Please quote as: Khosrawi-Rad, B., Keller, P. F., Benner, D., Grogorick, L., Borchers, A., Janson, A., Leimeister, J. M. & Robra-Bissantz, S. (2025). Promoting students' motivation in language education with gamified pedagogical conversational agents. Computers & Education, 238, 1-24. doi: 10.1016/j.compedu.2025.105374

Computers & Education 238 (2025) 105374

Contents lists available at ScienceDirect

Computers & Education

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

Promoting students' motivation in language education with gamified pedagogical conversational agents

Bijan Khosrawi-Rad [a,*], Paul Felix Keller [a], Dennis Benner [b], Linda Grogorick [c], Arne Borchers [a], Andreas Janson [d], Jan Marco Leimeister [b,d], Susanne Robra-Bissantz [a]

[a] *Institute for Business Information Systems, TU Braunschweig, Mühlenpfordtstr. 23, 38106, Braunschweig, Germany*
[b] *Research Center for Information Systems Design (ITeG), University of Kassel, Pfannkuchstraße 1, 34121, Kassel, Germany*
[c] *Didactics of Computer Science, University of Wuppertal, Gaußstraße 20D-42 097, Wuppertal, Germany*
[d] *Institute of Information Systems and Digital Business, University of St. Gallen, Müller-Friedberg-Strasse 8, 9000, St. Gallen, Switzerland*

ARTICLE INFO

Keywords:
Gamification
Pedagogical conversational agent
Design science research
Language learning
Artificial intelligence

ABSTRACT

Pedagogical conversational agents (PCAs) like chatbots are a novel approach to technology-mediated language learning with artificial intelligence. They convey learning content interactively and accompany students in their education. However, many users find conversations with PCAs unmotivating. Gamification is a suitable solution to these motivational hurdles due to its playful nature. Given the difficulty of selecting the appropriate game elements and the scarcity of design recommendations for gamified PCAs, we propose the GNPL framework including a cohesive set of four design principles: goal-setting and reflection, novice-expert relationship, performance-related motivation, and learning story narration. In two design cycles, the article shows the application of the design principles in English learning – a domain commonly associated with motivational challenges – by implementing and evaluating a gamified PCA. The results show that the design principles significantly foster learners' motivation and that learners perceive a solid language learning experience, expressed by higher perceived value and social factors. They highlight the relevance of aligning the PCA's social role, the motivational impact of gamification, and the educational goals of the learning context. The design principles guide educators and developers in gamified PCA design. The paper contributes to the theory stream of PCAs by investigating learners' motivation enhancement when using PCAs. In addition, the paper provides new knowledge on meaningful gamification in an unexplored context and practical insights to solve the design challenges of selecting game elements in this context. Furthermore, it shows how language education can be supported by educational technology.

* Corresponding author.
E-mail addresses: b.khosrawi-rad@tu-braunschweig.de (B. Khosrawi-Rad), paul-felix.keller@tu-bs.de (P.F. Keller), benner@uni-kassel.de (D. Benner), grogorick@uni-wuppertal.de (L. Grogorick), arne.borchers@tu-bs.de (A. Borchers), andreas.janson@unisg.ch (A. Janson), janmarco.leimeister@unisg.ch (J.M. Leimeister), s.robra-bissantz@tu-bs.de (S. Robra-Bissantz).

https://doi.org/10.1016/j.compedu.2025.105374
Received 1 August 2024; Received in revised form 28 May 2025; Accepted 30 May 2025
Available online 5 June 2025
0360-1315/

1. Introduction

1.1. Background

1.1.1. Relevance of language learning in the 21st century

Learning a foreign language is a key 21st-century skill (The Partnership for 21st Century Learning, 2024; Wu et al., 2024). In today's globalized world, proficiency in foreign languages, such as English, is a key factor for personal and professional success (ibid.). Learning a new language is often challenging due to its complexity, requiring continuous effort and frequent repetition (Gao & Liu, 2022; Y. Wang et al., 2021). These challenges lead to high drop-out rates in language learning (Vázquez Mendoza et al., 2024). These challenges underscore the need for innovative support tools, which the following subsection explores in more detail.

1.1.2. Potentials and motivational challenges of pedagogical conversational agents

To support learners in this regard, higher education increasingly uses technology-mediated learning instead of non-interactive material like slides or textbooks (Haleem et al., 2022). This approach is useful because students are often delegated to out-of-class learning tasks (Behr et al., 2021). One possibility for technology-mediated learning is to employ pedagogical conversational agents (PCAs) that act as human-like learning partners (e.g., chatbots) (Hobert & Meyer von Wolff, 2019). Technological advances in artificial intelligence (AI) have led to people increasingly using chatbots for daily tasks (X. Huang et al., 2023). PCAs can teach learning content, accompany students in their studies, and track their progress (Wollny et al., 2021). Usually, tutors or teachers are in charge of supporting their students in this regard, which is not always possible due to time constraints (Behr et al., 2021). PCAs can take over these roles from tutors or teachers due to their independence in time and location (Hobert & Meyer von Wolff, 2019). Despite their potential, PCAs are not yet widely used in educational practice. For instance, a practice analysis by Janssen et al. (2021) found that all of the PCAs studied were no longer operated by their providers. One major issue is that PCAs often fail to motivate learners if they are not designed adequately (Chiu et al., 2023; Janssen et al., 2021). They often neglect to meet learners' expectations for engaging interactions, which limits the PCAs' motivational effectiveness (ibid.). Empirical studies reveal that learners wish for more motivational support, especially in digital learning (Rinn et al., 2022; Behr et al., 2021). Motivating students in digital learning is a key challenge for educators, including the selection of the right motivational incentives and their thoughtful implementation (Delaney et al., 2019). To offer a motivational incentive, this research project draws on gamification, and the next subsection delves into gamification's role in fostering motivation in the PCA context.

1.1.3. Gamification for supporting motivation within pedagogical conversational agents

To address PCAs' motivational challenges, gamification is a promising approach in education. Gamification is a common way to engage learners in educational activities by employing game elements like rewards or storytelling (Deterding et al., 2011). Gamification's playful nature engages learners in topics that they would otherwise find unappealing (Manzano-León et al., 2021). Gamification suits low-threshold learning settings, such as mobile and interactive learning with PCAs (Benner et al., 2022). Especially in language learning, motivating dialogs are essential due to the learning domain's complexity (Gao & Liu, 2022; Y. Wang et al., 2021). For instance, interactive and practice-relevant game stories could help learners train a new language with a PCA in an engaging way. However, there is limited knowledge on how to use gamification for PCAs effectively and which game elements should be integrated into PCAs for a meaningful gamification approach. Studies highlight gamified PCAs' potential (Ruan et al., 2019; Smutny & Schreiberova, 2020). Nevertheless, research lacks theoretically grounded recommendations that guide educators and developers in deciding on an adequate design of gamified PCAs, given the range of options (Díaz & Nussbaum, 2024; Wittmann & Morschheuser, 2022). While we investigate these challenges in the domain of language education, the motivational shortcomings are also present in other areas of technology-mediated learning (Schöbel et al., 2023), which highlights the broad relevance of the prescriptive design knowledge presented in this paper. This paper addresses these design challenges by guiding gamified PCA developments. Therefore, it derives DPs consolidated under the "GNPL" framework for gamified PCAs and empirically tests their impact on learners' motivation in language education. To situate our work within current academic discourse and to lay the groundwork for deriving the theory gaps, the following section reviews key prior language learning, gamification, and PCA studies.

1.2. Literature review on gamified pedagogical conversational agents

PCAs are intelligent dialog systems used for educational purposes (Hobert & Meyer von Wolff, 2019). PCAs evolved from intelligent tutoring systems, the pioneers of dialog-based learning (Atkinson, 1968). These have been enhanced with technical advancements for instruction and feedback over the past decades (Kulik & Fletcher, 2016). PCAs extend the mere teaching of content and support learners in solving challenges and monitoring their progress (Wollny et al., 2021). The *Computers & Education* community is increasingly exploring their effective application. For example, Schneider et al. (2022) demonstrated that human-like PCA features foster learning outcomes. Sikström et al. (2022) emphasized the need for better relational communication. Nguyen (2023) showed that PCAs foster complex skills like systems thinking and emphasized the need for research on designing their social roles. Lo et al. (2024) highlighted AI advancements with ChatGPT and the importance of designing PCAs with motivational elements, while also considering potential user disengagement. PCAs interact with their users in natural language, either text-based as chatbots or voice-based (Hobert & Meyer von Wolff, 2019). For language learners, PCAs offer 24/7 scalability for continuous learning beyond the classroom and location-independent mobile learning (X. Huang et al., 2023; Tai, 2024). Additionally, PCAs provide targeted vocabulary and grammar practice support through personalized assessments and immediate feedback (Jeon et al., 2023). PCAs support students in learning through self-regulation, which means that learners control the learning process themselves (Hien et al., 2018). For PCAs to be accepted, students need to be motivated to learn with them (X. Huang et al., 2023). However, PCAs do not yet sufficiently exploit their potential for motivational support. For instance, Fryer et al. (2017) revealed that while the interest in solving tasks with a human partner does not decrease over time, it declines when learning languages with a chatbot.

Gamification is “the use of game design elements in non-game contexts” (Deterding et al., 2011, p. 10). Gamification studies have mainly shown positive outcomes. Gamification could, for instance, promote problem-based learning in math education (Alt, 2023), students’ foreign language speaking abilities (Hwang et al., 2024), or their motivation in Japanese language learning (C.-M. Chen et al., 2023). To achieve positive outcomes, educators should adopt a “meaningful design” approach, which includes carefully selecting the game elements (Nicholson, 2015). If gamification is not designed meaningfully, it can have the opposite effect and negatively influence students’ psychological, emotional, and social well-being (Melo et al., 2020). An inadequate game element design, for instance, leads to frustration and performance pressure (Sailer et al., 2017). Gamified learning systems require a balance between game and educational design, as overemphasizing learning can reduce enjoyment (Zichermann & Cunningham, 2011). Excessive focus on gaming may, on the other hand, promote addictive behavior (Melo et al., 2020). The gamification design is influenced by the context in which it is applied, as in this case, the context of PCAs (Hamari et al., 2014). We base the PCA’s gamification design in this paper on Schöbel et al.’s (2020) taxonomy. This taxonomy includes the following game elements: points, badges, virtual goods, levels, progress bars, goals, feedback, reminders, narratives, avatars, rankings, and time pressure. Points, badges, and virtual goods reward learners for certain activities. Levels and progress bars document users’ progress. Goals are achievable steps learners can reach, and feedback provides information on their performance. Reminders visualize learners’ previous behavior. Narratives tell a story, and avatars are visual representations of the user or the PCA. Rankings compare learners’ performance, and time pressure challenges them when completing activities. While gamification research has a long tradition for digital learning, its combination with AI-based PCAs is an emerging research topic (Schöbel, Saqr, & Janson, 2021). Some game elements are used often for PCAs (e.g., feedback or avatars), while others are used rarely (e.g., levels or progress bars) (Benner et al., 2022).

Regarding **gamified PCAs**, Smutny and Schreiberova (2020) showed that the most successful Facebook Messenger PCAs use gamification. In an analysis of Facebook Messenger chatbots, the authors found that all of the top ten applications employ some form of gamification (Smutny & Schreiberova, 2020). However, these implementations primarily rely on isolated elements such as quizzes, achievement points, or daily challenges rather than a comprehensive gamification concept. Thus, PCA developers have so far integrated game elements only sporadically and not in a systematic manner, a gap we address in this paper by deriving theory-supported DPs. Initial prototype evaluations moreover show that gamification suits mobile learning via Telegram and promotes learning

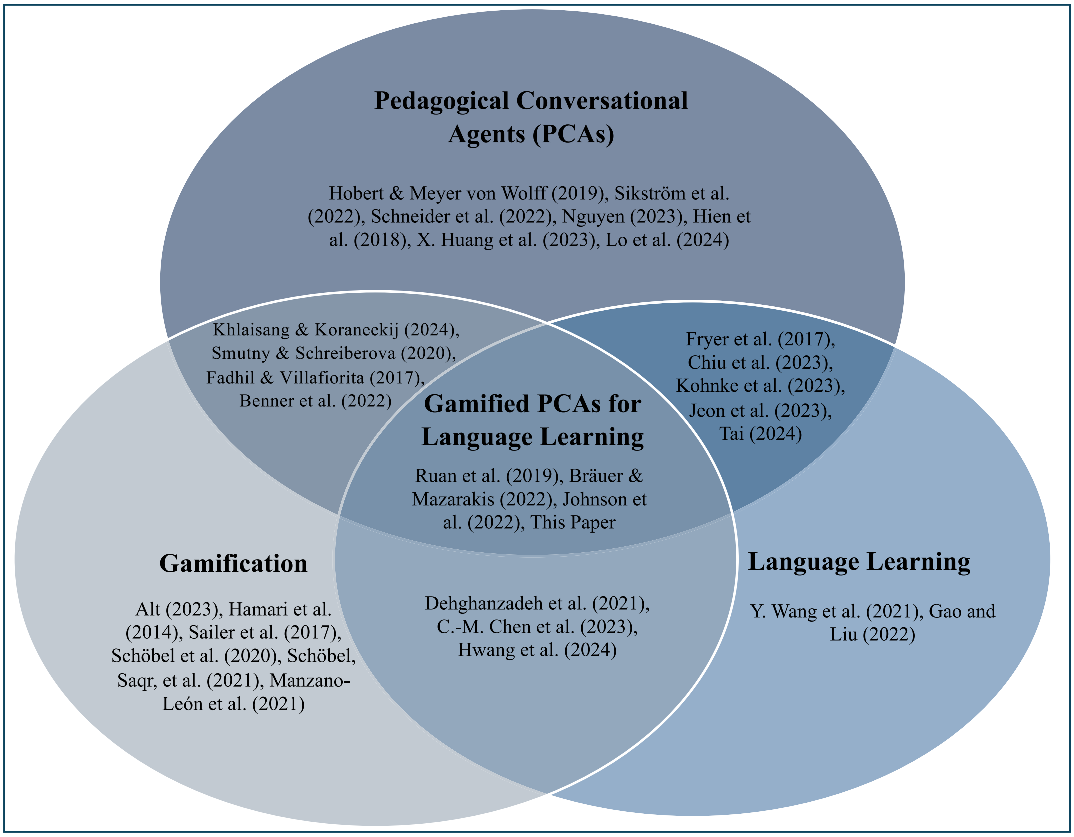

Fig. 1. Exemplary related studies and positioning of our paper.

satisfaction (Fadhil & Villafiorita, 2017; Johnson et al., 2022). In addition, a pilot study demonstrates its potential for engagement in massive open online courses (Khlaisang & Koraneekij, 2024). Related studies indicate positive effects of gamification for PCAs, but little is known about their intersection in language education. Initial findings show the effectiveness of gamified PCAs, with Ruan et al. (2019) noting positive effects on vocabulary learning and Bräuer and Mazarakis (2022) reporting long-term benefits for German language learning with Amazon Skill. Since learners interact with PCAs in open dialogues, this setting is highly interactive and natural, which is vital for language learning and goes beyond the possibilities of classic tools like "Duolingo" (Hwang et al., 2024). Benner et al. (2022) proposed a gamified PCA framework, advising designers to consider educational (learning goal), social (PCA role), and motivational layers (motivational outcome). Further PCA literature supports this framework and emphasizes that educators and developers should carefully select PCA design elements with these layers (Weber et al., 2021). However, this framework does not yet provide design guidelines or DPs. Related work often focused on single game elements like quizzes or individual reward elements instead of complete gamification concepts (e.g., Smutny & Schreiberova, 2020). This paper addresses this issue by deriving the GNPL framework comprising four DPs for the complex interaction of educational, social, and motivational layers. Fig. 1 visualizes the framing of our study compared to exemplary related work (Appendix A includes a detailed overview). The following section examines the main research gaps and questions arising from this literature review.

1.3. Research gaps and research question

The literature review revealed the following implications and research gaps related to this work.

First, employing gamified PCAs is complex. It requires careful design decisions regarding the social user-PCA interaction, gamification's motivational impact, and language learning's educational goals (Benner et al., 2022; Weber et al., 2021). As this intersection has rarely been considered, it is unclear how gamification effectively enables PCAs to address the motivational challenges of language education (B. L. Schroeder et al., 2020). Second, there are only few theoretically-grounded design recommendations on gamification for PCAs (Wittmann & Morschheuser, 2022). Initial approaches to gamified PCA design recommendations either focus only on the technical choice of architecture (Griol & Callejas, 2016), remain at a meta-level by describing overarching design layers instead of functional requirements (Benner et al., 2022), or marginally detail gamification design (Winkler et al., 2020). The community emphasizes the need for theoretically sound design recommendations about AI systems like PCAs (Díaz & Nussbaum, 2024; Weber et al., 2021). Analyses of practical applications such as PCAs for Facebook Messenger show that gamification has mainly been implemented sporadically so far, with developers only implementing individual game elements instead of complete gamification concepts (Smutny & Schreiberova, 2020). A recent literature review (Khosrawi-Rad et al., 2023a) also indicates that existing publications mainly implement only a few game elements without justifying their selection or providing concrete implementation details. This finding reveals a lack of systematic approaches that link meaningful design options to sound theories – a gap that the present paper addresses by offering theory-driven design principles (DPs) and concrete implementation guidance that build on our earlier work. Regarding the motivational effect of PCA design elements, Weber et al. (2021, p. 12) moreover emphasize: "Further research is needed to establish a theoretical link between the design elements and the outcome variables." Influential studies on AI in education focus on technological innovations rather than educational theories, signifying the gap in theory-driven approaches (X. Chen et al., 2020). While there are design frameworks for gamified learning management systems (e.g., Liu et al., 2017; Schöbel et al., 2020), there is little knowledge about gamification design for PCAs. As the effects of gamification vary in different contexts (Hamari et al., 2014), we provide new knowledge on how gamification should be implemented in language learning PCAs. Design knowledge in the form of DPs helps to solve the challenges of a meaningful approach to gamification in PCAs. These comprise "prescriptive statements that indicate how to do something to achieve a goal" (Gregor et al., 2020, p. 1622). These statements contain recommendations for developing IT systems that serve as abstract and reusable rules (Gregor et al., 2020). Thus, they offer validated guidance on how to design gamified PCAs that promote learners' motivation. This validated guidance helps educators and designers to choose from a wide range of gamification design options. Third, even though many researchers call for gamification of PCAs (Ciuchita et al., 2023; Wittmann & Morschheuser, 2022), few studies show whether gamification actually has a positive outcome for PCAs and how positive effects can be promoted. In digital learning, game elements yield mixed results, including positive motivational benefits (Alt, 2023; C. -M. Chen et al., 2023) and negative outcomes like frustration that affect learners' well-being (Melo et al., 2020; Sailer et al., 2017). Research analyzing game elements' effects is the foundation for examining meaningful gamification due to the risks of inappropriate design (Sailer et al., 2017; Seaborn & Fels, 2015). This includes research that uses gamification in new contexts (Hamari et al., 2014), such as in the case of PCAs in a setting with high interactivity and the influence of the PCA's social role (Wittmann & Morschheuser, 2022). Given the interactive nature of PCAs, this approach means to design the social role in a way that highlights gamification's advantages (ibid.). Similar to the role of non-player characters in classic games, this includes integrating feedback or guiding learners with hints in quizzes or learning stories (Ho & Ng, 2022). For example, frustration from learning could be mitigated by the PCA's social role through mimicking a learning partner through personal feedback when learners solve challenging quests or by narrating interactive game stories. The experimental investigation of DPs helps to determine whether and how the game elements should be linked to the social role of the PCA. By applying gamification as an established approach, we contribute to an advanced understanding of how to effectively design PCAs in educational practice. We address the following **research question**: *How should gamified PCAs be designed to motivate learners?*

We derive four DPs for motivating PCA design based on interviews and literature, which we summarize in a cohesive framework (the "GNPL" framework), implement in a text-based chatbot and evaluate in gamified English learning in higher education. This approach demonstrates the design of a gamified PCA for language learning as our main application case, but it could also be applied to other educational fields that face comparable motivational challenges. It shows how to overcome motivational barriers using AI and the PCA's social role. This approach thus extends existing applications like "Duolingo." It addresses broad implications for the effective use of educational technology by contributing new design knowledge and empirical insights to the theory streams of PCAs, gamification, and technology-mediated learning in language education.

The remainder of the paper is structured as follows. Chapter 2 outlines the methodological approach and our individual research steps. Chapter 3 presents the core results, including the problem statement, the derivation of the GNPL framework and its DPs for gamified PCAs, and the findings of two iterative development and evaluation steps. Chapter 4 discusses studies' results, and Chapter 5 concludes the paper with the implications of this research, the limitations, and directions for future research.

2. Methodology

To methodically position our work, we first introduce our design-oriented approach, rooted in design science research as the overarching research paradigm, and then describe how we proceeded in two iterative design cycles.

2.1. Design science research procedure

We follow the design-oriented research paradigm. This paradigm is a common method in pedagogical research, characterized by iterative cycles of developing, analyzing, and refining educational practices and tools (Euler, 2014; F. Wang & Hannafin, 2005). We classify our design science research project within Gregor and Hevner's (2013) knowledge contribution framework, which serves to define the nature and impact of design-oriented research contributions. It distinguishes contributions based on solution and application domain maturity: routine design (known solutions to known problems, no contribution), improvement (new solutions to known problems), invention (new solutions to new problems), and exaptation (extending known solutions to new problems). We apply the exaptation procedure, combining the established gamification approach with PCAs to address motivational challenges in language learning.

In addition to the definition of our contributions according to Gregor and Hevner's (2013) framework, we applied a methodological framework to systematically plan and execute the process steps of our design science research project. Therefore, we follow Kuechler and Vaishnavi's (2008) framework for designing innovative artifacts while ensuring rigor and practical relevance. We applied this process-oriented framework since it offers a step-by-step outline for design science research projects and is recognized by the community (Gregor & Hevner, 2013). We have conducted two design cycles (Fig. 2), which we will explain next. In design cycle 1, we conceptualized a gamified PCA (Khosrawi-Rad et al., 2023b).[1] In design cycle 2, we developed and evaluated a mature gamified PCA.

2.2. Design cycle 1

We interviewed eight learners to assess the learners' initial requirements and wishes for gamified PCA design (four males and four females, mean age = 26). Following Guest et al. (2006), we conducted interviews until we reached saturation after eight interviews. This approach ensured that the sample size met the qualitative research criteria. All participants are university students and belong to our target group of higher education learners. The participants were enrolled in business and technical degrees such as business information systems. We selected the subjects from a pool of learners from an overarching research project on motivating learners in vocational education with a focus on learning parallel to a job. We chose these participants because they faced the challenge of language learning in addition to their jobs or university studies (with up to 37 working hours per week). The participants practice English in their free time (e.g., with applications such as Duolingo). We selected this target group because they have difficulties staying motivated in language learning due to the combined burdens of studying, working, and learning in their free time. These challenges include balancing professional demands, coursework, and personal commitments. The participants had at least a B1 level of English proficiency and focused on maintaining and improving their language skills. Most participants had previous experience with chatbots (e.g., Amazon customer support) but had little experience with language learning PCAs. Overall, the participants are familiar with digital tools due to their predominantly technical study background and young age. The combination of language learners who are working students with limited free time and prior chatbot familiarity highlights their suitability as a sample for providing suggestions on motivational PCA design. We used a semi-structured interview guide on learners' study challenges and suggestions for gamified PCAs. We asked them which PCA functions and behaviors would motivate or demotivate them, their suggestions for the role design of the PCA, and their preferred design of the game elements for PCAs according to Schöbel et al.'s (2020) taxonomy. The interview guide captured both the problem and solution space (vom Brocke et al., 2020). Hence, the interviewees provided insights into their desired motivational PCA learning experience. This procedure allowed us to design the PCA to meet our target group's needs and requirements. The procedure, moreover, ensured user-centered design and reduced researcher bias in artifact development. We analyzed the interviews following Mayring (2015) using deductive coding based on a coding guide and adding further sub-categories inductively. We formulated user stories and assigned them to the game elements. Following Möller et al. (2020), we synthesized these with supporting literature on gamification and education to meta-requirements (MRs) as the basis for the DPs and the GNPL framework. The literature review included kernel theories from motivation, education, and human-computer interaction, as well as empirical work that

[1] The conference paper by Khosrawi-Rad et al. (2023b), published in Mindtrek '23 Proceedings, CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/), presented an earlier version of this research project. The current article expands the earlier version through a more extensive theoretical grounding, a focus on DPs for self-regulated language education, adapted DP formulations, a revised data analysis for design cycle 1, and the addition of design cycle 2.

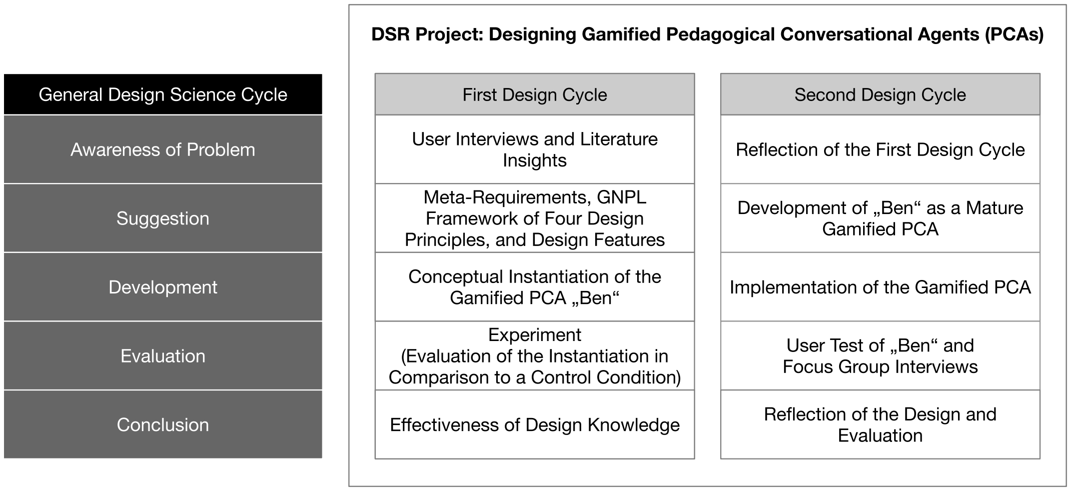

Fig. 2. Design science research procedure.

applied these theories in the gamification and PCA context. Specifically, we incorporated motivational theories, such as goal-setting theory (Latham et al., 2011; Locke & Latham, 1990) and self-determination theory (Ryan & Deci, 2000). We drew on social-cognitive perspectives (Bandura, 1977; Vygotsky, 1978) to rationalize the design of supportive, peer-like interactions. We applied the interactive, constructive, active, and passive framework (Chi & Wylie, 2014) to foster deeper engagement. Bloom's (1956) taxonomy informed us to align the complexity of PCA-guided tasks with learners' progress. We then formulated design features (DFs) based on the DPs and created a mapping diagram of MRs, DPs, and DFs. Each DP specifies the PCA role and highlights gamification's unique use in conjunction with the PCA's social role. Following design science research, we instantiated and evaluated our design knowledge (Kuechler & Vaishnavi, 2008). A conceptual prototype was built with the "Figma" software. We evaluated it in our first study with 76 higher education learners in Germany (a detailed description of the procedure and constructs used follows in Section 3.3). The evaluation constituted a within-subjects experiment in which learners evaluated the prototype with DPs compared to a control condition without DPs. To determine the appropriate sample size, we conducted a statistical power analysis. Following Baroudi and Orlikowski (1989), we aimed for a statistical power of 0.8. The analysis revealed a minimum sample size of 64 (input variables $\alpha = 0.05$; mean effect size $= 0.35$). All participants navigate the challenges of learning English as a second language—most of them major in business or technology-oriented courses like business information systems. We evaluated how they perceive our artifact compared to a control condition that does not satisfy the DPs. The evaluation of design cycle 1 served to test the DPs' effectiveness in a controlled setting. The evaluation results revealed a significantly higher perception of the prototype with DPs compared to the control condition for all analyzed constructs. Thus, the first design cycle demonstrated the DPs' effectiveness. Building on these findings, we proceeded in design cycle 2 by transforming the conceptual prototype into a fully functional prototype and evaluating it through real user interactions.

2.3. Design cycle 2

We followed up on the conceptual prototype of our first design cycle by implementing the gamified PCA "Ben" as a mature software artifact. While design cycle 1 had already proven the effectiveness of our GNPL framework in a controlled setting, the aim of design cycle 2 was to identify how a fully functional gamified PCA designed with the GNPL framework impacts learners' motivation. Design cycle 2 included the selection of suitable technical tools to implement the functional requirements. Specifically, we utilized the conversational AI framework "Rasa" to handle natural language understanding and dialog management. We integrated the grammar-checking tool "LanguageTool" to ensure linguistic accuracy and employed the messenger "Slack" as the communication platform for user interactions. We designed the PCA very similarly to the conceptual prototype from design cycle 1 to allow for comparability of our first study and final evaluation. In addition, we integrated improvement suggestions from participants in design cycle 1, which included minor adjustments, such as the selection of emojis or the wording of specific chatbot responses. In design cycle 2, we demonstrate how designers can implement the GNPL framework and the associated DFs. We finally evaluated our software artifact with 70 participants, again all navigating learning English as a foreign language and most majoring in business and technology-oriented degree programs. The users interacted with the PCA and then completed a questionnaire to assess the PCA. The questionnaire contained established quantitative scales for measuring PCA perceptions with a special focus on motivational constructs. In addition, we asked open-ended questions regarding the positive and negative aspects of the interaction (a detailed description of the procedure and the constructs used follows in Subsection 3.4.3). Moreover, we conducted focus group interviews with 17 voluntary participants from this group. As for design cycle 1, we ended the data collection of the focus groups after achieving content saturation (Guest et al., 2006).

3. Design and evaluation of the gamified PCA

In this chapter, we first introduce the problem statement (Section 3.1), then derive the GNPL framework and its DPs (Section 3.2), and finally present two iterative evaluations of the gamified PCA (Sections 3.3 and 3.4).

3.1. Problem statement

Our design science research project's core goal is to promote motivation. The interviewees expressed the desire to be motivated by a PCA since they find it difficult to continuously motivate themselves, for example: "*I think that a study companion should be able to motivate me. Studying means a lot of self-organization, and of course, that also takes a bit of the stress out of having everything in your head. You can rely on someone*." Motivational support is crucial in language learning since it is complex, requires many repetitions, and is often perceived as unpleasant (Derakhshan et al., 2021; Y. Wang et al., 2021). The participants described learning as a burden: *"I have a basic motivation problem. Like probably so many others. You think to yourself: Wow, I could do better things with my time now and do something else."* We aim to shift this perception towards enjoyment in language learning (Botes et al., 2022; Y. Wang et al., 2021). Motivational barriers are accompanied by other challenges that the gamified PCA should address. The students wished to track their learning progress to feel unburdened and focus on the learning itself rather than the learning organization. They expressed: "*It's simply better if the system knows where I am in the learning process, what I've already done, and what I haven't done. I don't have to memorize it myself, so I can focus on studying*." In addition, learning should be designed interactively (Chi & Wylie, 2014), as many learning materials at educational institutions, such as text materials or videos, are often not interactive (Haleem et al., 2022). This requirement is crucial in language learning, which needs immediate feedback and active engagement (Dörnyei, 2014; Gao & Liu, 2022). Learners noted that lectures are often not adapted to their individual level and pace because of the high number of participants. As a result, self-regulated learning becomes more difficult (Behr et al., 2021). Having established the motivational challenges that learners face, in the following we introduce the GNPL framework that helps educators and developers design meaningful gamified PCAs that address these issues.

3.2. Derivation of the GNPL framework

We derive DPs and refer to the related MRs and DFs, using supporting literature from human-computer interaction, motivation, and education. This procedure ensures a multi-layered view of gamified PCA design, as called for in the literature (Benner et al., 2022; X. Chen et al., 2020). This interplay of abstract MRs from theory and practice, generalized DPs, and their conversion into easily instantiable DFs constitutes an overarching framework for designing gamified PCAs. We have named this framework after the initial letters of the four DPs detailed below, which represent the core contribution of our work. We refer to it as the "GNPL framework" to give the DPs a cohesive conceptual identity: G – Goal-Setting and Reflection, N – Novice-Expert Relationship, P – Performance-Related Motivation, and L – Learning Story Narration. For each DP, we present the unique features of the design of the gamified PCA elements in combination with the social role of the PCA. We would like to emphasize that this combination differs from classic gamification applications without a PCA in that the game elements are combined with social learning, a need highlighted in the gamification literature (Melo et al., 2020).

Design Principle 1 (Goal-Setting and Reflection): The interviewees desired to set their own learning goals to incorporate personal interests and find them valuable (Scheu & Benke, 2022) (MR1). According to goal-setting theory, setting specific and challenging goals leads to learners being able to learn in a self-regulated manner and accept their goals (Latham et al., 2011; Locke & Latham, 1990). Learners, thus, influence their learning process, which causes greater motivation (Boekaerts, 1999; Lunenburg, 2011). Learners desired the PCA to suggest learning goals if they had difficulties defining an appropriate learning goal (MR2). For example, one participant said: "*It would be useful if the PCA already provides different, pre-defined goals to choose from. For example, you could choose an option that says, 'I'd like to be able to talk to locals on my vacation'*." Suggesting goals helps students who struggle to formulate goals and overarching visions due to a lack of self-regulation skills described above (Wong et al., 2021). Many learners set unrealistic goals for themselves, diminishing their motivation (Hew et al., 2023; Schlimbach, Windolf, & Robra-Bissantz, 2023). The students suggested setting learning goals at multiple levels (long-term to short-term) to see their progress (MR3). In that way, the gamified PCA could reward learners for short-term success and working toward long-term goals, promoting motivation (Krath et al., 2021; Ryan & Deci, 2000). One participant addressed: "*You should earn points not only for completing tasks but also for reaching weekly goals or learning units as more substantial milestones. You don't just get the short-term motivation like, okay, now I've completed another task, and now I've completed another task. Instead, the learning companion suggests that there are long-term or medium-term goals, so you are more motivated*." To further support self-regulated learning, the PCA can reflect on students' progress and adjust the goals if desired (Hew et al., 2023) (MR4). The opportunities to reflect on goals with the PCA enable learners to plan their learning and learn from mistakes (Zimmerman, 2013), positively influencing their confidence and motivation (Keller, 1987). This requirement is vital as Kim et al. (2006) demonstrate that a PCA acting as a personal learning facilitator like a human learning partner can enhance motivation. This finding is grounded in the theoretical perspectives of Bandura (1977) and Vygotsky (1978) who emphasize the importance of support from a learning partner. The requirement aligns with the interactive, constructive, active, and passive framework (Chi & Wylie, 2014). This framework stresses that interactive and constructive learning activities, such as reflecting on and adjusting goals, enhance cognitive engagement. Designers can implement the MRs by letting learners set their own goals during onboarding (DF1). For learning goal recommendations, the students desired the PCA to help learners define "SMART" goals (DF2). SMART is an acronym which states that goal definitions should be specific, measurable, attainable, realistic, and time-related (Doran, 1981). Setting clear goals for learners can increase the likelihood of being satisfied with their learning process and completing the course (Hew et al., 2023; Schlimbach, Windolf, & Robra-Bissantz, 2023). To implement goals with different time frames, the PCA can first formulate an overarching vision as a long-term goal and divide this into medium- and short-term milestones (DF3). After individual study units, the PCA can reflect on goal fulfillment (DF4). Overall, DP1 emphasizes the role of the PCA as a mentor in goal setting and reflection, like a human learning companion. Furthermore, DP1 lays the groundwork for rewarding learners, as emphasized in subsequent DPs of the GNPL framework. By aligning rewards with personally set goals, the PCA ensures that these are meaningful and motivational (Chi & Wylie, 2014; Kim et al., 2006). The PCAs' mentor role is unique to gamified PCAs compared to traditional gamification in non-PCA contexts. In the PCA context, feedback on goals and subsequent rewards are not only personalized but also accompanied by social communication of a mentor. Thus, learning becomes a social process, with learners feeling supported by a social actor rather than left alone with their challenges (Bandura, 1977). Consequently, learners experience the feeling of social recognition, which makes the learning process feel more collaborative and engaging (Sailer et al., 2017). Hence, we derive the following proposition for *DP1: Design the PCA with the ability to help learners set their own goals and reflect with them on their progress to promote self-regulated learning.*

Design Principle 2 (Novice-Expert Relationship): DP2 focuses on learners acquiring and applying knowledge in interaction with the PCA as a more experienced classmate. The interviewees preferred the PCA to act as an eye-level classmate that is knowledgeable at the same time so that they perceive it as helpful and trustworthy. The finding coincides with evidence from further empirical studies (Kim & Baylor, 2016; Schlimbach & Zhu, 2023). Moreover, a literature review emphasized PCAs' effectiveness when they balance friendly, peer-like interactions with the ability to deliver substantial language support (W. Huang et al., 2022) (MR5). Studies show that the classmate role is beneficial, as the PCA exhibits a low power distance (ibid.). However, this is specific to our target group (foreign language learners in Germany), as German students have a stronger cultural understanding of low power distance than students from other cultures, like Chinese students (Schlimbach & Zhu, 2023). DP2 emphasizes a nuanced view of the PCA's classmate role and suggests it performs as a more experienced student to exhibit a competent role model (Liew et al., 2013). The PCA acts as a friendly expert who supports a novice (the user) by expressing its knowledge through its role, thus fostering self-efficacy (Kim & Baylor, 2016). The PCA shapes this role through targeted didactic support in language learning that students would rather expect from a more experienced learning partner (ibid.). To express this role, students suggested that the PCA imparts knowledge at the beginning of the learning units and queries them afterward if they understood it. The interviewees wished that the difficulty of learning tasks increases with learners' progress to cover higher learning goals like applying knowledge according to Bloom's taxonomy (Anderson & Krathwohl, 2001; Bloom, 1956) (MR6). The tasks should therefore include low-level learning (e.g., vocabulary training) and higher-level learning goals (e.g., open tasks) (Chamot & O'Malley, 1994). To avoid overwhelming learners, the PCA can assist in mastering tasks (e.g., Winkler et al., 2020) (MR7). Through scaffolding, learners achieve higher learning outcomes when they receive hints that help them to engage cognitively and figure out the right solution (Janson, Söllner, & Leimeister, 2020; Vygotsky, 1978). Continuous practice with the PCA as a more experienced classmate can lead to self-efficacy, i.e., that learners exercise to accomplish tasks (Bandura, 1977; van Dinther et al., 2011). The PCA should remind learners to repeat the quiz later to test their knowledge (Pham et al., 2018) (MR8). The interactive nature of learning with the PCA aligns with the interactive, constructive, active, and passive framework, emphasizing interactive and constructive activities for deeper cognitive engagement (Chi & Wylie, 2014). To implement DP2, the users wished the PCA to provide insider knowledge in its role as a more experienced classmate (DF5). For instance, one participant said: *"I think of the PCA a bit like somebody who has already experienced the same thing as me. Someone who can understand my situation and can be empathetic. Someone who can give me some hints like what to focus on.*" In this role, the PCA should deliver content and prepare it in short summaries for studying in microlearning units (DF6). Quizzes at the end of learning units are essential to review what students have learned and cover low-level learning goals (Anderson & Krathwohl, 2001; Bloom, 1956) (DF7); for example, "*Once I have completed a lesson, I should always have the option of a quiz* via *the learning companion to complete it together. I should be able to do the quiz from the basics eight months ago once again, receive a summary of the content from the learning companion, and take another quiz*." More complex, open-ended tasks or use cases help to apply the knowledge (ibid.) (DF8). The PCA should constantly repeat the quiz so learners can study on the go (Yin et al., 2021) (DF9). Overall, DP2 emphasizes the social role of a knowledgeable classmate that offers support and guidance. This social interaction is a key differentiator from traditional gamification in non-PCA contexts, where feedback and rewards are mostly standardized (Schöbel, Schmidt-Kraepelin, et al., 2021). By receiving support in learning tasks from a knowledgeable classmate, learners feel socially recognized and supported, which enhances their sense of belonging with a learning partner and motivation (Ryan & Deci, 2000). By facilitating both simple and complex tasks interactively, the PCA promotes self-efficacy and motivation through social interaction (Kim et al., 2006; Ryan & Deci, 2000). Hence, we derive the following proposition for *DP2: Design the PCA as a more experienced classmate that engages with learners in both simple and complex learning tasks so that they practice for different levels of learning goals to gain and apply new knowledge.*

Design Principle 3 (Performance-Related Motivation): The interviewees considered it essential that the PCA rewards them for positive performance (MR9). Research has proven that rewards promote perceptions of competency awareness and greater motivation according to self-determination theory (Ryan & Deci, 2000; Sailer et al., 2017). The PCA's role as a social actor is crucial, as the PCA not only provides rewards but also offers personalized encouragement and feedback. These functions support the social interaction and foster motivation (Bandura, 1997; Vygotsky, 1978). Rewards should correspond to the users' performance to support motivation (Abramovich et al., 2013). Therefore, the PCA should award reward elements related to the effort (MR10). The rewards should be adapted to short-, medium-, or long-term goal achievements to link them to the previously set learning goals (DP1) according to goal-setting theory (Latham et al., 2011; Locke & Latham, 1990). Aligning game elements with learning goals leads to a meaningful gamification concept, i.e., learners perceive game elements as relevant and valuable (Abramovich et al., 2013). For instance, one interviewee said: "*I think you should get quite a few, but constant points for doing certain tasks. And then even higher rewards when completing weekly or monthly goals or certain learning units. So, you get bigger rewards for longer-term goals.*" Interview and literature findings revealed that learners desire to visualize their progress to motivate themselves (Sailer et al., 2017) (MR11). The visual representation of learners' performance via game elements like progress bars or levels can help them stay motivated and appreciate their efforts: "*Well, it is the progress that I need to see to know what I am getting out of it. When I can see that I have already achieved a certain percentage and might only be missing another 40 %, I am more motivated to keep saying: 'Okay, I will finish it now'.*" By providing performance-based feedback, the PCA promotes motivation for deeper cognitive activities (Chi & Wylie, 2014). Designers can implement DP3 by having the PCA award points regularly as straightforward game elements for short-term successes (e.g., correct answers to quizzes) (DF10), rewards like stars for medium-term milestones (DF11), and badges for remarkable achievements (DF12). The PCA can include progress bars (DF13) and levels (DF14) to visualize learning progress, showing learners how close they are to their goals. Overall, DP3 emphasizes the social role of the PCA as a motivator, which motivates learners through rewards and personalized feedback. The role of the PCA as a motivator distinguishes the implementation of gamification in PCAs from that in traditional learning applications. As for why, rewards are not standardized but are handed over by a PCA in the role of a learning partner. This function gives the rewards a social meaning. This social recognition promotes students' motivation to learn (Ryan & Deci, 2000). Hence, we derive the following proposition for *DP3: Design the PCA to* award *performance-related game elements and visualize their learning progress to keep them motivated in the long run.*

Design Principle 4 (Learning Story Narration): DP4 focuses on the role of the PCA as narrator. The participants recommended the PCA introduce a practice-relevant learning story in which learners solve tasks interactively (Forsyth et al., 2020) (MR12). By working on an interactive learning story, learners can practice their problem-solving skills (ibid.). It is crucial that the story relates to practice for learners to perceive the story to be relevant and engage with it in constructive and interactive learning (Chi & Wylie, 2014). They emphasized that the PCA should playfully introduce practical situations by fulfilling its narrator role through engaging storytelling: "*These can be very specific fictional stories in which you are perhaps supposed to act as a character in a certain situation. There should always be something meaningful in the story so that the learners can empathize with the situation. The PCA could say, 'Imagine you are in this situation …* '." This open communication through the PCA distinguishes the suggested design from related gamification applications such as "Duolingo," as the narratives can be made more lively and interactive using AI (Schöbel, Schmidt-Kraepelin, et al., 2021; Söbke & Streicher, 2016). The interviewees considered it important that the PCA gives the learners hints during the story to foster their self-efficacy and motivation to complete it (Bandura, 1977). These hints can be provided by AI-supported verification of the language input to provide targeted feedback (X. Huang et al., 2023; Kohnke et al., 2023). They highlighted that the PCA should merely introduce

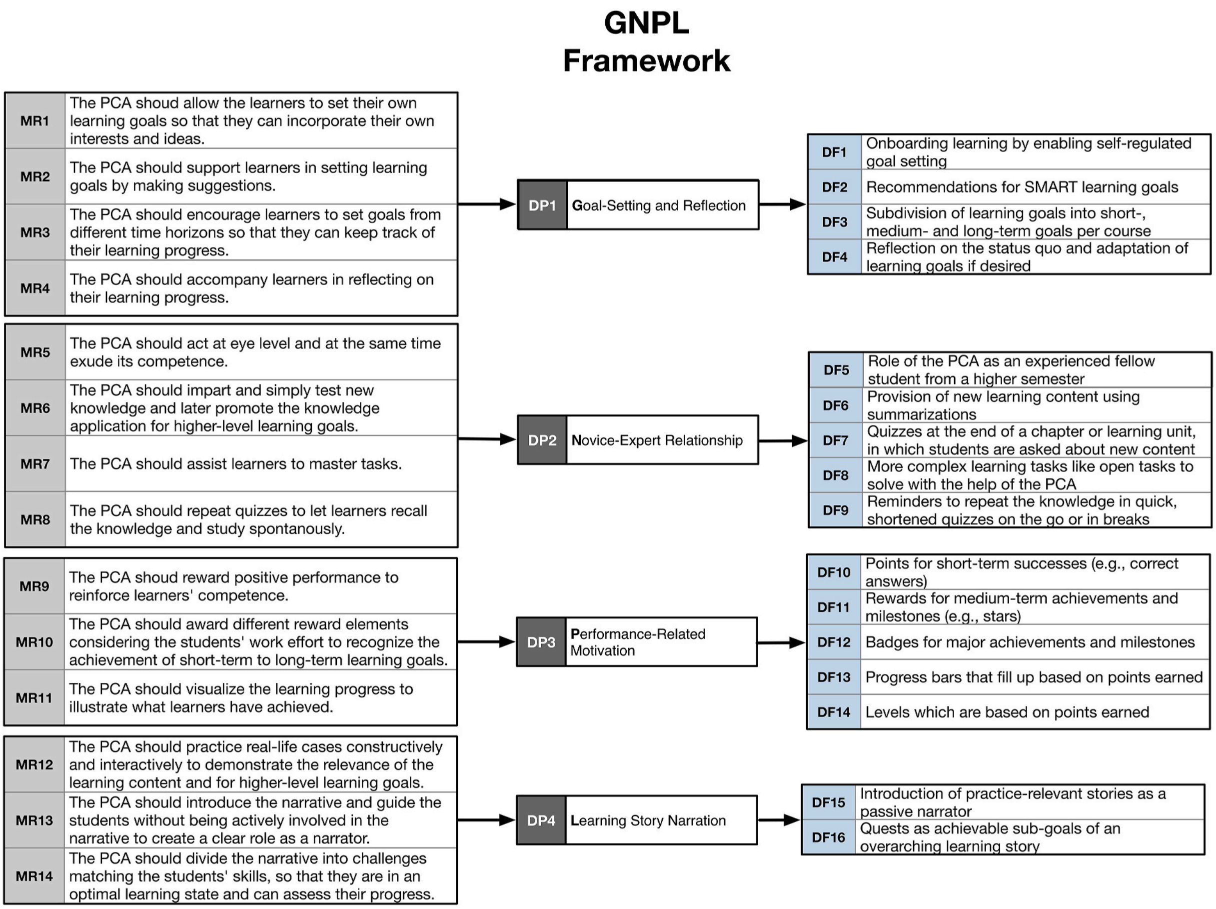

Fig. 3. Mapping diagram of the GNPL framework.

and accompany the story without being actively involved as a player (MR13). They justified this statement by explaining that they would consider it contradictory if the PCA participated in the story while also checking the correctness of the answers. For example, users may perceive the behavior of the PCA as untrustworthy if it pretends to be unaware of the story's content, for instance, "*Afterward, this system must somehow verify what I have just said. And if I explain some nonsense and it still accepts it unreflectively, then that does not help me. It feels strange to me to explain something to a person or program it already knows anyway.*" The learning task should be divided into subgoals corresponding to the learners' abilities (MR14). Subgoals enable students to continuously perceive successes and work out the story step by step (Tondello et al., 2018). By working on sub-goals step by step, learners experience the feeling of continuous success (Krath et al., 2021). Matching the learning story to the learners' abilities ensures they are neither under nor over-challenged (Locke & Latham, 1990). Designers can implement DP4 by having the PCA introduce and accompany a practice-relevant learning story, such as a case study, without being directly involved (DF15). Different quests confront learners with feasible individual tasks (Krath et al., 2021) (DF16). Overall, DP4 emphasizes the PCA in the role of a narrator, that promotes interactive learning. This narrator role differentiates the realization of the game stories from classic gamification, as they become more tangible. Hence, we derive the following proposition for *DP4: Design the PCA as a narrator of learning stories, which are divided into quests of varying difficulty to promote interactive learning.*

Fig. 3 summarizes the MRs, the four DPs of the GNPL framework, and DFs in a mapping diagram following Möller et al. (2020). The mapping diagram displays on the left side the MRs that capture learners' core needs and desires, for example, setting personal goals or receiving feedback. On the right, we show the DFs, which include classic game elements (e.g., points, rewards, goal-setting, and storytelling). The GNPL framework consisting of four DPs in the middle bridge these two perspectives by showing how to design gamification in PCAs so that key motivational needs are addressed. These DPs specify how to make the PCA interaction motivating, socially supportive, and pedagogically meaningful. Accordingly, the figure demonstrates how gamification, the process of enriching systems with game elements (Deterding et al., 2011) merges the social role of PCAs with pedagogical requirements, resulting in DPs that can generalize across multiple contexts. For instance, DP1 (Goal-Setting and Reflection) shows how to implement personal SMART goal-setting as a playful element, with the PCA acting as a mentor who provides feedback and prompts reflection. DP2 (Novice-Expert Relationship) integrates quizzes and scaffolding mechanisms, where the PCA functions as an experienced peer by applying challenge-based gamification in a social context. In DP3 (Performance-Related Motivation), typical reward systems (points, stars, badges, level-ups) make learning progress visible and rewarding, while the PCA offers personalized encouragement. DP4 (Learning Story Narration) connects narrative quests to practical learning contexts, placing learners within a continuous story supported by playful interim goals (quests) and thematic challenges. By intertwining game elements at the feature level (such as rewards, narration, and challenges) with a socially engaging PCA interaction and tailored learning tasks, the system enhances learners' motivation, thereby increasing student engagement and improving user experience throughout the learning journey. This approach creates comprehensive gamification concepts that go beyond simply adding points or badges by fully harnessing the strengths of PCAs.

3.3. First iteration: design and evaluation of the conceptual prototype

3.3.1. Description of the instantiation

In our first iteration, we instantiated the gamified PCA Ben as a conceptual chatbot prototype that supports students in learning English. The PCA Ben operationalizes the GNPL framework. The PCA features tasks designed to accomplish low-level goals. These include single- and multiple-choice quizzes focusing on vocabulary or grammatical tenses. The PCA furthermore contains open tasks in which learners apply the knowledge they have learned and formulate sentences in certain tenses on their own. The PCA includes interactive learning stories like "The Roman Colosseum." In these learning stories, learners apply the knowledge interactively and playfully, for example, by completing quests like booking a taxi to the Colosseum or purchasing discounted tickets for entry. Moreover, we have created a control condition of a baseline gamified PCA designed according to the current industry standard. It covers a basic gamified PCA design but does not fully fulfill the MRs of the GNPL framework. Our evaluation strategy in the first iteration aimed to initially test whether learners perceive a PCA in line with the GNPL framework better than the control condition.

3.3.2. Manipulation strategy

In the following, we will explain our manipulation strategy with respect to the GNPL framework. For DP1, learners can formulate learning goals independently with Ben's support and visually track their progress toward the goal (Latham et al., 2011; Locke & Latham, 1990). Students define their vision statement as a playful component of learning, which they break down into sub-goals with the PCA's help. In the control condition, the PCA determines the learning objectives in advance, so students have no influence on them and can only keep track of their progress through messages from the PCA but not visually. In line with DP1, the PCA allows learners to influence the learning process and visualizes their progress, unlike the control condition. For DP2, the PCA behaves like a more experienced classmate with a low power distance who has already completed the exam. Ben can provide hints, for instance, regarding exam preparation. The PCA functions as a tutor with high power distance in the control condition and is unable to provide hints. In line with DP2, the PCA Ben represents a novice-expert relationship, which is still eye-level rather than a relationship with a higher-level tutor. In DP3, learners receive different rewards depending on the learning objectives' long-term orientation (from short-to long-term) (Abramovich et al., 2013). They receive points for correct answers to quiz questions, coins for medium-term successes like multiple correct answers in a row, and badges for achievements, such as passing a quiz without errors the first time. In the control condition, the achievements do not align with the time horizon of goal achievement; for instance, the PCA only verbally congratulates successes and does not include points, coins, and badges. In line with DP3, the PCA applies performance-related rewards to ensure the gamification appears meaningful to the learner, unlike the control condition. For DP4, the learning story is subdivided into sub-challenges (quests) in compliance with the DP. The PCA furthermore narrates the learning story in an introductory role (e.g., "Imagine …"). Moreover, the chatbot helps students to solve the tasks when they are stuck or make a mistake. In the control condition, the story does not include quests, and the PCA has a participatory role, so it cannot give hints. The PCA, in line with DP4, thus, accompanies the learning story step by step as a narrator, unlike in the control condition. Fig. 4 shows exemplary screenshots of both prototypes: the control condition on the left and the DP-aligned artifact on the right (all dialogs translated from the original language to English). The screenshots related to the control condition show exemplarily that the PCA sets the learning goals without allowing the learner to influence them, has a basic knowledge assessment without performance-related rewards, and congratulates successes verbally without progress visualization. The DP-compliant variant shows the exemplary implementation of individual DFs, i.e., learners can set self-regulated goals (DF1), receive suggestions for SMART goals as a playful vision statement (DF2), the PCA divides the learning goals into subgoals to challenge learners (DF3), and helps them reflect on their progress (DF4). Learners are rewarded for successes based on their performance, for instance, with points for short-term success (DF10), and visualize their development in progress bars (DF13). Appendix B shows further excerpts of how we instantiated DP1-4, highlighting all DFs.

3.3.3. Experimental procedure and measures

We evaluated the DPs in an online survey with a within-subjects design (duration: 25 min). Following Aguinis and Bradley (2014), we designed the experiment as a vignette-based study in which the learners experienced simulated interactions with both chatbot variants (control condition and with DP compliance) and captured the functionalities of these prototypes. Learners watched demonstrations of both conditions in randomized order. To test the DPs, we formulated testable propositions (Gregor & Jones, 2007). Learners rated both prototypes on the same constructs on a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). We applied constructs from motivation as our project's focus, complemented by constructs from learning and human-computer interaction. With this decision, we align with the current trend in the *Computers & Education* community, which emphasizes that designing gamification artifacts requires consideration not only of traditional motivational constructs but also of pedagogical effects and social factors (e.g., C.-M. Chen et al., 2023; Melo et al., 2020). We will explain the choice of these constructs in the following and reference the related sources in the subsequent sentences. These sources also appear in the results table below (Table 1). We used "competency awareness" (IJsselsteijn et al., 2013) to evaluate how the reward design empowers users' motivational need to feel competent. We assessed the PCA variants' relationship-building potential with "friendship (helpfulness)" (Mendelson & Aboud, 1999) and "trusting beliefs" (Qiu & Benbasat, 2009). "Friendship (helpfulness)" is a construct initially developed for interpersonal relationships (Mendelson & Aboud, 1999), and researchers have applied it to PCA interactions to measure a PCA's willingness to help the users (Paiva et al., 2017; Strohmann et al., 2022). Trusting beliefs encompass a chatbot's competence, benevolence, and integrity, which are vital for social relationships with a chatbot (Komiak & Benbasat, 2006; McKnight et al., 2002; Strohmann et al., 2022). Furthermore, we used "self-efficacy" (Knispel et al., 2021), relevant to self-regulated learning, which can be induced by setting goals and autonomously determining the learning process (Bandura, 1977; van Dinther et al., 2011). "Relevance and goal-orientation" is based on Chen and Chan's (2008) motivational scale and Keller (1987). It measures whether learners perceive learning in line with the defined goals relevant. The motivational construct "value/usefulness" (Ryan, 1982) measures how the respondents perceive added value in learning with the gamified PCA. We selected the constructs according to their fit with our DPs' intended effects, i.e., goal-setting and reflection (DP1), novice-expert relationship (DP2), performance-related motivation (DP3), and learning story narration (DP4). This approach ensures that the constructs capture specific aspects of the PCA's role and functionality. For instance, "goal-setting and reflection" (DP1) aligns with constructs such as "relevance and goal orientation" and "value/usefulness." These constructs measure learners' ability to set meaningful goals and achieve them on their own. The "novice-expert relationship" (DP2) is reflected in constructs such as "friendship (helpfulness)" and "self-efficacy." These constructs evaluate the PCA's effectiveness as a supportive, approachable classmate. For "performance-related motivation" (DP3), "competency awareness" measures how rewards empower learners to feel competent. "Learning story narration" (DP4) ties constructs like "relevance and goal-orientation" or "trusting beliefs." These assess how narrative-based learning with a PCA as a social learning partner enhances learners' perception of task relevance and trust in the PCA's support. Together, these constructs comprehensively evaluate the PCA's ability to address the motivational, relational, and educational outcomes associated with gamified learning. Appendix C presents all constructs and items used in the study.

3.3.4. Results

The experiment involved 76 participants (39 female, 35 male, and two diverse; average age = 26). For the experiment, we recruited non-native English-speaking students from Germany with an intermediate to advanced level of English. They mainly come from technology-oriented and business-related degree programs like business information systems, which underlines their affinity for new technology like PCAs. Cronbach's α was >0.7 for all constructs, and composite reliability (CR) was always >0.7, so our data set met the requirements for data analysis (Hair et al., 2014). Specifically, the constructs achieved the following values: relevance and goal-orientation ($\alpha = 0.94$, CR = 0.96), value/usefulness ($\alpha = 0.98$, CR = 0.98), self-efficacy ($\alpha = 0.96$, CR = 0.97), friendship (helpfulness) ($\alpha = 0.96$, CR = 0.97), trusting beliefs ($\alpha = 0.97$, CR = 0.98), and competency awareness ($\alpha = 0.92$, CR = 0.94).

We first determined the descriptive results, including the mean values (MV), medians (Mdn), and standard deviations (SD). Fig. 5 provides a boxplot visualization of the results, showing that the "with DP condition" exhibited higher **Mdn** for every construct compared to the control condition. To enrich these findings, we also examined the MV (see Table 1 below). The MV for the DP condition all exceed the scale mean and are even >5. For the control condition, all MV are <5 and, in some cases, below the scale mean. The SD across constructs were relatively consistent, indicating similar variability around the MV for both prototypes. For instance, "friendship (helpfulness)" had an MV of 5.54 in the DP condition compared to 2.85 in the control condition. "Trusting beliefs" reached an MV of 5.60 in the DP condition compared to 3.76 in the control condition.

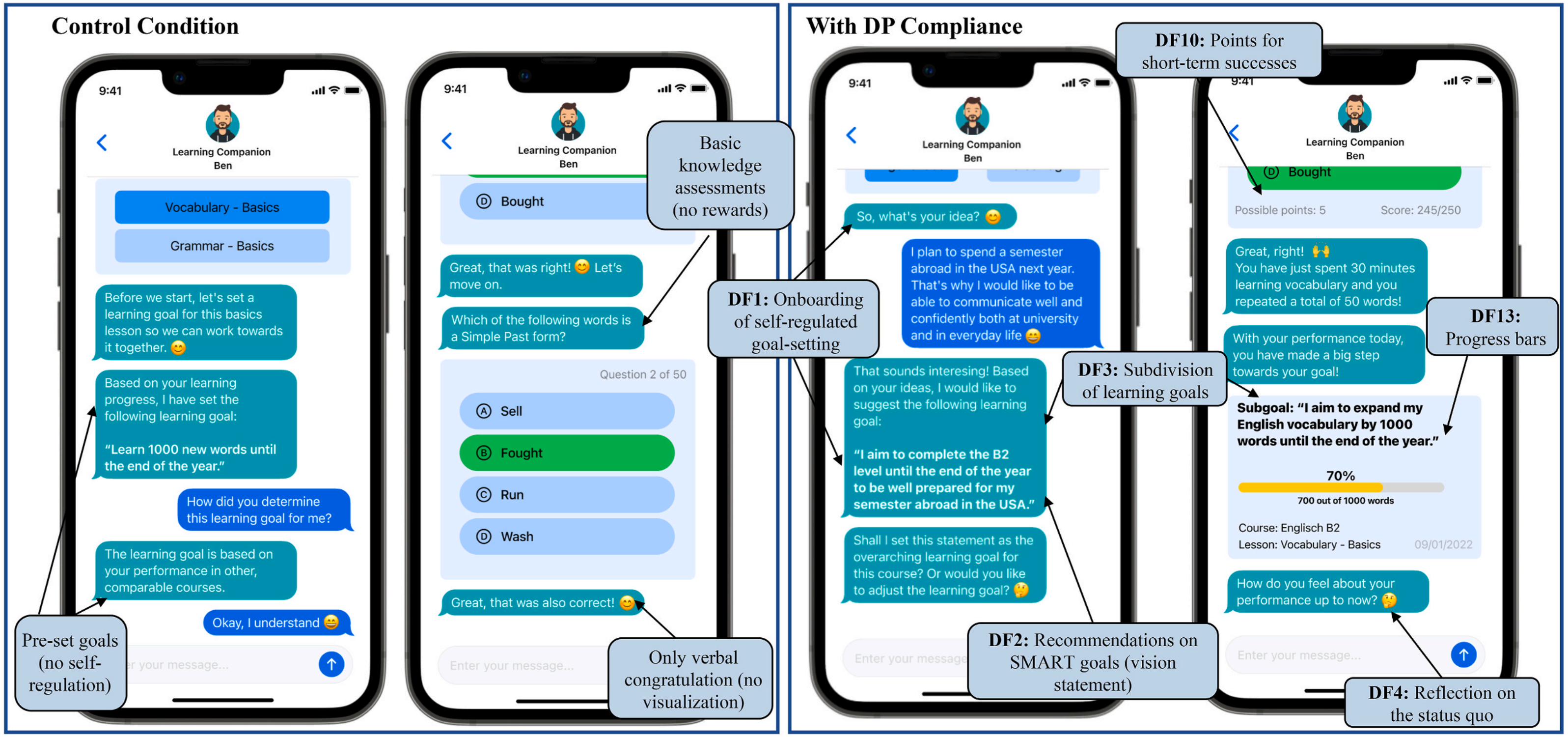

Fig. 4. Exemplary design of our prototype compared to the control condition.

Table 1
Wilcoxon signed-rank test results of the prototype with DPs compared to the control condition.

Constructs	Mdn With DP	MV With DP	SD With DP	Mdn Control	MV Control	SD Control	Z value	p	r
Relevance & Goal-Orientation (Z.-H. Chen & Chan, 2008)	6	5.71	1.48	5	4.71	1.38	5.139	<0.001	0.589
Value/Usefulness (Ryan, 1982)	6	5.81	1.36	5	4.85	1.48	5.306	<0.001	0.609
Self-Efficacy (Knispel et al., 2021)	6	5.06	1.48	4	3.65	1.47	5.945	<0.001	0.682
Friendship (Helpfulness) (Mendelson & Aboud, 1999)	6	5.54	1.31	2	2.85	1.50	6.751	<0.001	0.774
Trusting Beliefs (Qiu & Benbasat, 2009)	6	5.60	1.50	4	3.76	1.51	6.145	<0.001	0.705
Competency Awareness (IJsselsteijn et al., 2013)	5	5.14	1.36	4	4.39	1.35	4.889	<0.001	0.561

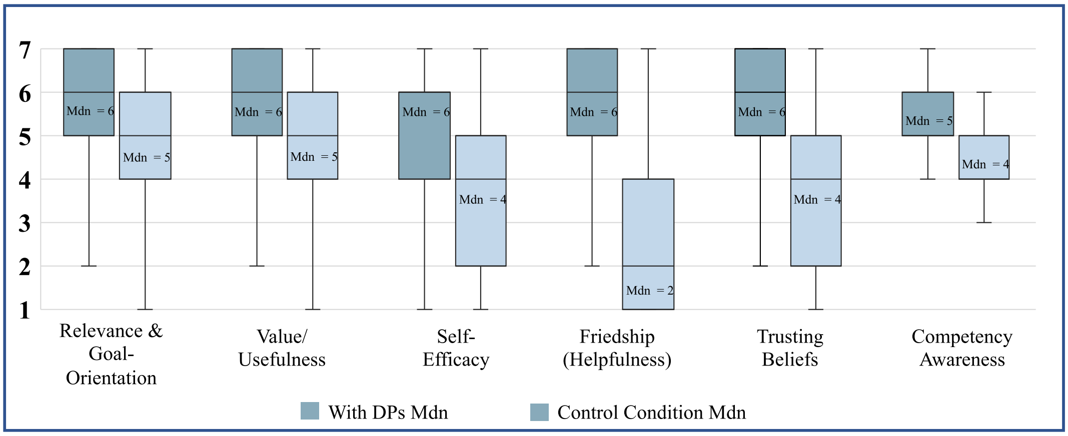

Fig. 5. Boxplot results for the investigated constructs related to the conceptual prototype.

To test the effect of the "with DPs" condition compared to the control condition, we conducted a Wilcoxon signed-rank test for each construct ($\alpha = 0.05$). We confirmed that the assumptions for conducting the Wilcoxon signed-rank test were met (paired and dependent measurements, independence of pairs, and symmetry of differences between paired measurements). Table 1 shows the Wilcoxon signed-rank test results along with the descriptive statistics.

The Wilcoxon signed-rank test results reveal that the prototype with DPs significantly outperformed the control condition across all scales ($p < 0.05$ and even <0.001). The results show clear advantages for the DP condition, with all constructs demonstrating higher Mdn values than the control condition. The effect size analysis shows that this is a strong effect for all constructs ($r > 0.5$). Notably, constructs related to social dimensions showed substantial effects. "Friendship (helpfulness)" had the largest effect, with an Mdn value difference of four scale points of the variant with DP and the control condition (Mdn = 6 vs. Mdn = 2; Z = 6.751; r = 0.774). This finding indicates that learners perceived the PCA as helpful due to its role presentation and the support during the knowledge application and learning story. "Trusting beliefs" showed the second largest effect of the constructs (Mdn = 6 vs. Mdn = 4; Z = 6.145, r = 0.705). This result indicates that the learners perceived the role of the PCA as an eye-level classmate to be pleasant. "Self-Efficacy" demonstrated substantial improvement as well (Mdn = 6; Mdn = 4; Z = 5.945; r = 0.682). This finding underscores the motivational impact of the gamified PCA's role in supporting students in their learning tasks. "Relevance and goal-orientation" (Mdn = 6; Mdn = 5; Z = 5.139; r = 0.589) and "value/usefulness" (Mdn = 6 vs. Mdn = 5; Z = 5.306; r = 0.609) exhibited lower, but still significant, results. "Competency awareness" had the lowest but still strong effect (Mdn = 5 vs. Mdn = 4; Z = 4.889, r = 0.561). These results underscore the relevance of social dimensions in the effectiveness of gamified PCAs. Overall, the students perceived the design positively, so we built on it in the next design cycle.

3.4. Second iteration: implementation, design, and evaluation of the software prototype

Since the first iteration showed that learners positively evaluated the DPs of the GNPL framework in a conceptual prototype, we implemented the fully functional PCA in the second iteration. In the following, we report the design of the PCA architecture, the instantiation of the DFs, and the evaluation results.

3.4.1. Implementation of the gamified PCA

For implementing the artifact in line with the GNPL framework, we utilized the "Rasa" framework (Rasa, 2023) and its software development kit (Bocklisch et al., 2017). The natural language understanding pipeline handled open-ended questions, for example, about the user's score or specific verb tenses. During goal setting, the PCA provides various suggestions for a language learning goal.

The iterative process, beginning with the initial determination of the goal and subsequent specification of the timeframe, is made possible using "Rasa forms," a function of the Rasa framework for conversational design. This function, moreover, enables the PCA to validate the correctness of users' answers using an API call to an AI grammar-checking tool called LanguageTool (2023). The function enables the user to respond again in case of an incorrect answer. The LanguageTool enables the assessment of the correctness of the response and the calculation of a score. Before granting this opportunity, our PCA gives the user hints through the LanguageTool API interface. With this guidance (scaffolding), users can independently acquire knowledge and receive assistance from the PCA to solve the language learning tasks. The communication channel between the user and the PCA is via the messaging tool "Slack." In Slack, we integrated the visual game elements (badges, level advancements, progress bars, and quests as part of the learning stories) as images into the dialog. This decision was vital to reduce the learners' cognitive load. As learners would otherwise lose track of the many incoming and outgoing messages in the interactive gamification setting, they can visually see core milestones of the communication. Therefore, the PCA strategically alternates between social interactions and gamification elements. The PCA ensures that motivational components enhance rather than distract from the learning experience. Rewards were supplemented by text messages in which the PCA personally greeted the learners. We also implemented other design elements, such as goal-setting functions, feedback, and learning tasks, through text communication and buttons to control the dialog. Appendix D illustrates the architecture of our prototype[2]. The further subsection will dive deeper into how we implemented each DP of the GNPL framework.

3.4.2. Representation of the design principles in the software prototype

Regarding DP1, learners set SMART (specific, measurable, attainable, realistic, and time-related) goals with the PCA (Doran, 1981). The PCA indicates to learners during the dialog that it helps them set a suitable common goal *("Let's set an overall learning goal together")*. Through this social support, the PCA gives learners the feeling of being supported by a mentor in formulating and pursuing their goals. Ben initially asks learners to formulate an overall vision for the course and provides several options like preparing for a semester abroad or achieving better grades. Users can set a timeline for achieving their goal to fit their learning pace (Latham et al., 2011; Locke & Latham, 1990). Learners then set subgoals, like how much vocabulary to learn in a month. The overall goal and the subgoals guide their English learning journey. The PCA then starts a quiz aligned with the learning goals, such as practicing the new vocabulary, and communicates the score after completion. The PCA, as a language learning mentor, reflects on the learning pace, asking if the learner wants to continue at the same pace or adjust the time frame. This feedback helps learners adapt their approach to ensure continuous progress (e.g., the number of questions they want to tackle next). By reflecting on the progress, learners are encouraged to learn in a self-regulated manner (Scheu & Benke, 2022).

For DP2, Ben acts as a more experienced classmate and states that it has already completed the exam and can offer helpful tips. For example, the PCA states at the beginning of the conversation: *"I am already in the* 4th *semester and if you would like, I can give you tips on this course as your learning companion."* The PCA initially summarizes the grammatical structure relevant to the upcoming quizzes, followed by simple-choice questions to identify tenses. These questions ensure that learners internalize the schema of the tenses. Following Bloom's taxonomy (Anderson & Krathwohl, 2001; Bloom, 1956) and goal-setting theory (Locke & Latham, 1990), the difficulty of the questions increases. The PCA uses open tasks to ask learners to form a sentence using given words. We did not include time pressure because it may induce cognitive stress (Ackerman & Lauterman, 2012). Ben provides feedback on the correctness of the learners' written sentences and awards points based on the number of attempts to answer the question. If learners do not answer correctly, the PCA helps them using scaffolding (Vygotsky, 1978). The PCA points out if the user has not used all the given words. It suggests corrections if the answer contains spelling errors using the LanguageTool (2023). The support provided by the PCA in the gamified tasks is a crucial function. On the one hand, it ensures that learners are not cognitively overloaded as they work through tasks step by step. On the other hand, learners perceive the PCA's presence as a supporter that helps them when they get stuck. This approach ensures that learners do not feel left alone with challenges (Kim et al., 2006). At each session's end, the PCA emphasizes the importance of regularly reviewing the knowledge and reminds users to repeat tasks regularly. For instance, it articulates: "*I'll keep track of your answers to the questions and will remind you to repeat the quiz."*

For DP3, the PCA rewards learners' individual performance by acting as a motivator. The PCA balances textual messages with visual elements. Therefore, the chatbot limits the use of images to key achievements (level upgrades, major progress in the progress bar, badges) and incorporates concise text to maintain a clear conversational flow. We did not implement rankings, as the interview and literature results (Super et al., 2019) suggest they could cause frustration. Ben rewards points based on the learner's attempts to answer the question: five points for correct first attempts and two points for correct second attempts. The PCA communicates these rewards via a text message (e.g., "*You have collected 5 more points. Keep it up!*"). Stars recognize their accomplishments in a specific field and serve as a marker of their medium-term progress. The PCA provides stars through emojis in text messages and remembers the number of stars collected. Learners who reach long-term milestones, such as mastering two different tenses, receive a badge as a visual representation. They level up by completing a certain number of tasks correctly, and a progress bar indicates how much of the lesson remains until completion. Positive reinforcement through different game elements aligned to short-, medium- and long-term learning goals aims to foster learners' competency awareness (Ryan & Deci, 2000; Sailer et al., 2017).

For DP4, the PCA helps students apply knowledge in real-life situations (Forsyth et al., 2020). Therefore, Ben is a narrator and guides learners through a fictional case (traveling to Rome with a friend). In its role as narrator, the PCA introduces the story as an independent actor and guides the learner through individual, included quests, for example, "*Imagine you are traveling to Rome with your*

[2] The software code, including a documentation of the final version of the gamified PCA Ben, is available on GitHub: https://github.com/tubspaulkeller/Conversational-AI-Agent-For-English-Learning

best friend." The PCA presents open tasks like ordering a taxi or asking for student tickets to the Colosseum. If learners answer incorrectly, the PCA provides hints to help them correct their answers so they can practice the foreign language interactively (Chi & Wylie, 2014; Chiu et al., 2023). The story contains smaller subgoals (quests). Visual cards with images and text modules display the individual quests to maintain an overview of core tasks and avoid distraction from too many text messages.

Fig. 6 shows exemplary how we integrated the different DPs of the GNPL framework into our prototype. The following subsection presents the quantitative and qualitative evaluation results of our gamified PCA design.

3.4.3. Quantitative and qualitative evaluation of the gamified PCA

We conducted a user test to validate the practical benefits of the gamified PCA with DPs. Learners tested the four DPs for about 30 min. Afterward, participants completed a 20-min questionnaire assessing the interaction with Ben. We used the same constructs as in our first iteration and added intrinsic motivation (Plant & Ryan, 1985), extrinsic motivation (Pintrich, 1991), and enjoyment & satisfaction (Z.-H. Chen & Chan, 2008; Keller, 1987). We included these constructs for greater focus on motivational outcomes in the final PCA evaluation as our project's core goal. Thus, our selection of constructs aligns with the growing trend in the *Computers & Education* domain to integrate motivational, pedagogical, and social factors when studying gamification (e.g., C. -M. Chen et al., 2023; Melo et al., 2020). A 7-point Likert scale was employed for all the scales. Appendix C contains a detailed overview of the constructs and items used in the questionnaire. 70 students participated in our survey (48 male and 22 female; average age = 25). The participants belong to the same target group as in design cycle 1 and were non-native English-speaking students from predominantly business and technical degree programs. Both Cronbach's α and composite reliability (CR) were consistently above 0.7, confirming that our dataset is suitable for further analyses (Hair et al., 2014). Specifically, the constructs demonstrated the following values: relevance and goal-orientation (α = 0.87, CR = 0.92), value/usefulness (α = 0.96, CR = 0.96), self-efficacy (α = 0.91, CR = 0.93), friendship (helpfulness) (α = 0.91, CR = 0.93), trusting beliefs (α = 0.90, CR = 0.93), competency awareness (α = 0.95, CR = 0.96), intrinsic motivation (α = 0.84, CR = 0.89), extrinsic motivation (α = 0.86, CR = 0.91), and enjoyment and satisfaction (α = 0.78, CR = 0.86). We calculated each construct's mean values (MV), medians (Mdn), and standard deviations (SD). Each construct represents the aggregation of its individual scale items, with descriptive statistics computed on these aggregated scores. To enrich the results with qualitative feedback, we asked open-ended questions in our survey regarding the positive aspects of the interaction and suggestions for improvement. Of the 70 participants, 55 participants answered the open-ended questions, which we categorized according to related comments. We also conducted three focus group workshops with a total of 17 of the survey participants lasting approx. 30 min. We selected these participants because they had prior knowledge of the topic since they were involved in a university seminar about gamified digital education. The focus group interviews included feedback on Ben's strengths and weaknesses as well as the perceptions of the four DPs regarding motivation and usage intention. In the following, we will highlight the main results of our quantitative and qualitative evaluation.

As shown in Fig. 7, the boxplot results indicate that all constructs have Mdn above the midpoint of the 7-point Likert scale. Specifically, the Mdn values for the constructs are as follows: "relevance and goal-orientation" (Mdn = 6), "value/usefulness" (Mdn = 6), "self-efficacy" (Mdn = 5), "friendship (helpfulness)" (Mdn = 5), "trusting beliefs" (Mdn = 6), "competency awareness" (Mdn = 6), "intrinsic motivation" (Mdn = 6), "extrinsic motivation" (Mdn = 5), and "enjoyment and satisfaction" (Mdn = 5.5).

In addition to the median-based boxplot depiction, we also computed MV and SD for each construct, which are as follows: relevance and goal-orientation (MV = 5.65, SD = 1.16), value/usefulness (MV = 5.28, SD = 1.35), self-efficacy (MV = 5.19, SD = 1.16), friendship (helpfulness) (MV = 4.92, SD = 1.39), trusting beliefs (MV = 5.63, SD = 1.07), competency awareness (MV = 5.55, SD = 1.23), intrinsic motivation (MV = 5.74, SD = 1.24), extrinsic motivation (MV = 4.65, SD = 1.47), and enjoyment & satisfaction (MV = 5.23, SD = 1.22).

In the following, we highlight the main results of our quantitative evaluation. The quantitative results show a tendency towards a positive perception of the PCA. SD was lowest for "trusting beliefs" (SD = 1.05), indicating that most participants consistently felt trust towards the PCA. The SD was the highest for "extrinsic motivation" (SD = 1.47). The construct "extrinsic motivation" had the lowest rating (MV = 4.65). Instead, the highest score was for the construct "intrinsic motivation" (MV = 5.74). "Relevance and goal-orientation" (MV = 5.65) and "trusting beliefs" (MV = 5.63) also scored high, indicating a positive perception of the goal-setting function and the PCA's role. Regarding the analysis of the results at the level of individual items, statements such as „*I believe* Ben *dealt with me truthfully*" (MV = 6.01; SD = 1.21) and „*I would characterize* Ben*'s dealings with me as honest*" (MV = 6.00; SD = 1.10) received particularly high MV in terms of trusting beliefs. Constructs related to intrinsic motivation, such as „I was attentive when using Ben, and I was happy to do so" (MV = 5.94; SD = 1.27) or „*I was very relaxed when using* Ben" (MV = 5.83; SD = 1.61), also received high ratings. These participant assessments suggest that learners found the interaction engaging and perceived a trusting relationship with the chatbot. Somewhat lower values were observed, for instance, for items related to friendship perceptions or extrinsic motivation, though with higher SD values, such as „Ben *helped me when I needed someone else's opinion*" (MV = 3.99; SD = 1.82) or „*If possible, I would like to get more appreciation or rewards for learning with* Ben *compared to others*" (MV = 4.01; SD = 2.05).

Several respondents moreover stated in the open-ended questions that they found the PCA motivating, for example, "*The* awards *and the point system motivated me. Especially the graphical* awards *are really cool!*" Participants trusted the PCA and mostly perceived it as friendly. For example, one participant remarked it was a helpful and pleasant contact: "*I could warm up with* Ben *quickly and without spending a lot of time. It felt like* Ben *cared about my success in learning."* In addition, this perception could be due to the hints provided by the PCA: "Ben *was a very good learning companion (especially regarding the vocabulary). I enjoyed working with it.* Ben *made me aware of my mistakes."* However, participants wished that the PCA could provide more tangible support: "*I think it would be good that in case of a wrong answer, not only the mistake itself is marked, but in addition, an explanation of the violated rule is given.*" The participants assessed the PCA positively regarding "value/usefulness" and "relevance and goal-orientation." The learners rated positively that the individual

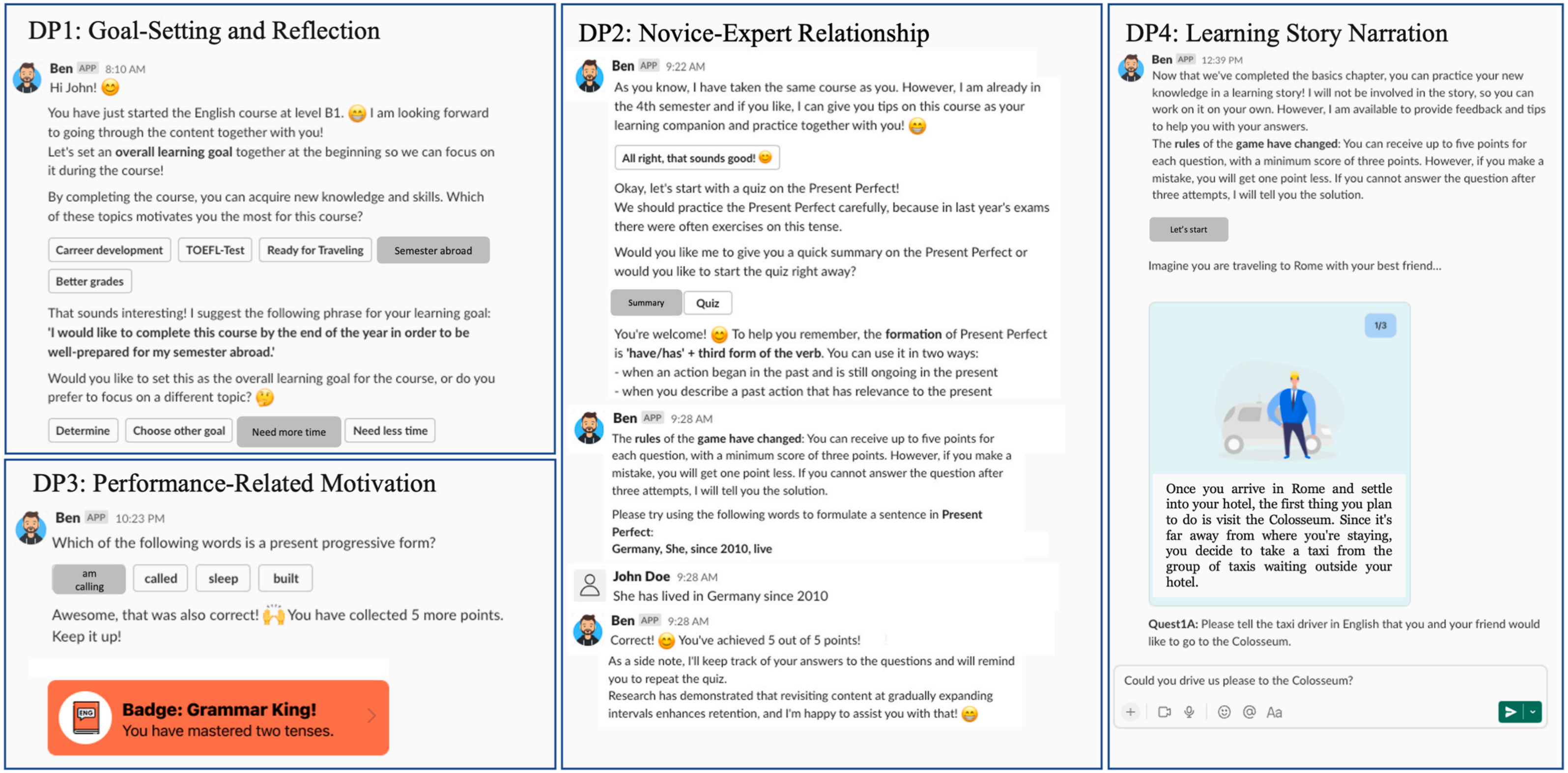

Fig. 6. Excerpts of the conversations with the gamified PCA related to the DPs.

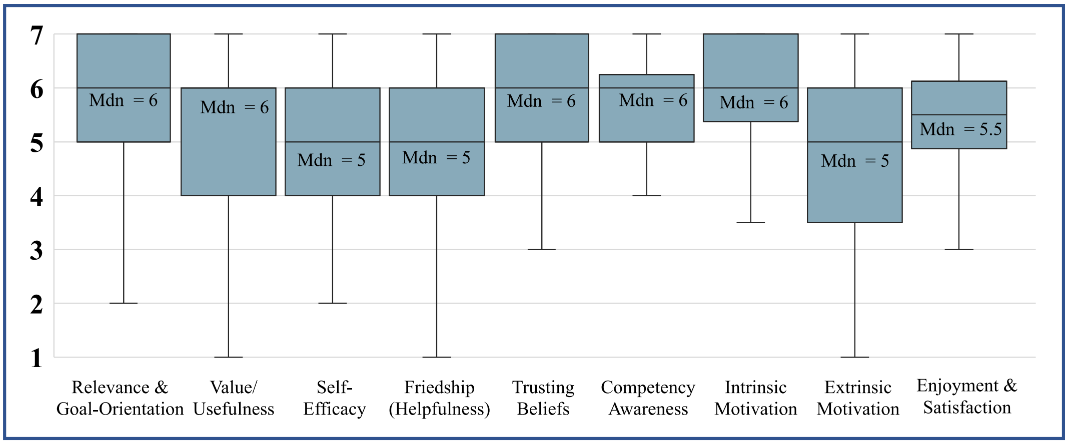

Fig. 7. Boxplot results for the investigated constructs related to the implemented gamified PCA.

DPs were well-aligned with each other. For instance, the learning tasks of DPs 2–4 were well aligned with the goals set in DP1: "*The alignment of the tasks with the goal was well done. The fact that it is aligned with the stated goal makes the whole thing more efficient, and you immediately see that what you are learning is useful. It increases motivation to continue and learn more together with* Ben." Respondents emphasized that they enjoyed the learning session: "Ben *made learning fun, and even if you didn't know something right away,* Ben *rewarded you when you got it right with its help.*" Despite gamification's risks to well-being, such as frustration and addiction, participants highlighted their positive perception towards the PCA's social role. For example, they addressed: *"I found its relaxed manner very pleasant. The chatbot also seemed very energetic and eager, which translated to me, and I was therefore more eager to get straight to the point.*" Participants also emphasized that "*Learning with* Ben *was fun,*" "Ben *was friendly,"* or "Ben *was a very helpful and nice contact.*" However, there was also criticism that the constant awarding of points was sometimes annoying, and participants desired to reduce the reward frequency (e.g., not rewarding points after every correct answer).

4. Discussion

Having evaluated the gamified PCA in two iterations, we now discuss our project's findings with exemplary qualitative insights from the final focus group.

First, gamified PCAs offer a straightforward way for technology-mediated learning and motivate language learning. The gamified PCA design, in line with the GNPL framework, outperformed the control condition for each construct, highlighting the effectiveness of the DPs. The implemented artifact showed a positive motivational perception, with intrinsic motivation being significantly stronger than extrinsic motivation. These findings coincide with the answers to the open-ended questions. In the open-ended questions, learners appreciated self-regulated learning opportunities, for example, "Ben *always gave options to choose from.*" Participants noted that the gamified PCA enables solving quizzes on the go, regardless of location. For example, the learners emphasized that they can easily practice the language with Ben without searching for tasks on the Internet themselves. They noted that this approach is an easy way to motivate themselves to learn the language despite having little time due to work commitments. Students found the chatbot's support beneficial: "Ben *made learning fun, and even if you didn't know something right away,* Ben *rewarded you when you got it right with its help.*" The increasing task difficulty was perceived as positive, which favors self-efficacy (van Dinther et al., 2011): "*The tasks are appropriate to the level and pleasantly entertaining so that they do not overwhelm and encourage to continue.*" The PCA thus succeeded in creating an appropriate balance between playful and learning activities. Furthermore, the PCA counteracted the problem of frustration when completing tasks, as is often the case with gamification (Sailer et al., 2017; Zichermann & Cunningham, 2011). The participants highlighted that rewards and progress indicators foster intrinsic motivation (Sailer et al., 2017). This construct was rated highest in the evaluation (see Chapter 3.4.3). However, some learners perceived extrinsic rewards as negative in terms of motivation, as described by Deci et al. (1999). In line with the ongoing debates in motivation research, different individuals perceive extrinsic rewards differently and extrinsic rewards may create performance pressure (Sailer et al., 2017; Toda et al., 2018). In the PCA context, too many rewards risk distracting learners from the actual content, especially in a mobile chat setting where they might lose track of messages and rewards. This could be a reason for the chatbot's highest SD and lowest MV rating in terms of extrinsic motivation (see Subsection 3.4.3). Therefore, we recommend limiting extrinsic rewards to key achievements and allowing learners to decide how frequently they like to receive these reward elements.

Second, participants expressed that gamified PCAs like Ben could serve as a supporting tool in higher education: „*I would use the chatbot as a supplement to my regular learning if* Ben *teaches a certain subject area that I want to learn.*" They also desired to use gamified PCAs to learn subjects other than English. However, some participants said that they could rather not imagine using the PCA in everyday life. They argued that learning with it is time-consuming and that they prefer to work on their studies on their own without digital PCA support. We assume that the intention to use the PCA depends on students' preferred learning format (e.g., with or without digital tool support).

Third, the participants valued practical knowledge applications with a learning partner and interactive story narration. These features align with the interactive, constructive, active, and passive framework (Chi & Wylie, 2014). The high rating of the PCA in the final study for constructs such as "competency awareness" and "self-efficacy" indicates that the PCA was able to support learners in mastering the learning tasks and stories in its role as a more experienced classmate and that they perceived the praise and rewards as confirmation of their successes (supporting DP2 - DP4). The high rating of "relevance and goal-orientation" emphasizes the PCA's ability to support learners in setting and pursuing meaningful goals (DP1). Learners perceived that it was motivating to solve challenges with Ben's help, highlighting the relevance of scaffolding for PCAs: "Ben *was able to give me good advice without revealing the answer directly.*" Furthermore, this underlines the value of immediate feedback using AI-supported language checking (LanguageTool, 2023). However, participants emphasized several ideas for further development to enable more in-depth assistance. For instance, the PCA might explain mistakes in detail or tell the user more hints if needed (e.g., for a score penalty). Implementing more open interaction requires reliable language processing and much training data. New large language models could facilitate training, reducing developer involvement (Kasneci et al., 2023; Lo et al., 2024). These advances might enable PCAs to handle open-ended questions more properly. For example, participants noted that the chatbot has difficulties in understanding complicated answers and in marking them as correct.

Fourth, we assume that PCAs are most effective when learning feels natural, enabling communication like with a real partner (Chi & Wylie, 2014). The positive perception of open tasks shows that learners prefer dialogues that feel like learning with a companion: "*I also liked its role as a classmate because it seemed as if it could empathize with my situation and had already worked on the same tasks.*" Our first evaluation cycle showed that the social role of the gamified PCA (DP2), expressed through higher "friendship (helpfulness)" and "trusting beliefs" values, had the most significant effect on the positive perception of the PCA. The PCA's effect on these constructs was stronger than the effect on competency awareness, often associated with reward elements (Cruz et al., 2017; Sailer et al., 2017) (DP3). Hence, we assume that the gamified PCA design goes beyond classical gamified learning applications, which focused on the effects of reward elements without a deeper focus on social outcomes (e.g., Cruz et al., 2017; Melo et al., 2020). Even though the second evaluation cycle's participants assessed "competency awareness" higher than "friendship" (helpfulness), they expressed a generally positive perception of the PCA on social constructs as well. The findings align with literature showing a PCA's friendly interaction promotes user experience (Sikström et al., 2022). The DP2 expands the existing literature's findings by proposing the role of a more experienced classmate. This role combines the advantages of the classmate and the expert role - unlike previous studies, which focused only on the two opposing poles of classmate and expert roles (Liew & Tan, 2021; Nguyen, 2023). It should be noted that the evaluation of the PCA's role focused on German students. Hence, DP2 only applies under this boundary condition since PCA role perception is culture-specific. Schlimbach and Zhu (2023), for example, emphasized in an experimental study on the perception of PCA roles that Chinese students prefer the tutor role (with high power distance) rather than the classmate role (low power distance) compared to German students. For their study, they relied on Hofstede's (1993) concept of cultural dimensions. This concept refers to power distance as an essential characteristic of cultures. Cultures with high power distance (e.g., China) tend to emphasize hierarchical structures, while cultures with lower power distance (e.g., Germany) value equality (Hofstede, 1993). Moreover, Schlimbach and Zhu (2023) highlighted that the learners' role preferences depend on the learning setting's formality. While German learners preferred the role of classmate in low-formality settings, such as mobile learning with PCAs, even German learners preferred the role of tutor in high-formality settings with high time pressure (Schlimbach & Zhu, 2023). Developers should, therefore, adapt the design of PCA tools to the learners' respective cultural background and the learning setting's formality (Ernst et al., 2016). To promote learners' autonomy (Ryan & Deci, 2000), it is also vital to give them opportunities to customize the PCA role. In addition, participants suggested further ideas for a more natural learning experience, like audio learning. Audio learning aligns with recent literature on gamified learning with Amazon Alexa (Bräuer & Mazarakis, 2022). However, participants warned against solely self-regulated digital learning, neglecting social exchange with "real classmates." We emphasize gamified PCAs as support tools, not substitutes for learning with peers to actually achieve positive social and learning outcomes (Melo et al., 2020). The participants furthermore highlighted that speaking the language with others is crucial, especially in language learning (Hwang et al., 2024). To address the users' feedback, future research can focus on extending gamified PCAs with a collaboration mode or audio learning.

5. Conclusion

5.1. *Key findings*

PCAs are an innovative opportunity for technology-mediated learning, supporting language education as a critical 21st-century skill. However, PCA applications to date have not yet realized their potential in terms of motivational support for learners (W. Huang et al., 2022; Janssen et al., 2021). Educators can enhance learners' motivation using gamification. Employing gamification is a challenging design task and includes the careful selection and implementation of game elements (Schöbel et al., 2020; Wittmann & Morschheuser, 2022). Considering this background, our paper followed the **research question** *How should gamified PCAs be designed to motivate learners in language education?* To answer this research question, we used a user-centered design science research approach to derive the GNPL framework including four DPs through an interview study with eight learners and a review of PCA and gamification literature: goal-setting and reflection (DP1), novice-expert-relationship (DP2), performance-related motivation (DP3), and learning story narration (DP4). We evaluated the DPs in two subsequent studies for English learning. In the first study with 76 learners, we tested the DPs in a conceptual prototype compared to a control condition. The PCA variant with DPs significantly affected learners' motivation and further constructs related to social and educational outcomes. In the second study, we implemented the DPs in a chatbot prototype that included AI-based language and grammar-checking functions. We evaluated the PCA in a user survey with 70 participants and focus groups with 17 learners. The PCA fostered learners' motivation, and the participants positively highlighted the interactivity of gamification in the PCA context, the relevance of the connection of the game elements with the PCA's social role, and the PCA's positive effect on learners' well-being. The study emphasizes the practical relevance of aligning the PCA's social role, gamification's motivational design, and the educational goals for language learning. This alignment is crucial to enable a meaningful use of this novel educational technology. In the following, we discuss how our findings advance current knowledge in PCAs, gamification research, and language education.

5.2. Contributions and implications

Our results first contribute to the theory stream of PCAs and their effective use in education by presenting the GNPL framework as a replicable, theory-grounded artifact. Research studies (Fryer et al., 2017; W. Huang et al., 2022) and analyses of applications from practice (Janssen et al., 2021; Smutny & Schreiberova, 2020) show that PCAs often fail to reach their full potential in motivating learners. To address this, we contribute new knowledge on designing PCAs adequately. Educational institutions started to offer PCAs for learning support that can be reached independently of time and place. However, these are not yet widespread (Hobert & Meyer von Wolff, 2019; W. Huang et al., 2022). To ensure successful PCA implementations, offering prescriptive knowledge in the early stages of these educational developments is critical. In the *Computers & Education* community, Smutny and Schreiberova (2020) demonstrated in their practical analysis of PCAs for the Facebook Messenger that game elements like quizzes and points foster motivation. However, they criticize that these game elements are not well integrated with the PCA interaction. Additionally, these existing applications incorporated gamification only sporadically by relying on isolated elements rather than employing a comprehensive, systematic approach (Smutny & Schreiberova, 2020). Our article adds value by examining the connection between the user and the PCA, for example, in terms of a novice-expert relationship (DP2). Schneider et al. (2022) and Sikström et al. (2022) also point out that more research is needed on relationship-oriented and adaptive communication. Our article addresses this by evaluating the implementation of relationship-oriented communication combined with gamification. The paper furthermore addresses the challenge that existing gamification applications often neglect social effects (Lo et al., 2024; Melo et al., 2020). The article presents a nuanced view of social interactions, as it focuses on the role of the PCA as a more experienced classmate that combines the advantages of the classmate and expert role, unlike further studies (Liew & Tan, 2021; Nguyen, 2023). We illustrate an approach for the next frontiers of learning with PCAs, which can be further developed in the near future as researchers recently highlighted new opportunities for PCAs with large language models (Kasneci et al., 2023; Lo et al., 2024). Our studies provide insights into the targeted use of AI-based PCAs. These insights include findings for the implementation of open language learning tasks, interactive game stories, and immediate scaffolded feedback (LanguageTool, 2023; Vygotsky, 1978). The paper shows how PCAs support learners in pursuing lower-to-higher-level learning goals in the social role of a learning companion (Anderson & Krathwohl, 2001; Bloom, 1956). Although this approach focuses on learning during leisure time, it can also complement traditional teaching. For instance, teachers can use gamified PCAs to support students to prepare for upcoming lectures or recall learning material afterward. Moreover, they could use this approach for in-class quizzes or learning stories which students solve collaboratively in-class with PCA support. Furthermore, gamified PCAs foster personalized learning, especially in large university courses where one-on-one support is limited (Behr et al., 2021). Through personalized feedback, varied learning tasks, and scaffolding, learners receive tailored assistance that complements instructor-led settings. Teachers can thus devote in-class time to more complex tasks, such as creative problem-solving. These benefits even extend beyond language learning and apply to various fields where learners must follow up on content or prepare for exams. These fields included, for example, programming (Aguirre et al., 2018), natural science (Graesser et al., 2018), or business-related topics (Schlimbach, Lange, & Robra-Bissantz, 2023).

Second, we contribute to the theory stream of gamification design. The GNPL framework demonstrates how gamification complements PCAs as a straightforward way of interactive learning (Chi & Wylie, 2014; Winkler et al., 2020). Previous articles focused on individual game elements in chatbots (e.g., Smutny & Schreiberova, 2020). They did not transfer complete gamification concepts, such as Schöbel et al.'s (2022) taxonomy, to PCAs. This paper addressed this issue by combining several game elements for PCAs along the user needs identified in the interviews. Our studies emphasize the relevance of a careful design that leverages the social role of the PCA, the motivational impact of gamification, and the educational goals specific to language learning (Benner et al., 2022; X. Chen et al., 2020). This view distinguishes our DPs' framing from ChatGPT with a very open dialog that the user must proactively control (Kohnke et al., 2023; Lo et al., 2024). Instead, the DPs target structured learning support by combining social, motivational, and educational theories and address the need for more theory-driven approaches to gamification for PCAs (Wittmann & Morschheuser, 2022). Unlike ChatGPT, gamification supports learners to pursue their goals, complete language learning tasks interactively, receive rewards from the PCA and engage in learning stories. This structured approach contributes to learners pursuing formal learning goals through these specific pedagogical features (Anderson & Krathwohl, 2001; Bloom, 1956). At the same time, it particularly helps students who are challenged to set learning goals on their own and actively pursue them (Wong et al., 2021). This approach sets the tool apart from ChatGPT as well, which, due to its openness, requires users to have higher self-regulation and prompting skills (Tolzin et al., 2024). The study results show that learners particularly emphasize the simplicity of solving ready-made quizzes in short units, for example, on the way to work. These features benefit our paper's target group of vocational students who have limited time due to work commitments. By exhibiting a social role aligned with gamification, the PCA design extends conventional language learning apps like “Duolingo” (Dehghanzadeh et al., 2021; Hwang et al., 2024). The study results show that social features such as the PCA's role as an experienced classmate are crucial for gamification to be effective. Through the PCA's social role, learners not only receive task-based support but also experience the social presence of a virtual companion (Kim et al., 2006). These features differentiate this paper's framing from existing language learning apps such as Duolingo and we recommend educational designers to adopt these functions. In particular, we contribute to meaningful gamification as an essential research direction in gamification (Nicholson, 2015). This research stream includes evaluating gamification in new contexts, as for PCAs in a highly interactive context (Hamari et al., 2014). Meaningful gamification research is vital since the appropriate gamification design is challenging. Meaningful gamification depends on the context and the users' needs to counteract negative effects like frustration, overchallenge, or neglection of social factors in learning (Melo et al., 2020; Sailer et al., 2017). The GNPL framework along with the DPs presented have a positive impact in this regard. The DPs show how meaningful gamified PCA design promotes motivational effects and avoids negative effects.

Third, we contribute to the theory stream of language learning as a crucial skill for personal and professional development. Language learning is complex and challenging, leading to drop-outs (Vázquez Mendoza et al., 2024). Hence, learners aim for motivational support in digital learning environments. We address the lack of individual language support due to the imbalance between learners and lecturers and the struggle many learners face in motivating themselves (Derakhshan et al., 2021). This approach considers the challenges of language education like practicing interactively. PCAs can help students by assuming the functions of teachers or tutors (Hobert & Meyer von Wolff, 2019). We show how PCAs can be combined with gamification using the GNPL framework to create a motivational learning environment, as called for in research (Botes et al., 2022; Y. Wang et al., 2021). Overall, we demonstrate the effectiveness of reusable DPs for the example of English language learning.

Fourth, our paper contributes practical knowledge for educators and developers who need to select and design appropriate educational tools. In particular, the choice of motivational incentives and gamification mechanisms is challenging (Schöbel et al., 2020; Weber et al., 2021). The paper responds to the community's call to create validated design recommendations for the use of AI-based educational technologies (Díaz & Nussbaum, 2024; Wittmann & Morschheuser, 2022). The GNPL framework and its four DPs, as abstract design recommendations, allow application not only in the context investigated in this study but also in further fields of language education and low-threshold learning. We furthermore show how practitioners can translate this abstract knowledge into a mature PCA. Therefore, we provide new knowledge both from a technical point of view (architecture) and a domain perspective (DFs). The four DPs and the associated DFs serve not only as guidance for the in-house development of educational tools but also as selection criteria for industry tools.

5.3. Limitations and future research directions

Although our findings offer new knowledge related to gamified PCA design, there are boundaries to our studies that future research can address. The evaluations in our two design cycles focused on demonstrating a proof of concept with high internal validity rather than evaluating PCAs under real-world conditions. Future research should investigate the DPs' long-term effectiveness, following the iterative nature of artifact development in design science research (Kuechler & Vaishnavi, 2008). Regarding the generalizability of our results, we have only shown the applicability of the GNPL framework and its four DPs to English learning. This application context enabled us to evaluate our design knowledge for a context many students struggle with. The PCA addresses DPs for low-threshold learning in a mobile setting. These are transferable to learning further foreign languages. Since gamification is particularly effective in low-threshold settings (Khosrawi-Rad et al., 2023a), our DPs are transferable to other subject areas that focus on low-threshold mobile practice of learning content, as outlined by the study participants. PCA designers need to adapt their implementations to the application context and learning content (Dai et al., 2022; N. L. Schroeder et al., 2013). Given DSR's iterative nature, the GNPL framework should also be discussed with product developers in future studies regarding developers' perceptions and the DPs' transferability into practice. We therefore recommend that further researchers instantiate the GNPL framework and consult developers about associated benefits and potential suggestions for future research in line with design science research's aim of ensuring projectability (vom Brocke et al., 2020). Such a study can evaluate whether the proposed DPs are sufficient for implementing gamified PCAs or if additional requirements for interaction design are needed. Moreover, we did not analyze the effect on more complex skills like critical thinking or problem-solving, for which serious games might be better suited than gamification (Khosrawi-Rad et al., 2023a). Future research can build on our design knowledge by analyzing whether it is transferable to further content areas and more complex learning competencies. In addition, further studies can investigate the culture-specific adaptation of gamified PCA roles (especially considering cultures with different levels of power distance). We moreover recommend follow-up studies with more technically advanced solutions, such as large language models, and to evaluate their scalability (e.g., in mass lectures). We acknowledge that approaches other than gamification may also promote learner engagement. Future studies could, for example, investigate how developers can improve conversational flows through user experience writing techniques (e.g., empathy, humor, or certain salutations). Moreover, this paper's studies focused on the applicability of our GNPL framework along with its DPs for text-based PCAs. PCA designers should consider adapting the implementation of the DFs. In voice-only solutions, some game elements like progress bars can only be implemented implicitly using voice commands (Bräuer & Mazarakis, 2022).

Overall, this paper contributes to understanding the effective use of PCAs through gamification as an innovative new educational technology. We address a solution to the challenges faced by learners in language education. Our DPs additionally provide a framework for future research in further learning contexts.

CRediT authorship contribution statement

Bijan Khosrawi-Rad: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Paul Felix Keller:** Writing – original draft, Software, Methodology, Data curation. **Dennis Benner:** Writing – review & editing, Writing – original draft, Data curation. **Linda Grogorick:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Arne Borchers:** Software, Methodology, Formal analysis, Data curation. **Andreas Janson:**

Writing – review & editing, Supervision. **Jan Marco Leimeister:** Supervision. **Susanne Robra-Bissantz:** Supervision.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This contribution results from the project StuBu (grant #21INVI06), which the German Federal Ministry of Education and Research (BMBF) funded. Furthermore, this contribution results from the projects Komp-HI funded by the German Federal Ministry of Education and Research (BMBF, grant #16DHBKI073) and Managing the Algorithm: Prompt Engineering for AI-based Systems as an Emerging Business Skill funded by the Swiss National Science Foundation (SNSF, grant #221281).

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2025.105374.

Data availability

The authors do not have permission to share data.

References

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education?: It depends upon the type of badge and expertise of learner. *Educational Technology Research & Development, 61*(2), 217–232.
Ackerman, R., & Lauterman, T. (2012). Taking reading comprehension exams on screen or on paper? A metacognitive analysis of learning texts under time pressure. *Computers in Human Behavior, 28*(5), 1816–1828.
Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods, 17*(4), 351–371.
Aguirre, C. C., Kloos, C. D., Alario-Hoyos, C., & Munoz-Merino, P. J. (2018). Supporting a MOOC through a conversational agent. Design of a first prototype. In *SIIE 2018 proceedings*. Scopus.
Alt, D. (2023). Assessing the benefits of gamification in mathematics for student gameful experience and gaming motivation. *Computers & Education, 200*, Article 104806.
Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist, 23*(4), 225.
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
Bandura, A. (1997). *Self-efficacy: The exercise of control New York*. WH Freeman and Company.
Baroudi, J. J., & Orlikowski, W. J. (1989). The problem of statistical power in MIS research. *MIS Quarterly, 13*(1), 87–106.
Behr, A., Giese, M., Teguim Kamdjou, H. D., & 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), 325–343.
Benner, D., Schöbel, S., Süess, C., Baechle, V., & Janson, A. (2022). Level-Up your Learning – Introducing a Framework for Gamified Educational Conversational Agents. Wirtschaftsinformatik 2022 Proceedings.
Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals* (1st ed.). Longman Group.
Bocklisch, T., Faulkner, J., Pawlowski, N., & Nichol, A. (2017). Rasa: Open source language understanding and dialogue management. *arXiv:1712*, Article 05181.
Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457.
Botes, E., Dewaele, J.-M., & Greiff, S. (2022). Taking stock: A meta-analysis of the effects of foreign language enjoyment. *Studies in Second Language Learning and Teaching, 12*(2), 205–232.
Bräuer, P., & Mazarakis, A. (2022). How to design audio-gamification for language learning with Amazon Alexa? - a long-term field experiment. *International Journal of Human-Computer Interaction*, 1–18.
Chamot, A. U., & O'Malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley Publishing Company.
Chen, Z.-H., & Chan, T.-W. (2008). Learning by substitutive competition: Nurturing my-pet for game competition based on open learner model. *IEEE DIGITEL Proceedings*.
Chen, C.-M., Ming-Chaun, L., & Kuo, C.-P. (2023). A game-based learning system based on octalysis gamification framework to promote employees' Japanese learning. *Computers & Education, 205*, Article 104899.
Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100002.
Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243.
Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments, 32*(7), 1–17.
Ciuchita, R., Heller, J., Köcher, S., Köcher, S., Leclercq, T., Sidaoui, K., & Stead, S. (2023). It is really not a game: An integrative review of gamification for service research. *Journal of Service Research, 26*(1), 3–20.
Cruz, C., Hanus, M. D., & Fox, J. (2017). The need to achieve: Players' perceptions and uses of extrinsic meta-game reward systems for video game consoles. *Computers in Human Behavior, 71*, 516–524.
Dai, L., Jung, M. M., Postma, M., & Louwerse, M. M. (2022). A systematic review of pedagogical agent research: Similarities, differences and unexplored aspects. *Computers & Education, 190*, Article 104607.
Deci, E., Koestner, R., & Ryan, R. (1999). A meta-analytic review of experiments examining the effect of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*(6), 627–668.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning, 34*(7), 934–957.
Delaney, D., Kummer, T., & Singh, K. (2019). Evaluating the impact of online discussion boards on student engagement with group work. *British Journal of Educational Technology, 50*(2), 902–920.
Derakhshan, A., Kruk, M., Mehdizadeh, M., & Pawlak, M. (2021). Boredom in online classes in the Iranian EFL context: Sources and solutions. *System (Linköping), 101*, Article 102556.
Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification.". *MindTrek 2011 Proceedings*.
Díaz, B., & Nussbaum, M. (2024). Artificial intelligence for teaching and learning in schools: The need for pedagogical intelligence. *Computers & Education*, Article 105071.
Doran, G. T. (1981). There'sa SMART way to write management's goals and objectives. *Management Review, 70*(11), 35–36.
Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.
Ernst, S.-J., Janson, A., Söllner, M., & Leimeister, J. M. (2016). It's about Understanding Each Other's Culture – Improving the Outcomes of Mobile Learning by Avoiding Culture Conflicts. ICIS 2016 Proceedings.
Euler, D. (2014). Design-Research–a paradigm under development. *Design-Based Research, 27*, 15–44.
Fadhil, A., & Villafiorita, A. (2017). An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. *25th conference on user modeling, adaptation and personalization (UMAP) proceedings*.
Forsyth, C. M., Graesser, A., & Millis, K. (2020). Predicting learning in a multi-component serious game. *Technology, Knowledge and Learning, 25*(2), 251–277. Scopus.
Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior, 75*, 461–468.
Gao, L., & Liu, H. (2022). Revisiting students' foreign language learning demotivation: From concepts to themes. *Frontiers in Psychology, 13*, Article 1030634.
Graesser, A. C., Dowell, N., Hampton, A. J., Lippert, A. M., Li, H., & Shaffer, D. W. (2018). Building intelligent conversational tutors and mentors for team collaborative problem solving: Guidance from the 2015 program for international student assessment. In *Building intelligent tutoring systems for teams* (Vol. 19, pp. 173–211). Emerald Publishing Limited.
Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). The anatomy of a design principle. *Journal of the Association for Information Systems, 21*, 1622–1652.
Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*(2), 337–355.
Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems, 8*(5), 312–335.
Griol, D., & Callejas, Z. (2016). Conversational agents as learning facilitators: Experiences with a mobile multimodal dialogue system architecture. In *Formative assessment, learning data analytics and gamification* (pp. 313–331). Elsevier.
Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Pearson.
Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers, 3*, 275–285.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? — a literature review of empirical studies on gamification. *HICSS 2014 Proceedings*.
Hew, K. F., Huang, W., Du, J., & Jia, C. (2023). Using chatbots to support student goal setting and social presence in fully online activities: Learner engagement and perceptions. *Journal of Computing in Higher Education, 35*(1), 40–68.
Hien, H. T., Cuong, P.-N., Nam, L. N. H., Nhung, H. L. T. K., & Thang, L. D. (2018). Intelligent assistants in higher-education environments: The FIT-EBot, a chatbot for administrative and learning support. *ACM SoICT 2018 Proceedings*.
Ho, J. C. F., & Ng, R. (2022). Perspective-taking of non-player characters in prosocial virtual reality games: Effects on closeness, empathy, and game immersion. *Behaviour & Information Technology, 41*(6), 1185–1198.
Hobert, S., & Meyer von Wolff, R. (2019). Say hello to your new automated tutor – a structured literature review on pedagogical conversational agents. *Wirtschaftsinformatik 2019 Proceedings*.
Hofstede, G. (1993). Cultural constraints in management theories. *Academy of Management Perspectives*. https://doi.org/10.5465/ame.1993.9409142061
Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257.
Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society, 26*(1), 112–131.
Hwang, G.-J., Rahimi, M., & Fathi, J. (2024). Enhancing efl learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a mall app: A microgenetic perspective. *Computers & Education, 214*, Article 105015.
IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). *The game experience questionnaire*. Eindhoven University of Technology Research Portal. https://research.tue.nl/en/publications/the-game-experience-questionnaire/fingerprints/.
Janson, A., Söllner, M., & Leimeister, J. M. (2020). Ladders for Learning: Is Scaffolding the Key to Teaching Problem-Solving in Technology-Mediated Learning Contexts? *Academy of Management Learning & Education, 19*(4), 439–468.
Janssen, A., Grützner, L., & Breitner, M. H. (2021). Why do chatbots fail? A critical success factors analysis. *ICIS 2021 Proceedings*.
Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education, 206*, Article 104898.
Johnson, C., Urazov, M., & Zanoli, E. (2022). Escapeling: A gamified, AI-supported chatbot for collaborative language practice. *The learning ideas conference (TILC) proceedings*.
Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., … Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274.
Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development, 10*(3), 2.
Khlaisang, J., & Koraneekij, P. (2024). Roles of chatbots in gamified self-regulated learning system to enhance achievement motivation of learners in massive open online courses. *Electronic Journal of e-Learning, 22*(8), 106–120.
Khosrawi-Rad, B., Grogorick, L., & Robra-Bissantz, S. (2023a). Game-inspired Pedagogical Conversational Agents: A Systematic Literature Review. *AIS Transactions on Human-Computer Interaction, 15*(2), 146–192.
Khosrawi-Rad, B., Borchers, A., Grogorick, L., & Robra-Bissantz. (2023b). Design Principles for Gamified Pedagogical Conversational Agents. MindTrek 2023 Proceedings.
Kim, Y., Baylor, A. L., & PALS Group. (2006). Pedagogical agents as learning companions: The role of agent competency and type of interaction. *Educational Technology Research & Development, 54*(3), 223–243.
Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 26*, 160–169.
Knispel, J., Wittneben, L., Slavchova, V., & Arling, V. (2021). Skala zur Messung der beruflichen Selbstwirksamkeitserwartung (BSW-5-Rev). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*.
Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550.
Komiak, S. Y. X., & Benbasat, I. (2006). The effects of personalizaion and familiarity on trust and adoption of recommendation agents. *MIS Quarterly, 30*(4), 941–960.
Krath, J., Schürmann, L., & von Korflesch, H. F. O. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior, 125*, Article 106963.
Kuechler, B., & Vaishnavi, V. (2008). On theory development in design science research: Anatomy of a research project. *European Journal of Information Systems, 17*(5), 489–504.
Kulik, J. A., & Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42–78.

LanguageTool. (2023). LanguageTool—Grammatik-, Stil- und Rechtschreibprüfung. https://languagetool.org/de.
Latham, G. P., Ganegoda, D. B., & Locke, E. A. (2011). Goal-setting: A state theory, but related to traits. In *The Wiley-Blackwell handbook of individual differences* (pp. 579–587). Wiley Blackwell.
Liew, T. W., & Tan, S.-M. (2021). Social cues and implications for designing expert and competent artificial agents: A systematic review. *Telematics and Informatics, 65*, Article 101721.
Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions, task-related attitudes, and learning achievement. *Journal of Educational Technology & Society, 16*(4), 275–286.
Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly, 41*(4).
Lo, C. K., Hew, K. F., & Jong, M. S. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education, 219*, Article 105100.
Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.
Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International Journal of Management, Business, and Administration, 15*(1), 1–6.
Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between level up and game over: A systematic literature review of gamification in education. *Sustainability, 13*(4), 2247.
Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (12th ed.). Beltz Verlag.
McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359.
Melo, C., Madariaga, L., Nussbaum, M., Heller, R., Bennett, S., Tsai, C.-C., & van Braak, J. (2020). Editorial: Educational technology and addictions. *Computers & Education, 145*, Article 103730.
Mendelson, M. J., & Aboud, F. E. (1999). Measuring friendship quality in late adolescents and young adults: McGill friendship questionnaires. *Canadian Journal of Behavioural Science, 31*(2), 130–132.
Möller, F., Guggenberger, T. M., & Otto, B. (2020). Towards a method for design principle development in information systems. *DESRIST 2020 Proceedings*.
Nguyen, H. (2023). Role design considerations of conversational agents to facilitate discussion and systems thinking. *Computers & Education, 192*, Article 104661.
Nicholson, S. (2015). A RECIPE for meaningful gamification. In T. Reiners, & L. C. Wood (Eds.), *Gamification in education and business* (pp. 1–20). Springer International Publishing.
Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems, 7*(3), 11: 1–11:40.
Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. *International conference on education and E-learning (ICEEL) proceedings*. Scopus.
Pintrich, P. R. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*.
Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality, 53*(3), 435–449.
Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145–182.
Rasa. (2023). Conversational AI platform. https://rasa.com/.
Rinn, H., Khosrawi-Rad, B., Schlimbach, R., Masurek, M., Robra-Bissantz, S., & Markgraf, D. (2022). Needs of Students in Further Education-A Mixed Methods Study. GeNeMe 2022 Proceedings.
Ruan, S., Jiang, L., Xu, J., Tham, B. J.-K., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., & Landay, J. A. (2019). QuizBot: A dialogue-based adaptive learning system for factual knowledge. *CHI Proceedings 2019*.
Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450–461.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67a.
Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380.
Scheu, S., & Benke, I. (2022). Digital assistants for self-regulated learning: Towards a state-of-the-art overview. *ECIS 2022 Proceedings*.
Schlimbach, R., Lange, T., & Robra-Bissantz, S. (2023). A longitudinal study on boosting students' performance with a learning companion. *ICIS 2023 Proceedings*.
Schlimbach, R., Windolf, C., & Robra-Bissantz, S. (2023). A service perspective on designing learning companions as bonding and mindful time managers in further education. *ECIS 2023 Proceedings*.
Schlimbach, R., & Zhu, X. (2023). Let's (not) Be friends! - an intercultural experiment with German and Chinese students on the perception of power distance in pedagogical conversational agents. *ECIS 2023 Proceedings*.
Schneider, S., Krieglstein, F., Beege, M., & Rey, G. D. (2022). The impact of video lecturers' nonverbal communication on learning – an experiment on gestures and facial expressions of pedagogical agents. *Computers & Education, 176*, Article 104350.
Schöbel, S., Janson, A., & Söllner, M. (2020). Capturing the complexity of gamification elements: A holistic approach for analysing existing and deriving novel gamification designs. *European Journal of Information Systems, 29*(6), 641–668.
Schöbel, S., Saqr, M., & Janson, A. (2021). Two decades of game concepts in digital learning environments–A bibliometric study and research agenda. *Computers & Education, 173*, Article 104296.
Schöbel, S., Schmidt-Kraepelin, M., Janson, A., & Sunyaev, A. (2021). Adaptive and personalized gamification designs: Call for action and future research. *AIS Transactions on Human-Computer Interaction, 13*(4), 479–494.
Schöbel, S. M., Janson, A., & Leimeister, J. M. (2023). Gamifying Online Training in Management Education to Support Emotional Engagement and Problem-solving Skills. *Journal of Management Education, 47*(2), 166–203.
Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49*(1), 1–39.
Schroeder, B. L., Fraulini, N. W., Van Buskirk, W. L., & Johnson, C. I. (2020). Using a non-player character to improve training outcomes for submarine electronic warfare operators. *HCII 2020 Proceedings*.
Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies, 74*, 14–31.
Sikström, P., Valentini, C., Sivunen, A., & Kärkkäinen, T. (2022). How pedagogical agents communicate with students: A two-phase systematic review. *Computers & Education, 188*, Article 104564.
Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook messenger. *Computers & Education, 151*, Article 103862.
Söbke, H., & Streicher, A. (2016). Serious games architectures and engines. In *Entertainment computing and serious games* (pp. 148–173). Springer.
Strohmann, T., Siemon, D., Khosrawi-Rad, B., & Robra-Bissantz, S. (2022). Toward a design theory for virtual companionship. *Human-Computer Interaction, 38*(3–4), 1–41.
Super, J., Keller, R. H., Betts, T. K., & Roach Humphreys, J. (2019). Simulation games: Learning goal orientations and norms for knowledge sharing. *Academy of Management Proceedings, 2019*.
Tai, T.-Y. (2024). Comparing the effects of intelligent personal assistant-human and human-human interactions on EFL learners' willingness to communicate beyond the classroom. *Computers & Education, 210*, Article 104965.
The Partnership for 21st Century Learning. (2024). Framework for 21st Century Learning—a unified vision for learning to ensure student success in a world where change is constant and learning never stops. *Battelle for Kids*. https://www.battelleforkids.org/insights/p21-resources/.

Toda, A. M., Valle, P. H. D., & Isotani, S. (2018). The dark side of gamification: An overview of negative effects of gamification in education. In A. I. Cristea, I. I. Bittencourt, & F. Lima (Eds.), *Higher education for all. From challenges to novel technology-enhanced solutions* (pp. 143–156). Springer International Publishing. https://doi.org/10.1007/978-3-319-97934-2_9.

Tolzin, A., Knoth, N., & Janson, A. (2024). *Leveraging Prompting Guides as Worked Examples for Advanced Prompt Engineering Strategies*. ICIS 2024 Proceedings.

Tondello, G., Premsukh, H., & Nacke, L. (2018). A theory of gamification principles through goal-setting theory. *HICSS 2018 Proceedings*.

van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review, 6*(2), 95–108.

Vázquez Mendoza, J. A., Ramírez, C. F., & Aranda, C. M. (2024). Analysis and discovery of procrastination patterns in a language learning MOOC. *Computers & Education.* , Article 105154.. https://doi.org/10.1016/j.compedu.2024.105154

vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special issue editorial –accumulation and evolution of design knowledge in design science research: A journey through time and space. *Journal of the Association for Information Systems, 21*(3), 520–544.

Wang, Y., Derakhshan, A., & Zhang, L. J. (2021). Researching and practicing positive psychology in second/foreign language learning and teaching: The past, current status and future directions. *Frontiers in Psychology, 12*.

Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research & Development, 53*(4), 5–23.

Weber, F., Wambsganss, T., Rüttimann, D., & Söllner, M. (2021). *Pedagogical agents for interactive learning: A taxonomy of conversational agents in education*. ICIS Proc.

Winkler, R., Hobert, S., Salovaara, A., Söllner, M., & Leimeister, J. M. (2020). Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. *CHI 2020 Proceedings*.

Wittmann, M., & Morschheuser, B. (2022). What do games teach us about designing effective human-AI cooperation? - a systematic literature review and thematic synthesis on design patterns of non-player characters. *Proceedings of the 6th international GamiFIN conference*.

Wollny, S., Schneider, J., Di Mitri, D., Weidlich, J., Rittberger, M., & Drachsler, H. (2021). Are we there yet? - a systematic literature review on chatbots in education. *Frontiers in Artificial Intelligence, 4*, Article 654924.

Wong, J., Baars, M., He, M., de Koning, B. B., & Paas, F. (2021). Facilitating goal setting and planning to enhance online self-regulation of learning. *Computers in Human Behavior, 124*, Article 106913.

Wu, T.-T., Silitonga, L. M., & Murti, A. T. (2024). Enhancing English writing and higher-order thinking skills through computational thinking. *Computers & Education, 213*, Article 105012.

Yin, J., Goh, T.-T., Yang, B., & 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), 154–177.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. Verlag nicht ermittelbar.

Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist, 48*(3), 135–147.