions for repeated assessment of cognition over time. *Journal of Clinical and Experimental Neuropsychology, 34*(7), 758–772. https://doi.org/10.1080/13803395.2012.681628

Hammer, M., Göllner, R., Scheiter, K., Fauth, B., & Stürmer, K. (2021). For whom do tablets make a difference? Examining student profiles and perceptions of instruction with tablets. *Computers & Education, 166*, Article 104147. https://doi.org/10.1016/j.compedu.2021.104147

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education, 153*, Article 103897. https://doi.org/10.1016/j.compedu.2020.103897

Hooshyar, D., Weng, X., Sillat, P. J., Tammets, K., Wang, M., & Hämäläinen, R. (2024). The effectiveness of personalized technology-enhanced learning in higher education: A meta-analysis with association rule mining. *Computers & Education, 223*, Article 105169. https://doi.org/10.1016/j.compedu.2024.105169.

Jacob, L., Lachner, A., & Scheiter, K. (2022). Do school students' academic self-concept and prior knowledge constrain the effectiveness of generating technology-mediated explanations? *Computers & Education, 182*, Article 104469. https://doi.org/10.1016/j.compedu.2022.104469

Jonassen, D. H. (2005). Tools for representing problems and the knowledge required to solve them. In S.-O. Tergan, & T. Keller (Eds.), *Knowledge and information visualization: Searching for synergies* (pp. 82–94). Springer.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Karst, K., Bonefeld, M., Dotzel, S., Fehringer, B. C., & Steinwascher, M. (2022). Data-based differentiated instruction: The impact of standardized assessment and aligned teaching material on students' reading comprehension. *Learning and Instruction, 79*, Article 101597. https://doi.org/10.1016/j.learninstruc.2022.101597

Keuning, T., & van Geel, M. (2021). Differentiated teaching with adaptive learning systems and teacher dashboards: The teacher still matters most. *IEEE Transactions on Learning Technologies, 14*(2), 201–210. https://doi.org/10.1109/TLT.2021.3072143

Knogler, M., Harckiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology, 43*, 39–59.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems. *Review of Educational Research, 86*(1), 42–78. https://doi.org/10.3102/0034654315581420

Lachner, A., Sibley, L., & Wagner, S. (2024). Practice recommendations or not? The LoGeT-model as empirical approach to generate localized, generalized, and transferable evidence. *Educational Psychology Review, 36*(36). https://doi.org/10.1007/s10648-024-09876-z

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology, 106*(4), 901–918. https://doi.org/10.1037/a0037123

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology, 114*(8), 1743–1772. https://doi.org/10.1037/edu0000764

Meurers, D., de Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics, 39*, 161–188. https://doi.org/10.1017/S0267190519000126

Molenaar, I., & Knoop-van Campen, C. (2019). How teachers make dashboard information actionable. *IEEE Transactions on Learning Technologies, 12*(3), 347–355. https://doi.org/10.1109/TLT.2018.2851585

Molin, F., de Bruin, A., & Haelermans, C. (2022). A conceptual framework to understand learning through formative assessments with student response systems: The role of prompts and diagnostic cues. *Social Sciences & Humanities Open, 6*, Article 100323. https://doi.org/10.1016/j.ssaho.2022.100323

Nievelstein, F., Van Gog, T., Van Dijck, G., & Boshuizen, H. P. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*(2), 118–125. https://doi.org/10.1016/j.cedpsych.2012.12.004

Obergassel, N., Heitmann, S., Grund, A., Fries, S., Berthold, K., & Roelle, J. (2025). Adaptation of quizzing in learning psychology concepts. *Learning and Instruction, 95*. https://doi.org/10.1016/j.learninstruc.2024.102028. Article 102028.

OECD. (2023). *PISA 2022 Results (Volume 1). The State of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist, 28*(1), 15–22. https://doi.org/10.1207/S15326985EP3801_3

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19–36. https://doi.org/10.1016/j.edurev.2015.12.003

Rigotti, T., Schyns, B., & Mohr, G. (2008). A short version of the occupational self-efficacy scale: Structural and construct validity across five countries. *Journal of Career Assessment, 16*(2), 238–255. https://doi.org/10.1177/1069072707305763

Roelle, J., Endres, T., Abel, R., Obergassel, N., Nückles, M., & Renkl, A. (2023). Happy togethers? On the relationship between research on retrieval practice and generative learning using the case of follow-up learning tasks. *Educational Psychology Review, 35*(102). https://doi.org/10.1007/s10648-023-09810-9

Roelle, J., Schweppe, J., Endres, T., Lachner, A., von Aufschnaiter, C., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (2022). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 54*(4), 142–150. https://doi.org/10.1026/0049-8637/a000261

Roschelle, J., Penuel, W., & Shechtman, N. (2006). Co-design of innovations with teachers: Definition and dynamics. In S. A. Barab, K. E. Hay, & D. T. Hickey (Eds.), *Icls '06: Proceedings of the 7th international conference on learning sciences* (pp. 606–612). International Society of the Learning Sciences.

Schmid, R., Pauli, C., Stebler, R., Reusser, K., & Petko, D. (2022). Implementation of technology-supported personalized learning: Its impact on instructional quality. *The Journal of Educational Research, 115*(3), 187–198. https://doi.org/10.1080/00220671.2022.2089086

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition Learning, 4*(33). https://doi.org/10.1007/s11409-008-9031-3

Shin, J., Chen, F., Lu, C., & Bulut, O. (2022). Analyzing students' performance in computerized formative assessments to optimize teachers' test administration. *Journal of Computers in Education, 9*. https://doi.org/10.1007/s40692-021-00196-7

Sibley, L., Lachner, A., Plicht, C., Fabian, A., Backfisch, I., Scheiter, K., & Bohl, T. (2024). Feasibility of adaptive teaching with technology: Which implementation conditions matter? *Computers & Education, 219*, Article 105108. https://doi.org/10.1016/j.compedu.2024.105108

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106*(2), 331–347. https://doi.org/10.1037/a0034752

Tetzlaff, L., Schmiedek, F., & Brod, G. (2021). Developing personalized education: A dynamic framework. *Educational Psychology Review, 33*(3), 863–882. https://doi.org/10.1007/s10648-020-09570-w

van Leeuwen, A., Rummel, N., & van Gog, T. (2019). What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *International Journal of Computer-Supported Collaborative Learning, 14*, 261–289. https://doi.org/10.1007/s11412-019-09299-x

VanLehn, K. A. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring Systems. *Educational Psychologist, 46*(4), 197–221. https://doi.org/10.1080/00461520.2011.611369

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Harvard University Press.

Wagner, S., Sibley, L., Weiler, D., Burde, J.-P., & Scheiter, K. (2024). The more, the better? Learning with feedback and instruction. *Learning and Instruction, 89*, Article 101822. https://doi.org/10.1016/j.learninstruc.2023.101844

Wittwer, J., Nückles, M., Landmann, N., & Renkl, A. (2010). Can tutors be supported in giving effective explanations? *Journal of Educational Psychology, 102*(1), 74–89. https://doi.org/10.1037/a0016727

Zheng, L., Niu, J., Zhong, L., & Gyasi, J. F. (2021). The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis. *Interactive Learning Environments, 31*(9), 5650–5664. https://doi.org/10.1080/10494820.2021.2015693.