A Longitudinal Study on Boosting Students' Performance with a Learning Companion

Conference Paper · December 2023

Ricarda Schlimbach
Hochschule Heilbronn

Tim Lange
RWTH Aachen University

Susanne Robra-Bissantz
Technische Universität Braunschweig

A Longitudinal Study on Boosting Students' Performance with a Learning Companion

Completed Research Paper

Ricarda Schlimbach
Hochschule Heilbronn
ricarda.schlimbach@hs-heilbronn.de

Tim C. Lange
TU Braunschweig
tim.lange@tu-bs.de

Susanne Robra-Bissantz
TU Braunschweig
s.robra-bissantz@tu-bs.de

Abstract

This study examines the impact of a coded virtual learning companion (LC) that interacts with students of an introductory information systems class throughout the semester. The LC is designed to motivate, advise on time management strategies, and study collaboratively. We conducted a between-subject longitudinal field experiment to investigate the LC's impact on student motivation, time management, and learning outcomes. Statistical analysis, including a PLS-SEM model, shows that the LC significantly (p < 0.05) improves extrinsic motivation, challenge, short-term planning, and time attitudes. A multiple mediator analysis confirms the role of motivation and time management as mediators between LC use and learning outcomes (subjective knowledge and exam scores). In addition, we conducted a qualitative workshop with the target group to identify barriers to LC adoption and derive mitigation strategies. Overall, our study reveals great potential to facilitate learning with LCs in higher education.

Keywords: Learning Companion, Conversational Agent, Longitudinal Study, Education

Introduction

Over the past years, the trend towards competence-oriented, interactive, and digital teaching for students has been accelerated by technological advancements and the forced transition to online courses during the global pandemic (Ngwacho 2023). The enhancement of remote learning since then has even intensified the perceived workload, uncertainties in the course of study as well as students' demand for assistance in coping with their learning material in line with their scarce time resources (Abdullah et al. 2022; Herrmann-Werner et al. 2021). Especially in distance education or mass lecturers, students complain to be left alone with the challenge of continuously motivating themselves and managing their time appropriately (Rinn et al. 2022) caused by lacking social support and interactivity (Dalipi et al. 2018; Herrmann-Werner et al. 2021). Consequently, these limiting factors often result in exhaustion, mental health issues, and high drop-out rates (Behr et al. 2021). A promising approach to counteract the mentioned challenges is the use of naturally interacting dialogue systems to accompany students individually in their learning process (Schlimbach et al. 2022). These agents simulate human language and adapt to human behavior (Diederich et al. 2022). Due to technological advancements in artificial intelligence (AI) and natural language processing (NLP) as well as increasing user acceptance, they have gained enormous popularity in both research and practice in recent years (Adamopoulou and Moussiades 2020; Diederich et al. 2022). Particularly in their manifestation as virtual, bonding *Learning Companions* (LCs) that go beyond pure assistance (Khosrawi-Rad, Rinn, et al. 2022; Strohmann et al. 2022), relationship-oriented chatbots have gained momentum. They might help students improve their time management collaboratively, form sustainable learning habits and motivate them to study (Luxton 2014; Schlimbach et al. 2023).

Furthermore, the release of ChatGPT in November 2022 has introduced a revolutionary conversational AI and has initiated a new era of deep learning (Sun 2022). This progress is of great significance as ChatGPT can engage in human-like conversations and is widely relied on as a learning assistant by students worldwide (Tlili et al. 2023). LCs have in common with disruptive chatbots like ChatGPT to be constantly available, easily scalable, and highly knowledgeable but complement this functional scope *(service value)* by their friendship-like role as companions *(relationship value)* – a characteristic that is crucial for recurring, enjoyable interactions and facilitating desired behavioral changes *(matching value)* for improving time management habits over the long term (Schlimbach et al. 2023; Skjuve et al. 2021).

Although LCs are expected to improve student motivation (e.g., Grivokostopoulou et al. 2020) and time management (e.g., Rodriguez et al. 2019) and thus knowledge for their purposeful design has been recently derived (e.g., Khosrawi-Rad, Schlimbach, et al. 2022), there remains a lack of empirical research on an LC's effectiveness in improving the digital learning experience and learning outcomes over a longer period of usage (Nißen et al. 2021; Zierau et al. 2020). Related research appears to be primarily task-driven, like focusing on scheduling services (Inie and Lungu 2021), goal setting with a rise in user efficiency (Chen et al. 2022; Du et al. 2021) or productivity increase at the workplace (Kimani et al. 2019) and is limited to short-term interactions, though neglecting the bonding and accompanying nature of LCs that emerges over time and that is needed for sustainable learning on the long run (Schlimbach et al. 2023). Therefore, this study aims to provide valuable insights into the effectiveness of an LC for enhancing the student's motivation and time management after accompanying them for a semester to potentially improve their academic performance compared to a control group that attends the same mass lecture but does not study with the LC. We thus strive to contribute to the following research questions (RQs):

RQ1: *To what extent does an LC improve students' digital learning experience in terms of motivation, and time management when accompanying freshmen in an IS course during the semester?*
RQ2: *How does the use of the LC affect students' academic performance, i.e., their subjective knowledge gain and points obtained in the final exam?*

We intend to gather implications on the design of digital learning environments and the development of LCs to better meet the needs of students in IT-enhanced digital learning in higher education.

Research Background

Learning Companions as Social Actors

LCs interact with students in various roles, such as tutor, moderator, mentor, organizer, and motivator (Wellnhammer et al. 2020), all aiming to facilitate learning. Students exhibit a preference for a human-like interaction with the computer, often facilitated by implemented social cues (Feine et al. 2019). Compared to task-oriented agents (like ChatGPT), LCs underscore relationship-oriented interactions. Thus, students instinctively apply social norms and bond socially, when interacting with the computer (Nass and Moon 2000). Winkler & Roos (2019) argue that adaptive and interactive dialogue-based learning systems promote learner engagement by adding a novel dialogue component to technology-based learning systems, in that they can trigger productive peer interaction in digital learning environments – an approach that is particularly crucial for establishing companionship in learning (Schlimbach et al. 2023).

Grounding Research on Motivation and Time Management (in Education)

We ground upon the self-determination theory (SDT) (Ryan and Deci 2000), which posits that environmental factors, such as gamification, influence an individual's basic psychological needs for autonomy, competence, and relatedness, ultimately shape motivation (Haque et al. 2018). The SDT has been widely supported empirically in various settings, including performance increase (Lichtenberg et al. 2021) and education (Niemiec and Ryan 2009). Concerning motivation, the SDT distinguishes between intrinsic and extrinsic motivation and amotivation. Intrinsic motivation is apparent when an individual is motivated by the task itself, whereas extrinsic motivation results when an individual is motivated by external factors, such as rewards or punishment (Amabile et al. 1994). Amotivation, on the other hand, is when an individual is indifferent to a behavior due to a lack of intentional regulation (Gagné and Deci 2005). We adopted the constructs' operationalization as suggested by Amabile et al.'s (1994) questionnaire, as it is well-grounded in SDT and widely used in educational research.

In addition to the SDT, our paper emphasizes the importance of effective time management in planning, organization, and prioritization of tasks to maximize productivity (Grover et al. 2020). Effective time management is essential for individuals to achieve their goals according to goal setting theory (Latham 2012) and maintain a healthy work-life balance (Michalke et al. 2022). Effective time management can enhance an individual's intrinsic motivation by providing a sense of accomplishment and autonomy over their tasks, leading to higher levels of job satisfaction (Parker et al. 2006). Conversely, poor time management can lead to increased stress, procrastination, and burnout, ultimately undermining motivation and productivity (Kimani et al. 2019; Tian et al. 2021).

Deriving Hypotheses

To better understand the effects of a semester-accompanying LC on students' motivation, time management and ultimately learning outcomes, we follow the three-part framework of Koivisto & Hamari (2019): *(1)* usage of the LC, *(2)* its perceived impact, and *(3)* behavioral outcomes. For this purpose, we formulate eleven hypotheses (derived in the following sub-sections) to investigate the relations between the LC application and time management, intrinsic motivation, extrinsic motivation, subjective knowledge gain, and obtained points in the final exam.

Increasing Motivation

Following SDT, LCs may increase learners' intrinsic motivation by simulating social interaction *(relatedness)* or promoting collaboration (Hayashi 2014), facilitating the active use of learning materials and resources to foster progress *(competence gain)* (Winkler et al. 2020), and empowering students to self-regulated learning, thereby promoting *autonomy* (Herrmann-Werner et al. 2021). To measure participants' motivation, we rely on Amabile et al.'s (1994) questionnaire, as it is widely used for measuring students' intrinsic and extrinsic motivation. It is ultimately based on the SDT and divides intrinsic motivation into the sub-scales of enjoyment (E) and challenge (CH). The use of LCs can potentially increase students' intrinsic motivation (Yin et al. 2021), which leads us to ***H1****: The LC has a positive influence on a) enjoyment and b) challenge of the students.*

Intrinsic and extrinsic motivation are closely linked (Amabile et al. 1994; Ryan and Deci 2000). For instance, reward systems, often associated with gamification, can increase both types of motivation (Kodalle and Metz 2022; Stöcklin 2018). Building on Deci and Ryan (1985), we developed independent scoring scales, recognizing the coexistence of intrinsic and extrinsic drivers. Differing from their study, our focus is on directly evaluating intrinsic and extrinsic motivations. Current research indicates the need to further investigate gamification's effectiveness and the effects of extrinsic reinforcement in learning (Kodalle and Metz 2022). Extrinsic motivation is represented through outward orientation (OS) and compensation orientation (CS) (Amabile et al. 1994), which leads us to ***H2****: The LC has a positive influence on a) outward orientation and b) compensation orientation of the students.*

Improving Time Management

Self-regulating abilities such as self-organization and goal-oriented learning are currently the biggest challenges faced by university students (Herrmann-Werner et al. 2021; Rinn et al. 2022). Challenging time management can be supported by to-do lists and reminders for short-term planning (Herrmann-Werner et al. 2021), procrastination-prevention systems or timed learning sessions (Inie and Lungu 2021), and assistance with long-term goals or spaced repetition learning units (Britton and Tesser 1992; Harley et al. 2018). Goal setting theory posits people are motivated by goals directly regulating actions (Latham 2012), impacting short-term planning. Hence, current research on designing LCs suggests integrating goal-setting and time-management support features into these learning facilitators (Schlimbach et al. 2023). Britton and Tesser (1992) conducted a prospective study that confirmed the hypothesis that students' grade point averages can be predicted by time management practices. We leverage their questionnaire as it measures time management in three sub-dimensions (short-range planning (SP), time attitudes (TA), and long-range planning (LP) in the context of student life). Building on their findings, we formulate **H3**: *The LC has a positive influence on a) SP, b) TA, and c) LP of the interacting students.*

Motivation and Time Management as Catalysts for Learning Success

We view learning as a dynamic process that encompasses the acquisition and assimilation of knowledge, skills, and attitudes, resulting in applicable outcomes. Although there is no consensus on the actual definition and measurement of learning success (Edelmann and Tippelt 2004), Plass and Pawar (2020) define the term as the objective of a learning activity that leads to a change in a learner's knowledge, skills, or attitudes. It is typically diagnosed by learning variables such as the student's current knowledge level (Plass and Pawar 2020). However, in higher education, the assessment of learning success is usually oriented towards standardized knowledge tests, even though tested knowledge is not enough to measure learning success overarchingly (Stratmann et al. 2009). Former research has revealed that students' subjective assessments are an indicator of their ability to apply what they have learned in different contexts (Hamminger 2020; Hattie 2012). Our study follows the suggestion of Eckardt et al. (2017) to operationalize learning success by combining measured knowledge gain criteria *(exam score)* and self-assessment *(subjective knowledge)*. We assess subjective knowledge using the five items presented by Flynn and Goldsmith (1999). Hsu et al. (2023) show that learning motivation leads to higher perceived knowledge transfer (Lechler et al. 2019) and thus the better achievement of course goals (Yin et al. 2021). Increasing the quality of the learning process through intrinsic motivation as well as augmenting the quantity of learning through extrinsic motivation improves learning outcomes (Becker et al. 2007; Cerasoli et al. 2014). Accordingly, we formulate **H5**: *The challenge (intrinsic motivation) has a positive influence on a) the subjectively perceived knowledge and b) the students' exam scores.* **H6**: *The outward orientation (extrinsic motivation) has a positive influence on a) the subjectively perceived knowledge and b) the exam results of the students.* **H7**: *The compensation orientation (extrinsic motivation) has a positive influence on a) the subjectively perceived knowledge and b) the students' exam scores.*

It has been shown that the success of students in (online) courses can best be predicted by good time management and motivation (Basila 2014). Abu Saa et al. (2019) identified both variables as significant predictors of students' performance at university in a comprehensive literature analysis, too. The ability to manage time not only strengthens the alignment of learning goals and learning outcomes (Hein et al. 2019) but is also directly related to academic performance (MacCann et al. 2012). Therefore, for the construct time management, we hypothesize **H8**: *Short-range planning has a positive influence on a) the subjectively perceived knowledge and b) the exam score of the students.* ***H9****: Time attitudes have a positive influence on a) the subjectively perceived knowledge and b) the exam score.* ***H10****: Long-range planning has a positive influence on a) the subjectively perceived knowledge and b) the students' obtained points in the final exam.*

Boosting Learning Outcomes with an LC

Adaptive learning systems have been proven to positively affect learning outcomes (Plass & Pawar, 2020). LCs are expected to motivate students (e.g., Grivokostopoulou et al. 2020) and facilitate time management (e.g., Rodriguez et al. 2019). Current research (e.g., Khosrawi-Rad, Schlimbach, et al. 2022) proposes design knowledge for reaching these intended effects. For example, by sending alerts and teaching time management strategies (Schlimbach et al. 2023), by using the scaffolding technique (Winkler et al. 2020), or by motivating with gamified elements to visualize learning progress (Kovisto and Hamari 2019). We anticipate **H11:** *The LC has a positive impact on a) the subjectively perceived knowledge gain and b) the obtained score in the final exam.*

Research Design

Designing Alex as a Situational Artifact

We developed LC Alex using the Rasa platform and DIET classifier, which we integrated into the digital learning platform Slack to enable text-based dialogues between students and the LC, as well as among peers. This integration provides device independence for users. When a student sends a message to Alex (the LC in Slack) to pursue a specific interaction intent, Slack sends an HTTP request to a Docker virtualized web server, which identifies the learner using Slack's anonymous user ID. Docker simplifies the development and deployment of server content and hosts the Rasa server, which recognizes the user's intent based on training data. To generate dynamic responses, the request is forwarded to a webhook that accesses and writes to a non-personalized database and external API interfaces (e.g., Google Search, YouTube). The webhook server, programmed in Python, is also hosted as a container in the Docker network (see Figure 1). Additionally, students can anonymously view their peers' solutions to quiz questions and tasks on a website programmed with Django. This external knowledge repository is primarily intended to promote collaboration and provide a useful preparation tool for the end-of-semester exam.

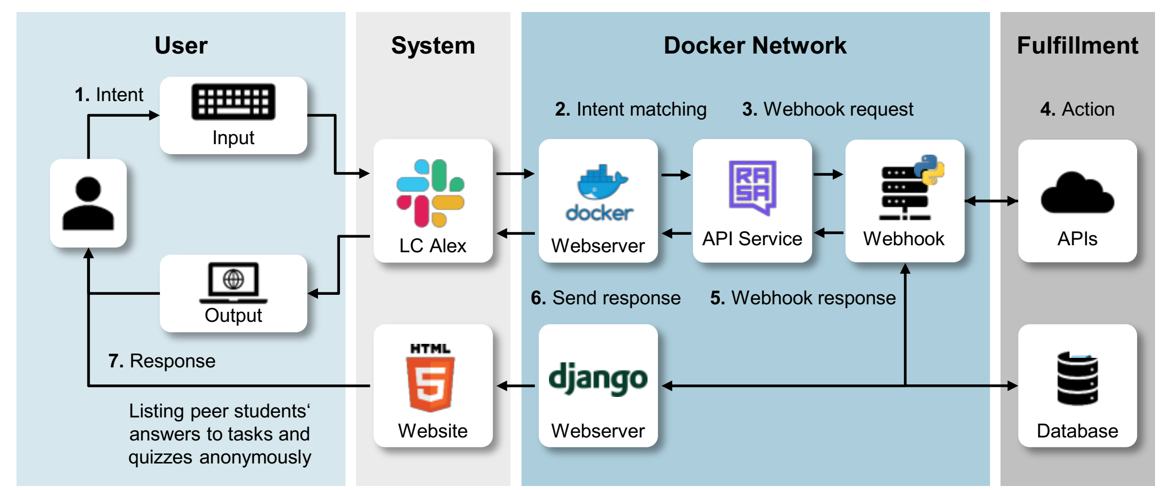


Figure 1. The Architecture of LC Alex

Current research suggests design knowledge for conversational agents to support time management (e.g., Kimani et al. 2019; Schlimbach et al. 2023) and goal setting (Chen et al. 2022; Du et al. 2021). Khosrawi-Rad et al. (2022) introduce nine design principles (DPs) for virtual LCs that teach knowledge, motivate and provide time management support for (university) students in a socially bonding manner. As the proposed design knowledge aligns with our purpose, we adopt their nine DPs as a grounding knowledge base and instantiate matching features for our LC 'Alex' to be implemented as summarized in Table 1.

No.	Design Principle	ID	Design Feature
DP1	Human-likeness and Dialogue Management	DF1.1	Personal salutation and self-referencing as w ell as natural communication (Alex: calling students by their first name, informal chat)
		DF1.2	Human-like avatar and its dynamic evolution (Alex as a female peer student)
		DF1.3	Human-like behavior & communication (Alex makes jokes, uses emojis, humor and memes and chats naturally like a peer student)
		DF1.4	Exposure of a personality (Alex as an older peer student w ith a lot of empathy for the student's challenges)
		DF1.5	Features of emotional intelligence (Alex motivates, asks for emotional state and reacts to the learner's mood)
DP2	Adaptabtability and Adaptivity	DF2.1	Adaptable LC features (defining a learning goal w ith Alex to then adapt it successively)
		DF2.2	Personality-adaptivity (not implemented)
		DF2.3	Preferred learning style (e.g., visual video content vs. text-based response)
		DF2.4	Context-aw areness (Alex relates to course schedule and remembers joint goals to w ork tow ard)
		DF2.5	Personalized informal conversations (e.g., small talk)
DP3	Proactive and Reactive Behavior	DF3.1	Proactivity (proactively shared learning tips, recommendations for task planning, proactive reminding of deadlines)
		DF3.2	Features for active distraction avoidance to foster concentration (Alex uses a timer and teaches Pomodoro technique)
		DF3.3	Reactivity (answ ering organizational questions, mental support for individual challenges)
DP4	Relationship Building	DF4.1	Emotional and mental support (Alex teaches strategies to stay motivated and avoid procrastination)
		DF4.2	Promoting common ground and a shared mental model (a shared SMART goal that Alex relates to)
DP5	Provision of Supportive Content	DF5.1	Teaching content on learning techniques and strategies as w ell as methods of time management (Alex provides videos and tools)
		DF5.2	Provision of learning content (here: 500 learn nuggets related to the lecture 'Introduction to Information Systems')
		DF5.3	Challenges (Alex challenges student in multiple-choice questions and open question in 6 progressive levels)
		DF5.4	Support in collecting learning materials (peer students share their solutions for the LC's challenges on a linked w eb space)
		DF5.5	Answ ers to specific learning questions (Alex collects FAQs in a specific Slack channel)
DP6	Fostering Learning Competencies	DF6.1	Learning advice and tips (Alex has an implemented feature to regularly share advice on time management)
		DF6.2	Encouraging communication behavior that stimulates self-reflection through a mentoring role (Alex reflects on progress made)
DP7	Motivational Environment	DF7.1	Gamified features (Alex motivates w ith levels and badges for accomplishments)
		DF7.2	Motivational, friendly and supportive communication (e.g.,congratulatory messages and gifs w hen a student levels up)
		DF7.3	Netw orking w ith peers (Alex invites peer students in Slack w ho are w orking on the same level to collaborate)
DP8	Ethical Responsibility	DF8.1	Explainability (provision of data processing information during the onboarding phase)
		DF8.2	Technostress avoidance (adapt settings like push-up notifications or learning goal and time frames for reminders)
		DF8.3	Features to fulfill ethics guidelines for AI & LCs (Alex's portrait represents a minority; human in the loop in monitoring the LC's training)
DP9	Purpose-oriented Functionality & Usability	DF9.1	Customizability of the application's functionality (adaptable schedules; alerts and reminders can be turned off or personalized)
		DF9.2	Mindful learning (Alex times a short break after learning for 25 minutes and informs about course schedule)
		DF9.3	Features for effective time management (to-do lists, Alex informs about next course session and reminds for learning together)
		DF9.4	Connectivity w ith internal and external interfaces (Slack used as a platform for collaboration that links to external ressources)
		DF9.5	Ergonomic design and ease of use (e.g., responsive design and easy navigation w ith symbols in home menu)

Table 1. The LC's Feature Set Aligned to Reused Design Principles

Alex is trained on approximately 500 small learning nuggets (DF5.2) from course-relevant content, as well as on small talk (DF2.5) and time-managing interactions (DF3.3). The learning content is conveyed using learning strategy methods (DF5.1), such as a Pomodoro time tracker (Almalki et al. 2020) (DF3.2), and is tested in six quizzes, each with 6-12 multiple-choice and free-text questions (DF5.3). Alex promotes not only subject matter competence, but also learning-supportive functions such as deep learning (Elbyaly and Elfeky 2022), spaced repetition (Tabibian et al. 2019), active recall (Cohn et al. 1996), interleaving (Rohrer 2012), and motivating communication (DF4.1), which contribute to the successful completion of a level (DF7.1). When selecting the DFs to be implemented, care was taken to ensure that all DFs are covered in a balanced manner while also consciously designing the various levels of valuable interactions, as the focus of this work is to apply previously evaluated design knowledge to the application environment to highlight the effects of an LC on motivation, time management, and ultimately, learning outcomes.

Alex establishes a close relationship with learners and adapts to their learning preferences to create a strong matching value and value in relatedness (Geiger et al. 2021). The LC introduces itself as an older student who has already passed the final exam (DF1.4), using a portrait of a student and "Alex" as an internationally common name (DF1.2), who addresses learners by their first name (DF1.1). Alex guides them in setting a personalized SMART goal (DF4.2) (Chen et al. 2022) and suggests useful tools to try out such as *Trello*, *Keepmeout*, or *ilovepdf*. To strengthen the virtual companionship, Alex uses emojis, humorous memes (DF1.3), and rewarding messages for reaching common goals or mastering a new level (DF7.2). To support self-directed learning, Alex aligns with lecture sessions and adapts quiz questions' difficulty levels throughout the semester. Within each of the six levels, tasks become more complex both formally and content-wise. Questions transition from multiple-choice to free-text to increase active recall. The questions not only test factual knowledge but also require students to make connections to previous levels. Alex provides personalized feedback on answers, offering praise and a new challenge for correct responses and text, image, or video help for incorrect answers. Its support becomes more extensive with repeated incorrect answers to dynamically adapt (Schlimbach et al. 2022). The LC facilitates time management by sending push notifications for upcoming lectures, exercises, spaced repetition sessions, and holiday greetings (DF3.1, DF9.3). These can be turned off anytime to maintain learners' autonomy. Alex also recommends a 5-minute break after every 25-minute study session in line with the Pomodoro technique (Almalki et al. 2020) and monitors learners' mental health (DF1.5) to promote well-being. To avoid cognitive overload, Alex restricts the use of multimedia content (DF9.4) and improves user-friendliness with a sleek start menu (DF9.5). The LC continuously encourages social interaction and collaborative learning by facilitating group chats among peers (DF7.3) and by sharing learning materials in the web application (DF5.4).

Figure 2 depicts some sample screenshots of the instantiated application to get a visual impression.

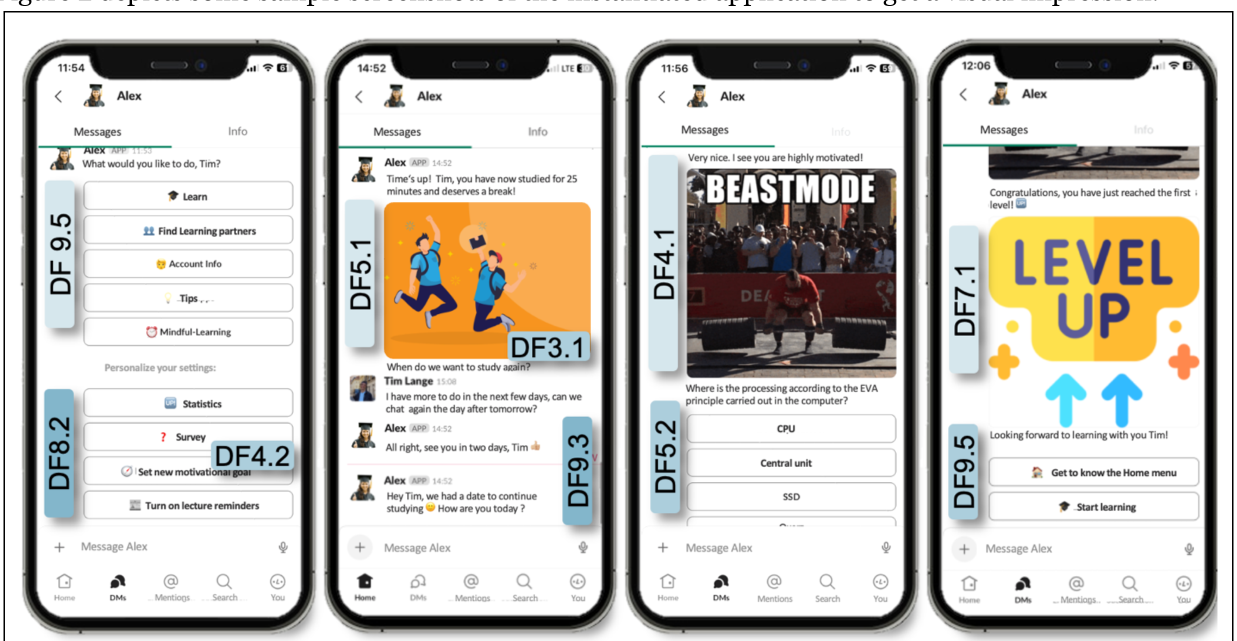


Figure 2. Sample Screens of the LC Application

The Field Experiment

Although we offered all 335 enrolled students of the mass lecture *'Introduction to Information Systems'* at TU Braunschweig (Germany) to learn with Alex throughout the semester by participating in the field study, only 41 students registered in the application, with 15 females and 26 participants completing the pre-test survey. Participation was voluntary and unpaid. These students then studied accompanied by Alex who was fed with knowledge on the content of each lecture and equipped with the features priorly illustrated in Table 1. After three months of Alex's go-live (at the end of the term and before the final examination), a post-test was conducted to measure the same constructs as in the pre-test, both for the students using Alex and the control group enrolled in the same course (same lecturer, peers, course material and exam) who had not interacted with the LC. Additionally, we measured the perceived usefulness and ease of use in the Alex test group following the technology acceptance model (Venkatesh et al. 2003).

As mentioned, we surveyed students as part of the registration process (pre-test) and after three months of usage (post-test). The survey takers answered questions about time management and intrinsic and extrinsic motivation based on reputable constructs with subscales for more granular analysis. As introduced in the section about hypotheses derivation (cf.p.3f.), we built upon Britton and Tesser's (1992) study to test the hypothesis that a student's GPA can be predicted by time management practices and adapted Amabile et al.'s (1994) questionnaire to our use case to measure participant motivation, covering its subscales for intrinsic motivation (enjoyment and challenge) and extrinsic motivation (compensation and outward scale). Besides, we measured subjective knowledge by reusing the five items presented by Flynn and Goldsmith (1999). All constructs were measured on a 7-point Likert scale. We conducted a reliability analysis and confirmatory factor analysis for the pre-and post-test, which confirmed that the respective dataset was suitable for analysis (Cronbach's alpha > 0.7 for all constructs; corrected item-scale correlation consistently > 0.3). In addition to reliability, we assessed the convergent and discriminant validity of the model in the post-test according to Hair et al. (2021b) (for details see the result chapter).

According to Hobert (2019), researchers often focus on specific aspects of conversational agents' design that are relevant to their use case and do not consider users' perceptions from a holistic perspective. To address this issue, we apply the Technology Acceptance Model (TAM) which is ascribed as the most commonly cited and used model for technology adoption (Warkentin et al. 2007). While several advanced adoption models have been proposed, most of these models are based on the original TAM constructs of Perceived Ease of Use (PEOU) and Perceived Usefulness (PU), which remain highly representative and influential predictors of technology adoption. Therefore, we measured the PEOU and PU constructs (Warkentin et al. 2007) of the LC group on a 7-point Likert scale to assess technology acceptance among students who had interacted with Alex. Complementarily, at the end of the semester, we conducted a workshop with nine participants from both groups as part of the explanatory mixed-method approach to gain a deeper understanding and interpretation of the results and to discuss reasons for the rather low participation rate in collaboratively studying with Alex throughout the semester.

Finally, we performed a partial least squares variance-based structural equation modeling (PLS-SEM) to test our previously derived hypotheses using *smartPLS*. We calculated the significance of path coefficients using a bootstrapping resampling method with 10,000 samples, following Chin's (1998) recommendation. The SEM was chosen for the research design with latent variables because it considers the multidimensional structure of the theoretical constructs (Bagozzi and Yi 1988) and enables empirical testing of theoretical statements about complex cause-and-effect relationships (Fuchs 2011). The PLS estimator has advantages regarding restrictive assumptions and is therefore often used in experimental research (Fombelle et al. 2016). Since non-continuous exogenous variables, like the categorical ones in this study, cannot be calculated by covariance-based structural equation modeling and this study aims to determine the presence and direction of the influence of the LC's usage on learning outcomes, following the PLS-SEM approach appeared particularly suitable for testing the hypotheses (Goodhue et al. 2012).

Results

Pre-Test, Post-Test, and Discriminant Analysis

Regarding the pre-test, we analyzed the results by calculating means and standard deviations for the control group compared to the Alex group. We then performed t-tests (see Table 2) to check whether the differences between the two groups were significant (α = 0.05) in the student's t-test or Whitney-Mann-U test (MWU) (which we selected based on the given distribution of the data per construct).

Factor F_1 | F_2		Dependent Variables	Mean Value F_1 | F_2	Std. Dev. s_1 | s_2	df	Statistic	Cohen's d	Rank Biserial	p	alpha
Control vs. LC Test Group	IM	E Enjoyment	3.27 | 3.51	0.87 | 0.86	39	MWU		0.165	0.184	0.801
		CH Challenge	3.02 | 3.18	1.03 | 1.02	39	Student's t	0.158		0.309	0.846
	EM	OS Outward Scale	2.15 | 2.40	1.01 | 1.08	39	MWU		0.179	0.165	0.702
		CS Compensation Scale	3.32 | 3.60	0.85 | 1.13	39	MWU		0.220	0.114	0.801
	TM	SP Short-Range-Planning	3.64 | 3.76	1.01 | 1.01	39	Student's t	0.122		0.350	0.750
		TA Time Attitudes	3.64 | 3.97	0.93 | 0.75	39	Student's t	0.396		0.107	0.888
		LP Long-Range-Planning	3.86 | 4.08	1.07 | 1.00	39	Student's t	0.208		0.255	0.83
	-	SK Subjective Knowledge	3.14 | 3.21	1.08 | 1.29	39	Student's t	0.063		0.421	0.731

Table 2. Pre-test Results to Compare the LC Test Group with the Control Group

We could neither find significant differences, nor large effect sizes between the two groups and thus assume comparable starting conditions for all constructs in both groups. Table 3 summarizes the results of reliability and validity analysis obtained in the **post-test** survey. Following Gefen and Straub's (2005) criteria, only items with a factor loading above the threshold of .60 were included in the analysis. All constructs showed sufficient composite reliability (CR > .80) and average variance extracted (AVE > .50) values, in line with Urbach and Ahlemann's (2010) recommendations, indicating that all measured constructs exhibit adequate reliability.

Dim.	Dependent Variables	Mean Value	Std. Dev.	CR	AVE	alpha	Dim.	Dependent Variables	Mean Value	Std. Dev.	CR	AVE	alpha
IM	E Enjoyment	4.360	1.060	0.863	0.681	0.802	TM	SP Short-Range-Planning	4.039	1.288	0.887	0.724	0.809
	CH Challenge	4.272	0.952	0.953	0.910	0.903		TA Time Attitudes	4.240	1.259	0.902	0.822	0.790
EM	OS Outward Scale	3.206	1.362	0.890	0.670	0.835		LP Long-Range-Planning	4.633	1.045	0.868	0.621	0.803
	CS Compensation Scale	4.985	1.054	0.884	0.721	0.806	-	SK Subjective Knowledge	4.580	0.820	0.908	0.768	0.848

Table 3. Post-Test

The model fulfills all measurement properties: all factor loadings are meaningful and significant, CR is above 0.8, the extracted average variance is above 0.5, and the Fornell-Larcker criterion is met (Fornell and Larcker 1981; Urbach and Ahlemann 2010). We depict the discriminant analysis' results in Table 4.

Dim.	Dependent Variables	E	CH	OS	CS	KP	ZE	LP	EX	SW	G
IM	E Enjoyment	**0.825**									
	CH Challenge	0.174	**0.954**								
EM	OS Outward Scale	-0.152	0.488	**0.819**							
	CS Compensation Scale	-0.012	0.669	0.384	**0.849**						
TM	SP Short-Range-Planning	0.167	0.700	0.492	0.557	**0.851**					
	TA Time Attitudes	0.236	0.244	0.052	0.096	0.068	**0.907**				
	LP Long-Range-Planning	-0.065	0.439	0.286	0.381	0.142	0.060	**0.788**			
-	EX Points in the final Exam	0.161	0.723	0.449	0.635	0.535	0.378	0.435	**n.a.**		
	SK Subjective Knowledge	0.244	0.846	0.505	0.766	0.654	0.247	0.473	0.858	**0.876**	
	G Group	0.011	0.468	0.369	0.462	0.450	0.357	0.177	0.548	0.543	**n.a.**

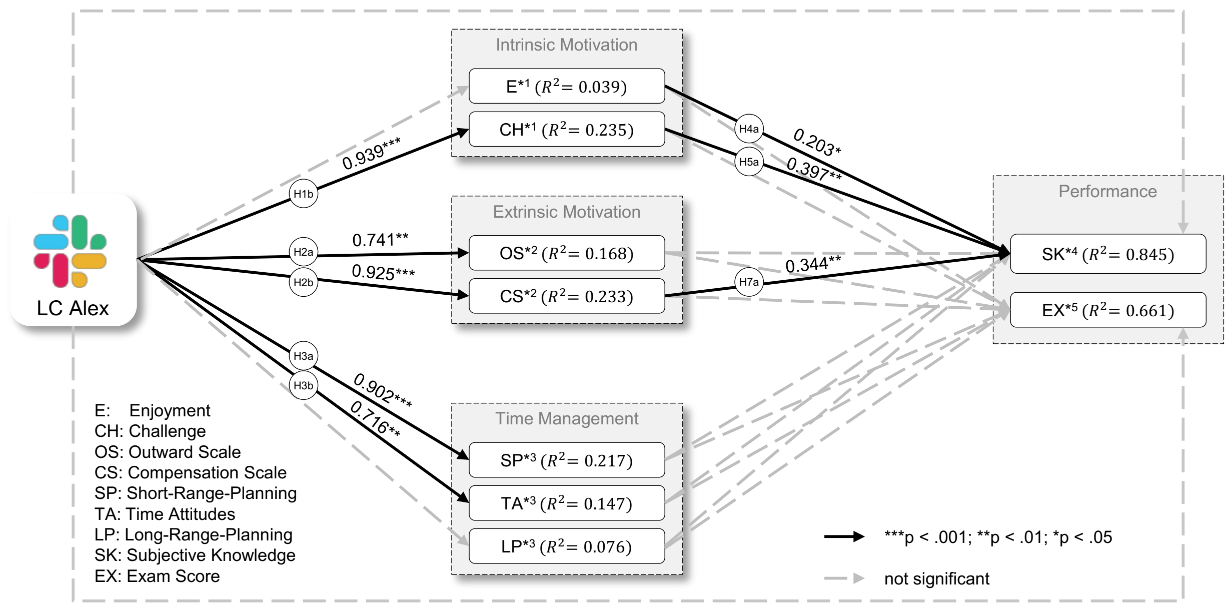
n.a. = not applicable

Table 4. Discriminant Analysis

Thus, all preliminary requirements were met for structural modeling.

Running a Structural Model

Obtained results with their respective path coefficients (ß-values), coefficients of determination (R^2-values), and significance levels (p) are visualized in Figure 3 and Table 5 on the next page.



*1,*2 Amabile et al. (1994), *3 Britton, B.K. und Tesser, A. (1991), *4 Flynn und Goldsmith (1999), *5 Student's Final Exam Score

Figure 3. Structural Model (PLS-SEM)

Our PLS-SEM analysis has revealed that intrinsic motivation, extrinsic motivation, and time management explain 84.5% (R^2 = 0.845) of the variance in perceived knowledge and 66.1% (R^2 = 0.661) of the variance in exam scores. According to Cohen (2013), only enjoyment (R^2 = 0.039) and long-range planning (R^2 = 0.076) have low explanatory power ($0.02 < R^2 < 0.13$). Challenge (R^2 = 0.235), compensation orientation (R^2 = 0.233), outward orientation (R^2 = 0.168), short-range planning (R^2 = 0.217), and time management (R^2 = 0.147) have moderate to high explanatory power. Subjective knowledge (R^2 = 0.845) and exam scores (R^2 = 0.661) have even high explanatory power ($R^2 > 0.26$).

We summarize the results of our tested hypotheses in Table 5, whereby the supported ones ($p < 0.05$) are printed in bold. Although we could not prove a significant direct effect of the LC on the learning outcome (H11), students of the LC test group achieved on average 61.25 points in the final exam, while students of the control group achieved only 45.25 out of 90 points in the same examination.

Hypothesis		Relation	β-value	t-value	p-value	Hypothesis		Relation	β-value	t-value	p-value
H1	a	LC → E	0.012	0.053	0.479	H7	**a**	**CS → SK**	**0.344**	**3.243**	**0.001**
	b	**LC → CH**	**0.939**	**3.950**	**< .001**		b	CS → EX	0.238	1.490	0.068
H2	**a**	**LC → OS**	**0.741**	**2.523**	**0.006**	H8	a	SP → SK	0.018	0.126	0.450
	b	**LC → CS**	**0.925**	**4.181**	**< .001**		b	SP → EX	0.028	0.122	0.451
H3	**a**	**LC → SP**	**0.902**	**3.906**	**<.001**	H9	a	TA → SK	0.012	0.141	0.444
	b	**LC → TA**	**0.716**	**2.864**	**0.002**		b	TA → EX	0.192	1.348	0.089
	c	LC → LP	0.355	0.939	0.174						
H4	**a**	**E → SK**	**0.203**	**1.752**	**0.040**	H10	a	LP → SK	0.123	1.557	0.060
	b	E → EX	0.087	0.624	0.266		b	LP → EX	0.149	1.081	0.140
H5	**a**	**CH → SK**	**0.397**	**2.625**	**0.004**	H11	a	LC → SK	0.237	1.213	0.113
	b	CH → EX	0.294	1.098	0.136		b	LC → EX	0.308	1.229	0.110
H6	a	OS → SK	0.117	1.093	0.137	**bold printed**: hypothesis is supported (p < 0.05)					
	b	OS → EX	0.104	0.739	0.230						

Table 5. Overview of Results for Tested Hypotheses

In addition, we conducted a mediator analysis following Hair et al. (2021a) to explore how enjoyment and challenge *(intrinsic motivation)*, compensation and outward orientation *(extrinsic motivation)*, and short-range planning, time attitudes, and long-range planning *(time management)* mediate the influence of LC on exam outcomes and subjective knowledge gain *(learning outcomes)*. Firstly, we analyzed the total indirect effects. The LC significantly affects subjective knowledge (LC → X → SK; $\beta = 0.851$; $p < .001$) and the exam score (LC → X → EX; $\beta = 0.791$; $p < .001$) by considering all specific indirect influences, while there are no significant direct effects from LC Alex to subjective knowledge gain (LC → SK; $\beta = 0.237$; $p = 0.113$) or exam scores (LC → EX; $\beta = 0.308$; $p = 0.110$). This analysis supports the mediating role of time management (SP, TA, LP) and motivation (E, CH, CS, OS) for both subjective knowledge and exam scores (Hair et al. 2021a). The product of direct and indirect effects for these two variables measuring learning outcomes has a positive algebraic sign, indicating complementary or supporting mediators (ibid.).

Our study proved a significant positive effect of the instantiated LC Alex to boost learning success, and the technology acceptance resulted also high within the user group that tested Alex for three months until the highest level. Among these, 19 students (12 men, and 7 women) completed the TAM questionnaire. Descriptive results revealed that the LC was perceived as easy to use, with a total PEOU score of 5.54 (of 7). The perceived usefulness (PU) was similarly positively rated with a total score of 6.11 (of 7). Positive technology acceptance is particularly important for learning tools to ensure that students perceive the tool as useful, and easy to interact with (Wambsganss et al. 2021). A positive perception promotes motivation and engagement in using the learning application and was observed within the LC test group. However, only a fraction of all enrolled students in the course (19 out of 335 students) actually decided on learning with Alex throughout the semester. Therefore, in the qualitative workshop, we strived for learning more about the underlying reasons for the scarce technology adoption.

Workshop Reflections

To better understand the low participation rate (5.7%) in studying during the semester with Alex, we conducted a Customer Journey Mapping (CJM) workshop with nine students; among these 7 who had interacted at least sporadically with Alex (U1-U7) and 2 who had not interacted with the LC at all (E1, E2). CJM depicts the user's journey from the first contact with a (digital) product along all touchpoints in practical application contexts (Pantouvakis and Gerou 2022). We outlined the developer-initiated customer journey on a digital whiteboard and asked students to write down barriers and frustrating situations for each step of the CJM. The reasons were then discussed, and students compared their experiences mapped to the touch points before clustering frustrating factors. The resulting categories align with Janssen et al.'s (2021) analysis, which systematically brought together twelve critical factors for the failure of chatbots from a total of 154 papers, 25 expert interviews, and 103 chatbots. We refer to these factors as a guiding framework, especially since they consider three core phases *(development, release, and usage)* applicable to our artifact (Janssen et al. 2021).

The user evaluation of the LC showed that students recognized the usefulness of the bot mainly for the exam preparation phase (U1): *"The final exam takes place on March 16th next year [...], once I actually start preparing, I will definitely use your bot, if it is still online, and depending on how much I like it, I will use it to practice/learn. [...] But I believe that it is worth its weight in gold in exam preparation [...]."* This suggests that either the intended use case - namely, recurring interactions along the semester - was not clearly communicated, or that students did not see the value in the idea of companionship initially. In addition, some students did not (physically) participate in the lectures (in Germany neither mandatory nor graded) and therefore missed important information (U3). In addition, three arguments countered students' conviction: first, downloading an unknown application (Slack) for which they had to create an account was a barrier that led to a decrease in technology acceptance for many of them. Second, students reported that Alex being presented by its developer, led to lower expectations towards its usefulness in learning compared to being introduced by the professor (U2, E2). Third, the service-level capabilities of Alex could not keep up with those of ChatGPT, which had just been released in November 2022 with its revolutionary technology maturity. All course participants confirmed that they had experience with ChatGPT and two of them used it for learning (U1, U5). In addition, data privacy concerns may have affected some students' voluntary use of Alex, although students registered with nicknames and remained anonymous in interacting with the LC. The analysis of anonymized user data revealed that only 56 out of 335 (16.72%) students decided to test the chatbot application. Two problems arose during the registration phase: first, it was not clear to students that they had to add the bot as an extra step in Slack to interact with Alex. Second, some students claimed to be deterred by the lengths of the mandatory survey to participate in the field experiment (twice 15-20 minutes for the pre-and post-test).

In live operations, students also reported technical difficulties that led to frustration. One student reported that Alex sometimes *"did not respond or got stuck in a loop"*, and he then *"feared being at fault himself" (U6)*. Some features, such as reminder messages for lectures or learning sessions, were positively rated, but students were not aware of the option to turn off these alerts. 20 students claimed they had gotten annoyed by push-up notifications and by emails sent by Slack for new messages (which could have been turned off as well), so they stopped using Alex right after the first day of usage. Seven more students interacted with Alex recurringly but did not reach the final level, since they did not master the quizzes to level up, which led to some frustration as well. Besides, students were often unsure where they could interact with Alex publicly, privately, or in channels of the new Slack workspace environment and were thus frightened to be *"secretly observed by the professor" (U7)*. Some students reported forgetting to use the application after the lecture, even though Alex proactively sent reminders. Some students had deactivated push-up notifications, and others had Alex only running in the background on their computers, so they missed notifications or reported a larger barrier to logging in to the system. In both cases, the developers' intention was not met and requires improved communication with usage recommendations in the onboarding process.

In Figure 4 we summarize students' feedback gathered in the application's feedback space and during the workshop, while also proposing mitigation strategies.

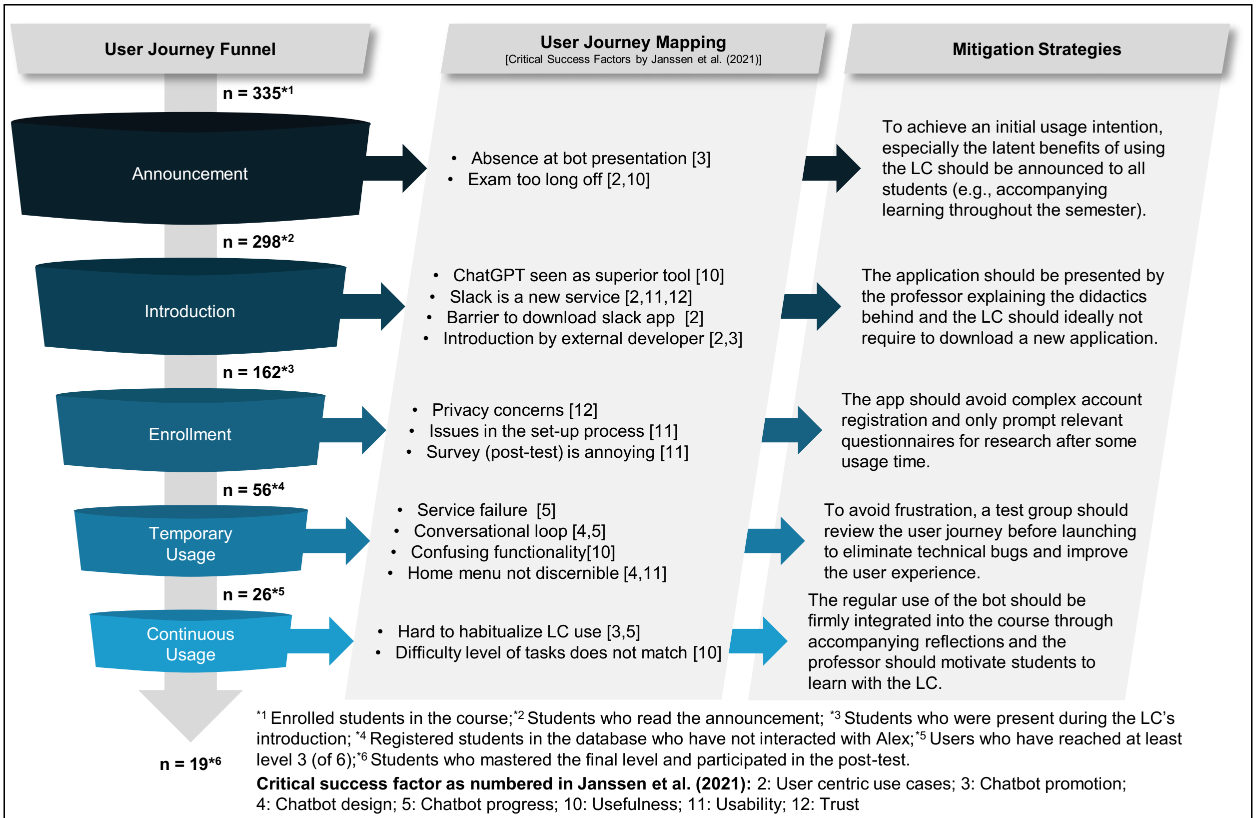


Figure 4. TAM Funnel with Mitigation Strategies

Overall, 19 participants (5.73%) used Alex steadily during the semester in its intended role as a companion and mastered the final level. These 19 participants however were quite enthusiastic about Alex and wished for additional learning content for exam preparation, as revealed by the built-in feedback chat interaction after completing the last level. While for some students, the questions seemed not challenging enough because *"so far it has been very easy to find the correct answers from the script or courseware contents,"* others tried to skip some tasks because they were *"too hard to grasp."* Thus, students request more (cognitively) adaptive elements. Students highly appreciated the LC's remote accessibility, allowing them to *"learn on the go, such as during morning commutes"*. Moreover, they praised Alex's swift retrieval of relevant lecture content that spared them the *"hassle of poring through lengthy course materials"*.

Discussion

This study aimed to investigate the influence of an LC on the motivation, time management, and learning outcomes of freshmen at TU Braunschweig. The results indicate that taking into account the post-test, short-range planning (SP), time attitudes (TA), extrinsic motivation, and challenge (CH) were significantly positively impacted by the use of LC Alex. Moreover, students learning with Alex on average achieved better grades in the final exam (0.7 grading points better on a 5-point grading scale), where motivation and time management acted as mediators. More concretely, the LC establishes an environment that empowers learners to enhance motivation and plan time effectively, culminating in improved learning outcomes. This underscores the LC's role not merely as an information conduit, but as a guiding catalyst for boosting (intrinsic) motivation and cultivating time management skills.

Regarding students' **intrinsic motivation**, our study found that the variable CH can be significantly increased through the use of an LC, possibly due to embedded quizzes that offer different difficulty levels and feedback. However, the use of LC Alex did not significantly increase enjoyment (E), which may be due to technical problems during live operations or little interest in the topic. Additionally, **extrinsic motivation** was significantly increased in both sub-dimensions (CS and OS), through the use of the LC. However, caution is advised when relying solely on extrinsic motivation should, as it may create physical pressure or undermine intrinsic motivation (Amabile et al. 1994). For that reason, Alex allows to individually formulate and adjust the goal, and turn off features like push-up notifications for reminders and alerts. Concerning **time management**, the LC intervention significantly improved SP compared to the control group but did not impact long-range planning (LP). According to Britton and Tesser (1992), LP involves setting goals for a longer period (at least a quarter), recording important dates, regularly reviewing material, and completing important tasks early. These skills may be more difficult to learn and change than short-term strategies in SP since they require changing routines and adapting to new circumstances. Based on workshop feedback, we believe students did not use Alex's goal-setting function effectively and require more guidance in its use. Since changing routine behavior with technical tools is challenging, we conclude that stronger long-term support and incentives may be necessary to achieve the intended effects in LP. These findings complement and extend the results of Kimani et al. (2019) to the educational sector, as their six-day investigation in a professional setting yielded limited data for a meaningful comparison of time management.

Regarding **learning outcomes**, significant predictors for subjective knowledge (SK) were compensation orientation, enjoyment, and challenge. Therefore, external motivation and time management dimensions play a minor role in increasing SK. Our study tested the exam score's prediction by seven constructs, resulting in a high coefficient of determination ($R^2 = 0.661$) with no significant predictors. This may be due to data overfitting, multicollinearity, or small sample sizes, which make small effects insignificant (Hair et al. 2021b). The largest effect size was 0.083 for time management, suggesting that the effects are too small with the small sample size (ibid.). However, CS had a nearly significant influence on the exam score (EX), which is supported by the multiple mediator analysis. Motivation and time management act as mediators between personal characteristics and EX. These findings are consistent with previous studies that have identified motivation and time management as significant factors affecting student performance, albeit as one among roughly 200 other factors (Abu Saa et al. 2019). Besides, we observed that all students of the Alex group took the exam, while a quarter of the control group postponed the examination to the next semester. Although we can only hypothesize reasons for that, students from the LC group might have felt more self-confident about passing the exam as they stated that *"Alex was a great buddy to feel ready for the exam"*.

Our findings hold several **implications**, encompassing both theoretical and practical dimensions. From a **research** viewpoint, our web-based educational LC Alex (i.e., situational artifact) and the more abstract knowledge (i.e., design features aligned to design principles) respond to recent calls for virtual companionship in learning (Hayashi 2014; Khosrawi-Rad, Schlimbach, et al. 2022). Unlike prior (design-oriented) studies, we not only anticipated a favorable effect of the LC (Chen et al. 2022; Du et al. 2021; Schlimbach et al. 2023) but actually gathered quantitative results that confirm its positive influence on learning in a field experiment. Compared to Kimani et al. (2019) who showed in a professional setting that conversational agents positively influenced time management in business settings, our artifact especially takes into account the educational context and thus implements organizing, teaching, and motivational features to increase engagement, facilitate time management and ultimately boost learning. Moreover, since existing chatbots in learning tend to lay focus on ad hoc or temporary advice (Nißen et al. 2021), our focus on companionship goes beyond the functional scope of Alex and highlights the relevance of the relationship value instead (Schlimbach et al. 2023). While current research often measures immediate effects after a one-time interaction, our work contributes to the research gap of lacking longitudinal studies that measure the impact on learning over a longer period with recurring interactions – an aspect that has not been covered in any of the 252 analyzed studies in a recent literature review of conversational agents in education (Khosrawi-Rad, Rinn, et al. 2022).

From a **practice** perspective, our work responds to educational needs in terms of skills to jointly learn with the LC by fostering critical thinking, problem-solving, and scaffolding mechanisms rather than having Alex react to prompts with fully elaborated responses. That way, we counteract the current trend to fully rely on disruptive technology (such as ChatGPT) and its knowledge base but empower students instead to learn interactively (and with embedded didactics like scaffolding or Pomodoro technique) to thus promote a deeper understanding of the learning content. Besides, our paper reports on a virtual application and thus adds an important component to asynchronous learning. Lastly, our instantiation goes beyond a conceptual prototype, as Alex is already being used in real-world university courses and might be transferable to other contexts as well (e.g., younger students or work environments). In this course, we identified critical factors for our artifact's rather poor technology adoption along the user journey. Since most studies finish after revealing design knowledge or after proving the artifacts' intended impact on user behavior, we view this final step as particularly crucial to actually leverage the potential of the artifact in practice.

Nevertheless, we do admit several **limitations**. The theoretical model analysis was performed using PLS-SEM due to a small sample size and non-normally distributed constructs. Therefore, further research should investigate the stability of the results with a larger sample. Also, the R^2 values only indicate the model's explanatory power within the sample, not its predictive power beyond it, which could be explored in another model that evaluates its predictive performance. In light of the mediator analysis indicating that LC Alex exerts no direct significant impact on learning success, but instead manifests its influence indirectly via time management and motivation, the iterative methodology advocated by Hair et al. (2021b) would suggest omitting these paths (LC→EX and LC→SK) for predictive considerations. Thus, the evaluation of the predictive significance of this PLS-SEM model could be undertaken, in consonance with Hair et al.'s framework (2021b). Additionally, the timing of data collection was influenced by organizational factors, and crucial junctures such as post-exam periods might have impacted students' attitude towards the LC. Although post-tests revealed no substantial disparities in motivation and time management between the control and test groups at the initial stages of the experiment, it remains plausible that divergences in motivation and time management skills could manifest at a later stage. Moreover, our findings emerge from a field study at a German technical university (TU Braunschweig) and may be limited by cultural bias.

Future research areas related to our findings include further exploring the benefits of continuous, semester-long learning to position LCs as a complementary, collaborative tool to human professors, rather than viewing them as competitive disruptive technologies such as ChatGPT that mediate knowledge but do not collaboratively elaborate on learning content, while also teaching time management strategies and motivate. Another area for future research is analyzing the students' preferred applications to integrate the LC into and thus bypass the download and registration barriers. In this course, the user journey should be tested more intensively to avoid technical frustration sources such as server response overflow and to include relevant questionnaires without deterring users. Finally, we suggest repeating the study in different course settings (e.g., digital vs. hybrid learning; mass lecture vs. seminar), (learning) cultures, and disciplines to compare outcomes respectively and investigate what boosts learning success in each setting.

Acknowledgements

RS and TCL share first authorship. The study was conducted at TU Braunschweig emerging from the project StuBu, funded by the German Federal Ministry of Education and Research (BMBF); Grant #21INVI06.

References

Abdullah, S. I. N. W., Arokiyasamy, K., Goh, S. L., Culas, A. J., and Manaf, N. M. A. 2022. "University Students' Satisfaction and Future Outlook towards Forced Remote Learning during a Global Pandemic," *Smart Learning Environments (9:1)*.**https://doi.org/10.1186/s40561-022-00197-8**

Abu Saa, A., Al-Emran, M., and Shaalan, K. 2019. "Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques," *Technology, Knowledge and Learning (24:4)*, pp. 567–598. **https://doi.org/10.1007/s10758-019-09408-7**

Adamopoulou, E., and Moussiades, L. 2020. An Overview of Chatbot Technology, (Vol. 584 IFIP), *Advances in Information and Communication Technology*, p. 383-394.

Almalki, K., Alharbi, O., Al-Ahmadi, W., and Aljohani, M. 2020. "Anti-Procrastination Online Tool for Graduate Students Based on the Pomodoro Technique," in Learning and Collaboration Technologies. *Human and Technology Ecosystems* (Vol. 12206), Lecture Notes in Computer Science, P. Zaphiris and A. Ioannou (eds.), Cham: Springer International Publishing, pp. 133–144.

Amabile, T. M., Hill, K. G., Hennessey, B. A., and Tighe, E. M. 1994. "The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations," *Journal of Personality and Social Psychology (66)*, US: American Psychological Association, pp. 950–967. **https://doi.org/10.1037/0022-3514.66.5.950**

Bagozzi, R. P., and Yi, Y. 1988. "On the Evaluation of Structural Equation Models," *Journal of the Academy of Marketing Science (16:1),* pp. 74–94. **https://doi.org/10.1007/BF02723327**

Basila, C. 2014. "Good Time Management and Motivation Level Predict Student Academic Success in College On-Line Courses," https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?, IGI Global.

Becker, C., Kopp, S., and Wachsmuth, I. 2007. "Why Emotions Should Be Integrated into Conversational Agents," in *Conversational Informatics: An Engineering Approach*, pp. 49–67.

Behr, A., Giese, M., Teguim Kamdjou, H. D., and Theune, K. 2021. "Motives for Dropping out from Higher Education—An Analysis of Bachelor's Degree Students in Germany," *European Journal of Education (56:2)*, pp. 325–343. **https://doi.org/10.1111/ejed.12433**

Britton, B. K., and Tesser, A. 1992. "Effects of Time-Management Practices on College Grades.," *Journal of Educational Psychology (83:3)*, US: American Psychological Association.

Cerasoli, C. P., Nicklin, J. M., and Ford, M. T. 2014. "Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis.," *Psychological Bulletin (140:4)*, pp. 980–1008.

Chen, Y., Zhu, T., Bhalla, S., Aggarwal, N., and Gee, L. 2022. A Conversational Goal Setting Buddy for Student Learning, presented at the *AMCIS 2022 Proceedings*, Minneapolis, USA.

Chin, W. W. 1998. "Commentary: Issues and Opinion on Structural Equation Modeling," *MIS Quarterly (22:1)*, Management Information Systems Research Center, University of Minnesota, pp. vii–xvi.

Cohen, J. 2013. Statistical Power Analysis for the Behavioral Sciences, (2nd ed.), Routledge.

Cohn, D. A., Ghahramani, Z., and Jordan, M. I. 1996. Active Learning with Statistical Models, arXiv.

Dalipi, F., Imran, A. S., and Kastrati, Z. 2018. "MOOC Dropout Prediction Using Machine Learning Techniques: Review and Research Challenges," in 2018 *IEEE Global Engineering Education Conference (EDUCON)*, Tenerife: IEEE, April, pp. 1007–1014.

Deci, E. L., and Ryan, R. M. 1985. Intrinsic Motivation and Self-Determination in Human Behavior, Boston, MA: Springer US. **https://doi.org/10.1007/978-1-4899-2271-7**

Diederich, S., Brendel, A., Morana, S., and Kolbe, L. 2022. "On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research," *Journal of the Association for Information Systems*.

Du, J., Huang, W., and Hew, K. F. 2021. "Supporting Students Goal Setting Process Using Chatbot: Implementation in a Fully Online Course," in 2021 *IEEE International Conference on Engineering, Technology & Education (TALE)*, December, pp. 35–41.

Eckardt, L., Körber, S., Becht, E. J., Plath, A., Al Falah, S., and Robra-Bissantz, S. 2017. "Führen Serious Games zu Lernerfolg? – Ein Vergleich zum Frontalunterricht," in Gamification und Serious Games: Grundlagen, Vorgehen und Anwendungen, *Edition HMD*, S. Strahringer and C. Leyh (eds.), Wiesbaden: Springer Fachmedien, pp.139–150.**https://doi.org/10.1007/978-3-658-16742-4_11**

Edelmann, D., and Tippelt, R. 2004. "Kompetenz – Kompetenzmessung: Ein (Kritischer) Überblick," Durchblick (3), pp. 7–10.

Elbyaly, M. Y. H., and Elfeky, A. I. M. 2022. "The Role of Metacognition in Promoting Deep Learning in MOOCs during COVID-19 Pandemic," *PeerJ Computer Science (8)*, PeerJ Inc., p. e945.

Feine, J., Gnewuch, U., Morana, S., and Maedche, A. 2019. "A Taxonomy of Social Cues for Conversational Agents," *International Journal of Human-Computer Studies (132)*, pp. 138–161.

Flynn, L. R., and Goldsmith, R. E. 1999. "A Short, Reliable Measure of Subjective Knowledge," *Journal of Business Research (46:1)*, pp. 57–66. **https://doi.org/10.1016/S0148-2963(98)00057-5**

Fombelle, P. W., Bone, S. A., and Lemon, K. N. 2016. "Responding to the 98%: Face-Enhancing Strategies for Dealing with Rejected Customer Ideas," *Journal of the Academy of Marketing Science (44)*, Germany: Springer, pp. 685–706.

Fornell, C., and Larcker, D. F. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research (18:1)*, American Marketing Association, pp. 39–50.

Fuchs, A. 2011. "Methodische Aspekte linearer Strukturgleichungsmodelle. Ein Vergleich von kovarianz- und varianzbasierten Kausalanalyseverfahren," *Working Paper No. 2/2011*, Working Paper, Research Papers on Marketing Strategy.

Gagné, M., and Deci, E. L. 2005. "Self-Determination Theory and Work Motivation: Self-Determination Theory and Work Motivation," *Journal of Organizational Behavior (26:4),* pp. 331–362. **https://doi.org/10.1002/job.322**

Gefen, D., and Straub, D. W. 2005. "A Practical Guide to Factorial Validity Using PLS-Graph: Tutorial and Annotated Example," *Communications of the Association of Information Systems (16:5)*.

Geiger, M., Jago, F., and Robra-Bissanzt, S. 2021. Physical vs. Digital Interactions: Value Generation Within CustomerRetailer Interaction, January 1, pp. 153–165.

Goodhue, D. L., Lewis, W., and Thompson, R. 2012. "Does PLS Have Advantages for Small Sample Size or Non-Normal Data?," *MIS Quarterly (36:3)*, Management Information Systems Research Center, University of Minnesota, pp. 981–1001. **https://doi.org/10.2307/41703490**

Grivokostopoulou, F., Kovas, K., and Perikos, I. 2020. "The Effectiveness of Embodied Pedagogical Agents and Their Impact on Students Learning in Virtual Worlds," *Applied Sciences (10:5)*, p. 1739.

Grover, T., Rowan, K., Suh, J., McDuff, D., and Czerwinski, M. 2020. "Design and Evaluation of Intelligent Agent Prototypes for Assistance with Focus and Productivity at Work," in *Proceedings of the 25th International Conference on Intelligent User Interfaces*, IUI '20, New York, NY, USA: Association for Computing Machinery, March 17, pp. 390–400.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., and Ray, S. 2021a. "Mediation Analysis," in Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook, Classroom Companion: Business, J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray (eds.), Cham: Springer International Publishing, pp. 139–153. **https://doi.org/10.1007/978-3-030-80519-7_7**

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., and Ray, S. 2021b. "Evaluation of the Structural Model," in Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook, Classroom Companion: Business, J. F. Hair Jr., G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray (eds.), Cham: Springer International Publishing, pp. 115–138.

Hamminger, L. 2020. "Lernerfolg aus Sicht berufstätiger Studierender. Erste Schritte auf dem Weg zu einer Messmethode," Magazin erwachsenenbildung.at (40), pedocs.

Haque, M. S., Isomursu, M., Kangas, M., and Jämsä, T. 2018. Measuring the Influence of a Persuasive Application to Promote Physical Activity, presented at the *PPT@PERSUASIVE*.

Harley, J. M., Taub, M., Azevedo, R., and Bouchet, F. 2018. "Let's Set Up Some Subgoals: Understanding Human-Pedagogical Agent Collaborations and Their Implications for Learning and Prompt and Feedback Compliance," IEEE Transactions on Learning Technologies (11:1), pp. 54–66.

Hattie, J. 2012. Visible Learning for Teachers: Maximizing Impact on Learning, Routledge.

Hayashi, Y. 2014. "Togetherness: Multiple Pedagogical Conversational Agents as Companions in Collaborative Learning," in *Intelligent Tutoring Systems*, Lecture Notes in Computer Science, S. Trausan-Matu, K. E. Boyer, M. Crosby, and K. Panourgia (eds.), Cham: Springer International Publishing, pp. 114–123. **https://doi.org/10.1007/978-3-319-07221-0_14**

Hein, J., Kohler, L., and Dickhäuser, O. 2019. "Berufliches Lernen von Wissenschaftler(inne)n: Zeitmanagement als Moderator des Zusammenhangs zwischen Lernzielen und Lernergebnis," Präsentation auf Konferenz presented at the *paEpsy - Fachtagung der Pädagogischen und Entwicklungspsychologie,* Leipzig, Germany.

Herrmann-Werner, A., Loda, T., Junne, F., Zipfel, S., and Madany Mamlouk, A. 2021. "'Hello, My Name Is Melinda' – Students' Views on a Digital Assistant for Navigation in Digital Learning Environments; A Qualitative Interview Study," *Frontiers in Education (5)*, p. 541839.

Hobert, S. 2019. "How Are You, Chatbot? Evaluating Chatbots in Educational Settings–Results of a Literature Review," in *DELFI 2019,* N. Pinkwart and J. Konert (eds.), Bonn: Gesellschaft für Informatik e.V., pp. 259–270. **https://doi.org/10.18420/delfi2019_289**

Hsu, T.-C., Huang, H.-L., Hwang, G.-J., and Chen, M.-S. 2023. "Effects of Incorporating an Expert Decision-Making Mechanism into Chatbots on Students' Achievement, Enjoyment, and Anxiety," *Educational Technology & Society (26:1)*, JSTOR, pp. 218–231.

Inie, N., and Lungu, M. F. 2021. "Aiki - Turning Online Procrastination into Microlearning," in Proceedings of the *2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 6, pp. 1–13. **https://doi.org/10.1145/3411764.3445202**

Janssen, A., Grützner, L., and Breitner, M. H. 2021. "Why Do Chatbots Fail? A Critical Success Factors Analysis," in *ICIS 2021 Proceedings*, Austin, TX, USA.

Khosrawi-Rad, B., Rinn, H., Schlimbach, R., Gebbing, P., Yang, X., Lattemann, C., Markgraf, D., and Robra-Bissantz, S. 2022. "Conversational Agents in Education – A Systematic Literature Review," in *ECIS 2022 Proceedings*, Timișoara, Romania.

Khosrawi-Rad, B., Schlimbach, R., Strohmann, T., and Robra-Bissantz, S. 2022. "Design Knowledge for Virtual Learning Companions," in *Proceedings on the International Conference on Information Systems Education and Research*, Copenhagen, Denmark.

Kimani, E., Rowan, K., McDuff, D., Czerwinski, M., and Mark, G. 2019. "A Conversational Agent in Support of Productivity and Wellbeing at Work," 2019 *8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Cambridge, United Kingdom: IEEE, pp. 1–7.

Kodalle, T., and Metz, M. 2022. "Das Konzept Gamification als spielerisches Lernelement," in Digitale Lernwelten – Serious Games und Gamification, W. Becker and M. Metz (eds.), Wiesbaden: Springer Fachmedien Wiesbaden, pp. 65–78. **https://doi.org/10.1007/978-3-658-35059-8_5**

Kovisto, J., and Hamari, J. 2019. "The Rise of Motivational Information Systems_ A Review of Gamification Research | Elsevier Enhanced Reader." **https://doi.org/10.1016/j.ijinfomgt.2018.10.013**

Latham, G. P. 2012. "Motivate Employee Performance through Goal Setting," in Handbook of Principles of Organizational Behavior, John Wiley & Sons, Ltd, pp. 161–178.

Lechler, R., Stöckli, E., Rietsche, R., and Uebernickel, F. 2019. "Looking beneath the Tip of the Iceberg: The Two-Sided Nature of Chatbots and Their Roles for Digital Feedback Exchange.," in *ECIS 2019 Proceedings,* Stockholm and Uppsala, Sweden.

Lichtenberg, S., Brendel, A., Riquel, J., Diederich, S., and Kolbe, L. 2021. "'Not All Treasures Are Silver and Gold' – Understanding the Gamification Element Lootbox and Its Influence on Motivation and Performance," in *ECIS 2021 Proceedings,* Virtual.

Luxton, D. D. 2014. "Recommendations for the Ethical Use and Design of Artificial Intelligent Care Providers," *Artificial Intelligence in Medicine (62:1)*, pp. 1–10.

MacCann, C., Fogarty, G. J., and Roberts, R. D. 2012. "Strategies for Success in Education: Time Management Is More Important for Part-Time than Full-Time Community College Students," *Learning and Individual Differences (22:5)*, Adult Learning, pp. 618–623.

Michalke, S., Lohrenz, L., Lattemann, C., and Robra-Bissantz, S. 2022. "Exploring Engagement, Well-Being, and Welfare on Engagement Platforms: Insight into the Personal Service Sector from the DACH Region," *Electronic Markets*, pp. 1–16. **https://doi.org/10.1007/s12525-022-00589-1**

Nass, C., and Moon, Y. 2000. "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues (56:1)*, pp. 81–103.

Ngwacho, G. A. 2023. Online Teaching Competencies for Efficacious Competency-Based Implementation in Higher Education, Kenya: IGI Global. (https://www.igi-global.com/gateway/chapter/www.igi-global.com/gateway/chapter/318580).

Niemiec, C. P., and Ryan, R. M. 2009. "Autonomy, Competence, and Relatedness in the Classroom: Applying Self-Determination Theory to Educational Practice," *Theory and Research in Education (7:2)*, pp. 133–144.

Nißen, M., Selimi, D., Janssen, A., Cardona, D., Breitner, M., Kowatsch, T., and Wangenheim, F. 2021. "See You Soon Again, Chatbot? A Design Taxonomy to Characterize User-Chatbot Relationships with Different Time Horizons," *Computers in Human Behavior (127:2)*.

Pantouvakis, A., and Gerou, A. 2022. "The Theoretical and Practical Evolution of Customer Journey and Its Significance in Services Sustainability," *Sustainability (14:15)*, p. 9610.

Parker, S., Williams, H., and Turner, N. 2006. "Modeling the Antecedents of Proactive Behavior at Work," The *Journal of Applied Psychology (91)*, pp. 636–52. **https://doi.org/10.1037/0021-9010.91.3.636**

Plass, J. L., and Pawar, S. 2020. "Toward a Taxonomy of Adaptivity for Learning," *Journal of Research on Technology in Education (52:3)*, Routledge, pp. 275–300.

Rinn, H., Khosrawi-Rad, B., Schlimbach, R., Masurek, M., Robra-Bissantz, S., and Markgraf, D. 2022. Needs of Students in Further Education -A Mixed Methods Study, *GeNeMe 2022 Proceedings*.

Rodriguez, J., Piccoli, G., and Bartosiak, M. 2019. "Nudging the Classroom: Designing a Socio-Technical Artifact to Reduce Academic Procrastination," in *HICSS 2019 Proceedings,* Maui, Hawaii, USA, p. 10.

Rohrer, D. 2012. "Interleaving Helps Students Distinguish among Similar Concepts," *Educational Psychology Review (24:3)*, pp. 355–367. **https://doi.org/10.1007/s10648-012-9201-3**

Ryan, R. M., and Deci, E. L. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American Psychologist (55:1)*, US: American Psychological Association, pp. 68–78.

Schlimbach, R., Rinn, H., Markgraf, D., and Robra-Bissantz, S. 2022. "A Literature Review on Pedagogical Conversational Agent Adaptation," *PACIS 2022 Proceedings*.

Schlimbach, R., Windolf, C., and Robra-Bissantz, S. 2023. "A Service Perspective on Designing Learning Companions as Bonding and Mindful Time Managers in Further Education," in *ECIS 2023 Proceedings*, Kristiansand, Norway.

Skjuve, M., Følstad, A., Fostervold, K. I., and Brandtzaeg, P. B. 2021. "My Chatbot Companion - a Study of Human-Chatbot Relationships," *International Journal of Human-Computer Studies (149)*, p. 102601.

Stöcklin, N. 2018. Vielfältige Möglichkeiten von Gamification. Framework Zur Kategorisierung von Gamification-Ansätzen Im Bildungskontext.

Stratmann, J., Preussler, A., and Kerres, M. 2009. "Lernerfolg und Kompetenz bewerten. Didaktische Potenziale von Portfolios in Lehr-/Lernkontext," *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung (18)*, pp. 1–19. **https://doi.org/10.21240/mpaed/18/2009.12.18.X**

Strohmann, T., Siemon, D., Khosrawi-Rad, B., and Robra-Bissantz, S. 2022. "Toward a Design Theory for Virtual Companionship," *Human-Computer Interaction*.

Sun, F. 2022. "ChatGPT, the Start of a New Era," December 23. (https://feisun.org/2022/12/23/chatgpt-the-start-of-a-new-era/).

Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., and Gomez-Rodriguez, M. 2019. "Enhancing Human Learning via Spaced Repetition Optimization," *Proceedings of the National Academy of Sciences (116:10)*, , pp. 3988–3993.

Tian, X., Risha, Z., Ahmed, I., Lekshmi Narayanan, A. B., and Biehl, J. 2021. "Let's Talk It Out: A Chatbot for Effective Study Habit Behavioral Change," *Proceedings of the ACM on Human-Computer Interaction (5:CSCW1)*, pp. 1–32. **https://doi.org/10.1145/3449171**

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., and Agyemang, B. 2023. "What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education," *Smart Learning Environments (10:1)*, p. 15. **https://doi.org/10.1186/s40561-023-00237-x**

Urbach, N., and Ahlemann, F. 2010. "Structural Equation Modeling in Information Systems Research Using Partial Least Squares," *Journal of Information Technology Theory and Application (11)*.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly*, pp. 425–478.

Wambsganss, T., Höch, A., Zierau, N., and Söllner, M. 2021. "Ethical Design of Conversational Agents: Towards Principles for a Value-Sensitive Design," *Wirtschaftsinformatik 2021 Proceedings*.

Warkentin, M., Shropshire, J., and Johnston, A. 2007. The IT Security Adoption Conundrum: An Initial Step Toward Validation of Applicable Measures., (Vol. 3).

Wellnhammer, N., Dolata, M., Steigler, S., and Schwabe, G. 2020. Studying with the Help of Digital Tutors: Design Aspects of Conversational Agents That Influence the Learning Process, presented at the *Hawaii International Conference on System Sciences*. **https://doi.org/10.24251/HICSS.2020.019**

Winkler, R., Hobert, S., Fischer, T., Salovaara, A., Soellner, M., and Leimeister, J. M. 2020. "Engaging Learners in Online Video Lectures with Dynamically Scaffolding Conversational Agents," *Proceedings of the 28th European Conference on Information Systems*.

Winkler, R., and Roos, J. 2019. Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors, p. 10.

Yin, J., Goh, T.-T., Yang, B., and Xiaobin, Y. 2021. "Conversation Technology with Micro-Learning: The Impact of Chatbot-Based Learning on Students' Learning Motivation and Performance," *Journal of Educational Computing Research (59:1)*, pp. 154–177. **https://doi.org/10.1177/0735633120952067**

Zierau, N., Elshan, E., Visini, C., and Janson, A. 2020. "A Review of the Empirical Literature on Conversational Agents and Future Research Directions," in *Proceedings of the 41th International Conference on Information Systems*, virtual, December 14.