Personalized Learning in Vocational Education

Inaugural Dissertation

to Obtain the Academic Degree of

a Doctor in Business Administration

at the University of Mannheim

submitted by

Herbert Thomann, M.Sc.

Mannheim

Dean:	Joachim Lutz
Referent:	Prof. Dr. Viola Deutscher
Co-referent:	Prof. Dr. Dirk Ifenthaler
Day of oral examination:	30.04.2025

Abstract

Personalized learning (PL) has become increasingly relevant in Vocational Education and Training (VET) due to changing workplace demands, the increasing heterogeneity of learner characteristics, and new opportunities presented by digital learning systems. Digital personalized learning (DPL) environments adapt to individual needs and provide targeted support through digital tools, promoting knowledge acquisition. Against this background, VET has undergone significant digital transformation, leading to the development of various personalized learning approaches. This dissertation examines the effectiveness of DPL approaches in VET through three interconnected studies. The first study presents a meta-analysis of 68 experimental studies on personalized digital learning prompts. It reveals that while prompts have a moderate positive effect on learning achievement (d = .394), publication bias analysis suggests a more conservative estimate (d = .22). Action-based prompts (d = .447) and group-targeted prompts (d = .513) significantly enhance learning achievement compared to standardized interventions. The second study's design-based research investigates the development and implementation of a personalized prompt design in the *Luca* Office Simulation. It demonstrates how cognitive, metacognitive, and non-cognitive prompts can be tailored to support individual learning processes based on log data in vocational business education. The third study evaluates personalization through immersive virtual reality (IVR) in VET using a randomized controlled trial with 72 students. Although IVR improved motivation, mood, and immersion, traditional methods were found to be more effective for immediate declarative knowledge acquisition. This highlights a significant discrepancy between perceived and actual learning gains. While revealing certain limitations in digital approaches, these findings suggest that successful personalization requires careful consideration of learning objectives, student characteristics, and design principles, while emphasizing the importance of balanced implementation approaches that combine traditional and digital learning methods. This research contributes to understanding how digital technologies can effectively support personalized learning in vocational education while acknowledging both their potential and limitations.

Acknowledgments

Wenn ich auf meinen bisherigen Weg zurückblicke, erscheint mir vieles davon noch immer überraschend und unwirklich. Auf diesem Weg haben mich viele Menschen begleitet, denen ich an dieser Stelle danken möchte.

An erster Stelle danke ich meiner Doktormutter, Prof. Dr. Viola Deutscher, die meine wissenschaftliche Entwicklung seit den ersten Schritten als studentische Hilfskraft im Bachelor bis zum Abschluss dieser Arbeit begleitet und gefördert hat. Danke für das Vertrauen von Tag eins an, die Unterstützung, und das wertvolle Feedback über all die Jahre. Als Mentorin hat sie nicht nur meine fachliche, sondern auch meine persönliche Entwicklung entscheidend beeinflusst. Die Balance zwischen gezielter Förderung und großen Freiheiten, die sie mir ermöglicht hat, hat wesentlich zum Gelingen dieser Arbeit beigetragen. Besonderer Dank gilt auch Prof. Dr. Dirk Ifenthaler der sich bereit erklärt hat, das Zweitgutachten zu meiner Dissertation zu verfassen.

Prof. Dr. Jürgen Seifried und Prof. Dr. Andreas Rausch danke ich für ihre konstruktiven Rückmeldungen und Ratschläge, die sie mir während des gesamten Forschungsprozesses zuteilwerden ließen. Die produktive Zusammenarbeit im „LUCA-Team" hat mir sehr viel Freude bereitet.

Darüber hinaus danke ich dem (ehemaligen) Mannheimer Lehrstuhlteam, Susanne Bechthold (für die unermüdliche Unterstützung in allen administrativen Belangen), Maximilian Krötz, Anke Braunstein und Ascha Ahmed sehr herzlich für die gemeinsame Zusammenarbeit und entspannte Atmosphäre. Den wissenschaftlichen Hilfskräften, insbesondere Julian Löhrer und Lea Dugrillon, danke ich für ihre zuverlässige Unterstützung. Allen Kolleginnen und Kollegen der AREA-Wirtschaftspädagogik danke ich für den fachlichen Austausch und die tollen gemeinsamen Konferenzreisen. Auch meinen neuen Kolleginnen und Kollegen in Göttingen danke ich für die herzliche Aufnahme und freue mich auf die kommende Zusammenarbeit.

Ein besonderer Dank gilt all meinen Freunden in Mannheim (und mittlerweile Mainz), sowohl aus der gemeinsamen Studienzeit als auch aus der gemeinsamen Zeit am Luisenring. Sie haben mich durch viele Höhen und Tiefen begleitet, waren für mich da und haben mir den nötigen Ausgleich zur akademischen Arbeit gegeben. Insbesondere danke ich meinem besten Freund, Mergim Mangjolli, dafür, dass er immer ein offenes Ohr für mich hatte und ich mich immer voll auf ihn verlassen konnte. Auch Dr. Verena Strempel danke ich, die mich zum Schritt meiner Promotion motiviert hat und ein unermüdlicher Sparringspartner auch in schwierigen Phasen war.

Meinen neuen Freundinnen und Freunden in Göttingen möchte ich für die wunderbar herzliche Aufnahme in ihren tierfreundlichen Kreis danken. Sie haben die finale Phase der Promotion bedeutend erleichtert.

Mein tiefster Dank gilt jedoch meiner Familie und insbesondere meinen Eltern. Dafür, dass ihr immer an mich geglaubt habt. Dafür, dass ihr alles in eurer Macht Stehende getan habt, damit wir es einmal besser haben. Ich kenne niemanden, der so hart wie ihr gearbeitet hat. Danke, dass ihr mir die Freiheit gegeben habt, einen Weg einzuschlagen, den ihr selbst nie gehen konntet. Ohne euch hätte ich es nicht bis hierher geschafft.

für meine Eltern

Table of Contents

List of Abbreviations

AI	Artificial intelligence
AR	Augmented reality
ASCOT	Technology-based assessment of skills and competences in vocational education and training *In German: Technologiebasierte Kompetenzmessung in der beruflichen Bildung*
ASCOT+	Technology-based assessment of skills and competences in vocational education and training (follow-up initiative of ASCOT) *In German: Technologiebasierte Kompetenzmessung in der beruflichen Bildung*
CAMIL	Cognitive-affective model for immersive learning
Cedefop	European center for the development of vocational training
CLT	Cognitive load theory
CTML	Cognitive theory of multimedia learning
DomPL-IK	Domain specific problem solving competence of industrial clerks *In German: Modellierung und Messung domänenspezifischer Problemlösekompetenz bei Industriekaufleuten*
DPL	Digital personalized learning
ECL	Extraneous cognitive load
FAIRI	Four-Component based, Adaptive, Immersive, Realistic, Intelligent
ICL	Intrinsic cognitive load
ITS	Intelligent tutoring system

IVR	Immersive virtual reality
LUCA	Learning by Using Competence Assessment
OECD	Organization of economic co-operation and development
PBL	Problem-based learning
PL	Personalized learning
PSA	Problem-solving analytics
SRL	Self-regulated learning
STEM	Science, technology, engineering, and mathematics
VET	Vocational education and training
VR	Virtual reality

List of Figures

List of Tables

List of Appendixes

1 Introduction

1.1 Motivation and Research Goals

Learning environments that address the needs of individual learners have been of great interest to learning designers and pedagogues since long before the emergence of digital learning (Plass & Pawar, 2020). For instance, educators usually personalize their teaching by providing additional assistance to struggling students and offering greater academic stimulation for high achievers (Holmes et al., 2018). The fundamental goal of providing each student with tailored learning experiences has gained new momentum with recent technological advances, particularly through log-file data analysis and artificial intelligence, creating unprecedented opportunities for digital learning environments (Hwang et al., 2020; Plass & Pawar, 2020).

Alongside this renewed emphasis on learning sciences, the modern labor market is undergoing a profound transformation. Demand for routine tasks is diminishing, while the need for skilled labor continues to grow, a trend highlighted in the "skill shift debate" (e.g., Cedefop & OECD, 2024; Lund et al., 2021; Binkley et al., 2012; Bughin et al., 2018). This shift results in increasingly complex work tasks, emphasizing the increasing importance of domain-specific problem-solving skills. Digital learning environments have emerged as valuable tools to address these changing circumstances (Rausch et al., 2024; Rausch et al., 2021).

These transformations in both education and the labor market are further complicated by increasing societal diversity, presenting additional challenges (Van Schoors et al., 2021; 2023). The intersection of technological change and growing diversity is particularly evident in large-scale educational settings where learners possess varying levels of prior knowledge and capabilities (Holmes et al., 2018). The Coronavirus Pandemic further accelerated these obstacles, forcing an unprecedented transition from face-to-face teaching to remote learning. This massive shift toward technology-driven education has made the question of personalized learning (PL) in digital environments more crucial than ever (Marzano et al., 2021). Consequently, there is an urgent need to shift from standardized, one-size-fits-all approaches toward more individualized and personalized learning solutions (Aleven et al., 2016; Cedefop, 2023; Van Schoors, 2021).

While these challenges affect all educational sectors, they are particularly pronounced in vocational education and training (VET), where the degree of diversity of learners is even more pronounced than in other educational systems (see Heinrichs & Reinke, 2019; Kremer et al., 2012).

Vocational competence as a central goal of VET includes not only technical skills but also action-oriented competence, which is a holistic framework that integrates methodological, personal, and social competencies for autonomous work and professional reflection (Deutscher & Winther, 2018; Frank & Schreiber, 2006; Klotz & Winther, 2016; Klotz, 2015). In recent years, developing professionals with these comprehensive competencies has become increasingly critical, particularly as the dual system in VET, prevalent in Germany and other nations, undergoes significant digital transformation (see Schumann et al., 2022). Modern VET students are required to adapt to the increasing digitalization in their industries, necessitating proficiency in contemporary digital tools and software for their professional responsibilities (Harteis & Billett, 2023). However, traditional uniform educational practices often fall short in addressing diverse student needs and adapting to this rapidly evolving job landscape. This limitation has given rise to personalized learning systems, establishing them as a legitimate educational framework for vocational education, specifically designed to accommodate learners' diverse demands while enhancing their learning achievements (Zhao et al., 2024; Rausch et al., 2021).

Given these complex challenges in VET, personalized learning represents a particularly promising solution within digital environments (Maier & Klotz, 2022; Kabudi et al., 2021). Creating a personalized digital learning environment that can self-adapt to provide learning support for different types of learners can overcome “the weakness of one-size-fits-all approaches” (Shemshack & Spector, 2020, p. 3). When properly designed and implemented, these environments can significantly contribute to learning achievements by dynamically adapting to learner behavior, mirroring strategies long employed by human tutors (Shemshack et al., 2020; Shemshack & Spector, 2021). Successful digital personalized learning environments have a significant impact on learners' anticipations as well as their ability to acquire, manipulate, construct, generate, and convey knowledge (see Dillenbourg et al., 2002; Green & Donavan, 2018). These include intelligent tutoring systems, computer-based simulations, virtual reality settings, educational games, and a variety of others.

Among these technological approaches, computer-based learning simulations have emerged as particularly effective tools for developing domain-specific problem-solving skills (see Chernikova et al., 2020; 2023; 2024). Their effectiveness stems from their ability to combine personalized learning experiences with authentic workplace scenarios, making them especially valuable for

VET (Plass & Pawar, 2020). In commercial training for industrial and office management clerks, computer-based office simulations such as *ALUSIM* (Sangmeister et al., 2019; Winther et al., 2016; Winther & Achtenhagen, 2010; Winther & Achtenhagen, 2009), *DomPL-IK* (Seifried et al., 2016) and *Luca* office (Rausch et al., 2021) have been developed with authentic problem-based working scenarios (see Braunstein et al., 2022). These simulations are part of the "Technology-Oriented Competence Measurement in Vocational Education and Training" research initiatives "ASCOT" and "ASCOT+" (Beck et al., 2016).

Additionally, adaptive digital learning environments with personalized support are integral to the cultivation of problem-solving competencies. Recent advances have enabled the seamless integration of scaffolding through 'prompts' in digital learning platforms (Serge et al., 2013; Puntambekar & Hubscher, 2005). Appearing as strategic questions, targeted suggestions, and timely support, these prompts actively promote the application of effective processing strategies (Bannert, 2009; Wirth, 2009). By providing tailored assistance across different proficiency levels, prompts have the potential to markedly improve the educational experience and accommodate diverse learner needs (Schumacher & Ifenthaler, 2021; Ifenthaler, 2012; Davis, 2003).

Another promising avenue of research in VET is utilizing immersive virtual reality (IVR) simulations (Radianti et al., 2020). IVR's unique features may be especially beneficial in VET for developing domain-specific knowledge, action-oriented skills, and applying specialized knowledge in new work contexts (Buchner & Mulders, 2020; Conrad et al., 2022; Mulders, 2022). Moreover, empirical studies consistently reveal that IVR surpasses conventional methodologies (such as paper-and-pencil approaches) in enhancing learner engagement, motivation, and immersive experiences (see Kolarik et al., 2024; Makransky & Klingenberg, 2022; Makransky & Lilleholt, 2018; Parong & Mayer, 2018).

Despite these promising developments in personalized learning technologies for VET, several critical gaps remain. These gaps span across the key technological approaches discussed above and require systematic investigation. First, while numerous studies have examined digital learning prompts, there is a notable lack of systematic understanding regarding which prompt features are most effective and under what conditions (Renkl & Scheiter, 2017). Existing research tends to focus on individual implementations rather than providing comprehensive analyses of prompt design principles that could guide future developments in vocational education.

Second, despite the recognized potential of computer-based simulations, there is a scarcity of research on creating and effectively utilizing personalized prompts in these settings (Guo, 2022; Zheng et al., 2022), especially for developing vocational competencies. Both earlier and recent studies primarily used standardized prompts, meaning that the prompt content remained identical for each learner, and the design of this content was mostly based on insights drawn from existing literature (Lim et al., 2024; Li et al., 2023; Guo, 2022; Bannert, 2009).

Third, while IVR shows promise for vocational education, particularly in developing skills and competencies, research on its effectiveness compared to traditional teaching methods remains limited and often contradictory (e.g., Conrad et al., 2024; Matovu et al., 2023).

Addressing these gaps is crucial for advancing the understanding of how digital technologies can best serve VET. Therefore, this dissertation seeks to (1) explore how prompts can serve as a tool for personalization in digital environments (e.g., *Luca* office) and identify which prompt characteristics most effectively enhance learning achievement. Moreover, it intends to (2) empirically assess an IVR simulation against traditional learning methods to evaluate their effects on cognitive and affective learning results. With these overarching aims, the dissertation seeks to illuminate the role of personalized learning in digital learning environments. Consequently, three distinct studies were conducted, which are detailed in the following section.

1.2 Outline and Research Questions

The dissertation explores the multifaceted nature of personalized learning in Vocational Education and Training through three interconnected studies. As illustrated in Figure 1, it progresses systematically from a foundational understanding of digital prompts to their practical implementation in both computer-based simulations and immersive virtual reality.

To visualize this research structure, Figure 1 illustrates the research progression linking the three papers, demonstrating how this dissertation advances from fundamental principles of digital prompting to specific VET applications while maintaining personalized learning as the unifying concept. According to Walkington and Bernacki's (2020) framework, which will be detailed in Section 2.3, PL is conceptualized not as a learning theory but as an overarching method that leverages existing theories to adapt educational environments to learner needs.

Following this conceptual structure, Paper 1 provides an exhaustive overview of the effectiveness of digital learning prompts across all domains and learning environments through a systematic review and meta-analysis (manuscripts 1 & 2). It establishes the theoretical and empirical basis for understanding how prompts can be used as a personalization tool and what prompt features are most effective in improving learning achievement (Figure 1, orange ellipse).

Applying these theoretical insights to practice, Paper 2 (Figure 1, blue line) employs a design-based research approach to develop and implement a personalized prompt design within the *Luca* computer-based simulation (learning environments) for vocational education. It focuses on how log data (learner characteristics) can be utilized to create adaptive prompting strategies.

Paper 3 (Figure 1, green line) empirically compares an adaptive immersive virtual reality (IVR) environment to traditional learning methods in vocational logistics education. The focus is on examining the impact on both cognitive and affective learning outcomes.

Figure 1: Research outline

(based on the model of personalized learning by Walkington & Bernacki, 2020)

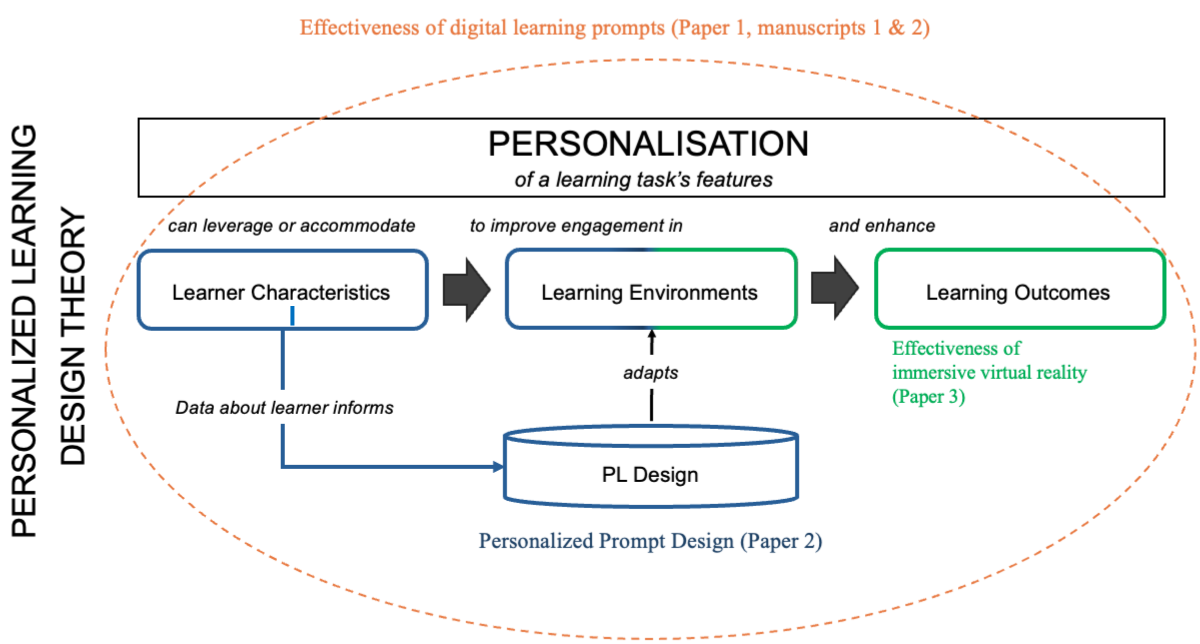


Table 1 below presents a concise overview of the three papers, including aspects such as publication status, research questions, data collection, samples, and methodologies.

Paper 1 forms the foundation of this dissertation. It consists of two manuscripts that present a systematic review and meta-analysis of digital learning prompts. This study addresses three key research questions: the types of prompts distinguished in the literature (regardless of the type of learning environment or domain), the overall effectiveness of prompts in enhancing learning achievement, and how prompt effectiveness is moderated by various features and study demographics.

Building on these findings, Paper 2 applies the concept of personalized prompts to a specific VET context. Using design-based research, it develops and implements a personalized prompt design within the *Luca* office simulation. While not explicitly stating research questions, this study demonstrates how insights from Paper 1 can be practically applied to create personalized learning experiences in a simulated environment. This study was conducted as part of the ASCOT+ project "PSA-Sim – Problem Solving Analytics in Office Simulations." Further research efforts (beyond this dissertation) included testing this prompt design, and the effects are currently under analysis.

Extending this investigation into another digital learning environment, Paper 3 explores personalized learning by examining immersive virtual reality. It presents an empirical study involving logistics trainees that compares the adaptive IVR environment *InGo* with traditional learning methods (paper-pencil). This research tackles questions concerning IVR's effectiveness in acquiring declarative knowledge, the link between objective and perceived learning, and its effects on learners' mood and motivation.

Collectively, this research progression examines personalized learning in VET. It starts with a broad understanding of digital prompts, moves to their specific application in a computer-based commercial office simulation, and culminates in comparing an IVR environment with traditional methods in warehouse logistics.

Drawing these elements together, the findings derived from the three studies ought to function as a guide, particularly for educators or technological facilitators, to inform their personalized learning strategies within digital learning environments, thereby facilitating the individualized learning trajectories of their students.

Table 1: Paper overview and publication status

Study	**Paper 1 (Manuscript 1 & 2)**	**Paper 2**	**Paper 3**
Reference	Thomann, H., & Deutscher, V. (2025). Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement. *Educational Research Review*, 47, 100686.	Deutscher, V., Seifried, J., Rausch, A., Thomann, H., & Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen-und Lehrerbildung-didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen (Vol. 68, pp. 107-121). *wbv*.	Thomann, H., Zimmermann, J., & Deutscher, V. (2024). How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects. *Computers & Education*, *220*, 105127.
Research questions	(1) Which types of prompts are distinguished throughout the literature? (2) What is the overall effect of prompts on learning achievement? (3) How is the effectiveness moderated by prompt features and study demographics?	What prompt designs can be exemplarily developed within the *Luca* office simulation?	(1) Does objectively and subjectively measured declarative knowledge acquisition differ between IVR and paper-based learning approaches? (2) How strong is the relation between objective knowledge acquisition and subjectively perceived knowledge acquisition in both test settings (paper-based versus IVR)? (3) To what extent do differences exist between the IVR and paper-based groups regarding mood, intrinsic motivation, and immersion during task completion?
Data & Methods	• PRISMA literature search and selection (manuscript 1) • What works Clearinghouse (WWC) • Meta-analysis (manuscript 2): • Sensitivity analysis • Publication bias • Correlation analysis • Moderator (regression) analysis • Graphical illustrations	• Design-based Research	• Questionnaire data • Power analysis • Testing for homogeneity of variances and normality • Independent sample t-tests and Welch t-tests • Correlation analysis • Holm-Bonferroni correction
Sample/Participants/ included studies	68 peer-reviewed journal articles (70 effect sizes) published between 1999 and 2022. The average sample size of the included studies was n = 98 (range 26 - 656).	Phase 1: 3 research scientists (DBR-Team) Phase 2: DBR-Team, 4 assistants, 2 teachers Phase 3: 647 VET students Phase 4: 222 VET students	72 participants (logistic trainees from four classes)
Data Collection	Own data collection	Part of the ASCOT+ Project “PSA-Sim – Problem Solving Analytics in Office Simulations”	Own data collection and contribution of master thesis student.

1.3 Structure of the Thesis

This inaugural dissertation is structured into five chapters. Chapter 1.1 introduces the motivation and significance of personalized learning approaches within digital learning environments in the context of VET. Additionally, research questions are developed and presented in Chapter 1.2, while the structure of the thesis is outlined in Chapter 1.3.

Chapter 2 of this dissertation serves as a framework for all three papers, detailing the theoretical underpinnings of the research. It covers the comprehension and approach to (digital) personalized learning (Section 2.1), reviews existing literature, identifies research gaps (Section 2.2), elaborates on the design process framework (Section 2.3), and examines relevant learning theories (Section 2.4). Emphasis is placed on cognitive load theory, the cognitive theory of multimedia learning, and the associated theory of interest. Following this, prompting will be discussed in detail as a personalized learning strategy (Section 2.5) rooted in the theoretical prompt model established in Paper 1 (Section 2.5.1) and the associated prompt feature categories (Section 2.5.2). Subsequently, personalized prompts within the computer-based simulation *Luca* office (Paper 2), guided by the 4C/ID Model, will be further developed in the context of VET (Section 2.6.1 and 2.6.2). The final section (2.6.3) presents a modified FAIRI framework (Section 2.6.4) that combines the prompt model with the 4C/ID model, tailored for personalized, immersive virtual reality environments in VET (see Paper 3).

Chapter 3 outlines the methodological framework, detailing the operationalization of both objective and subjective learning outcomes, emotional states, motivation, and immersion, alongside the statistical techniques utilized. Additionally, it elaborates on the design-based research method employed in paper 2.

The fourth chapter constitutes the primary section of this dissertation. It summarizes the three papers before presenting them in their original formats, with the exception of paper 1, which is included as a manuscript originally submitted in March 2024.

Chapter 5 outlines essential findings, underscoring their scientific and practical significance, limitations, and future research directions. It moreover delineates possible upcoming studies in a research outlook designed to empirically test *Luca* office and the respective prompt design from Paper 2, along with another IVR study that incorporates insights from Paper 3.

2 Theoretical Foundation

Personalized learning in vocational education, particularly when enhanced by digital tools, is grounded in several interconnected theoretical frameworks. This chapter begins by exploring the concept of personalized learning, examining the theoretical foundations that support research on digital prompts, the *Luca* office simulation, and the immersive virtual reality environment *InGo*, as discussed in the three papers in Chapter 4.

2.1 The Concept of (Digital) Personalized Learning

> *The parable of the elephant and the blind men illustrates how fragmented perspectives can obscure a unified understanding. Similarly, in education, personalized learning encompasses diverse approaches—each addressing a part of the whole. To grasp its full potential, these perspectives must be integrated, crafting a comprehensive framework that meets the multifaceted needs of learners.*

This integration is particularly relevant to understanding personalized learning's evolution as a longstanding topic in educational research. At its core, personalization is the process of creating tailored solutions to meet individual needs (Cheung et al., 2021). This approach has been prevalent for centuries, since well before the advent of digital learning linked to apprenticeship and mentoring (Bernacki et al., 2021; Shemshack & Spector, 2020). While learning designers and educators consistently strive to personalize their students' learning experiences, the international research literature reveals a complex landscape of terminology and definitions (Schmid & Petko, 2019). For example, in their review, Shemshack and Spector (2020) identified several terms, including “personalized learning,” “adaptive learning,” “individualized learning,” “customized learning,” and more. Nevertheless, the majority of existing literature focuses mainly on the terms adaptive and personalized learning (Hooshyar et al., 2024).

Bernacki et al. (2021) analyzed ten definitions of PL from authoritative sources, including the OECD, revealing substantial variations and conceptual ambiguities. While these definitions uniformly integrated attributes of learners, educational outcomes, and design frameworks, they varied significantly in their specific components and emphasis. The most influential definition comes from the US Department of Education: “[…] instruction in which the pace of learning and the instructional approach are optimized for the needs of each learner. Learning objectives, instructional approaches, and instructional content (and its sequencing) all may vary based on learner needs" (2016, p. 7). In contrast, adaptive learning dynamically modifies learning paths and scaffolds (prompts) in response to real-time performance, often utilizing data-driven learner modeling to optimize educational experiences (Shemshack & Spector, 2020). Although these concepts share a common goal of individualized instruction, they emphasize different dimensions. Personalized learning focuses on broader learner characteristics and agency, while adaptive learning primarily addresses performance-based adjustments (Hooshyar et al., 2024).

The metaphor of the elephant and the blind men aptly illustrates the fragmented yet interconnected nature of these definitions. Each framework, like the perspectives of the blind men, captures a vital but partial truth about the comprehensive nature of personalized learning. For instance, adaptive learning aligns with the tactile recognition of one part of the elephant, emphasizing precision in performance-driven adaptations. Personalized learning, on the other hand, embodies the broader holistic understanding, incorporating learner autonomy, goals, and socio-emotional characteristics. Together, these perspectives provide complementary insights into a unified vision of effective, learner-centered education.

The discourse surrounding digital technology highlights its potential for facilitating personalized learning experiences, a development that has gained significant attention in educational research (Lin et al., 2024; Zheng et al., 2022; Major et al., 2021; Van Schoors et al., 2021; Xie et al., 2019). Similar to the concept of personalized learning, a universally accepted definition of digital personalized learning also continues to be ambiguous (Major et al., 2021). Technology-facilitated or digital personalized learning (DPL) refers to the utilization of technology to tailor education to individual learner characteristics and requirements (Major et al., 2021; Van Schoors et al., 2021). In order to surpass the efficacy of one-size-fits-all approaches, personalized learning demands that a digital learning environment addresses the learner and a mix of their learning characteristics (e.g., prior knowledge, experience, or motivation). This can establish a learning experience attuned to these characteristics, promoting higher engagement with and performance of a learning task (Shemshack & Spector, 2020; Aleven et al., 2016).

The variability in definition impedes progress in research on DPL. It tends to be used as an ever-expanding umbrella term, and a clearly defined concept for (digital) personalized learning remains absent (Bulger, 2016). This complicates "the study of PL and the ways that designs can leverage student characteristics to reliably achieve targeted learning outcomes” (Bernacki et al., 2021, p. 1675). Therefore, this dissertation follows Hooshyar et al.'s (2024) understanding by referring to adaptive and personalized learning collectively as (digital) personalized learning (PL; DPL), reflecting their shared goal of catering to diverse learner needs. The focus is on PL implemented through personalized prompts and immersive virtual reality (IVR) environments, as examined in the three papers of this dissertation. These approaches highlight the convergence of theoretical constructs and practical innovation in addressing diverse educational needs in vocational contexts.

2.2 Existing Reviews and Research Gaps

The landscape of personalized learning research has evolved significantly, as revealed by both Hooshyar et al. (2024) and Lin et al. (2024). In the past five years, 18 reviews and meta-analyses have been published, addressing various aspects of PL effectiveness (see Table 2 for an overview).

Examining previous work in the field, several systematic reviews laid important groundwork. Xie et al. (2019) initiated systematic examinations by reviewing trends from 2007-2017, identifying limitations in device personalization across different platforms. Around the same time, Martin et al. (2020) focused on adaptive learning designs, examining methodologies and outcomes from 2009-2018.

Plass and Pawar (2020) developed a framework for adaptivity in learning, which significantly advanced theoretical understanding. Shemshack and Spector (2020) emphasized the ambiguity in defining personalized learning components, an ongoing challenge in the field. Zhang et al. (2020a, 2020b) examined various technologies facilitating personalized learning and contextual implementation factors.

Bernacki et al. (2021) advanced theoretical understanding by exploring diverse definitions and examining educational theories shaping development and implementation. Their work emphasized personalized learning's interdisciplinary nature and advocated for theory-guided design approaches. Moreover, their work aligned with Li and Wong's (2021) analysis of features and trends in personalized learning environments.

This theoretical framework was enhanced by Alamri et al. (2021), who pinpointed effective technology models in higher education, and by Van Schoors et al. (2021), who investigated differentiated learning in technology across primary and secondary education.

Major et al. (2021) and Zheng et al. (2022) provided early meta-analytic evidence, though with differing scopes and results. Major et al. (2021) focused on low-income countries, demonstrating positive learning outcomes though limited to 15 articles, while Zheng et al. (2022) found moderate effects on achievement but lacked quality appraisal measures.

Zhong et al. (2022) examined higher education contexts specifically, while Komalawardhana and Panjaburee (2023) reviewed DPL specifically in science education from 2010 to 2022. Their review illuminated patterns in the utilization of technology to enhance applied science education, encompassing individualized frameworks and domains of application. Li and Wong (2023) focused on STEM domains, offering additional insights into domain-specific applications.

Hooshyar et al. (2024) conducted a comprehensive meta-analysis addressing previous limitations by examining both personalized and adaptive learning approaches up to 2023. Their study incorporated gray literature, followed systematic review guidelines, and considered both cognitive and non-cognitive outcomes. Lin et al. (2024) complemented this work by specifically focusing on interest-based personalization from 34 publications. Their findings revealed medium-to-large effects on interest, cognitive load, retention, and transfer. The effectiveness was influenced by the diagnostic approach, grain size, and domain of interest. Additionally, geographical location and experimental design moderated retention effects.

Research shows that digital personalized learning is experiencing significant growth, not only due to its anticipated capacity to boost student motivation and learning results, as well as its effectiveness in helping teachers manage diverse classroom settings (see Lin et al., 2024; Van Schoors et al., 2023, 2021; Zheng et al., 2022; Bernacki et al., 2021; Xie et al., 2019). DPL environments include various inherent personalization mechanisms. These mechanisms create unique learning experiences by providing adaptive suggestions and personalized content for learners (Shemshack & Spector, 2020; Van Schoors et al., 2022). Numerous authors have established frameworks to elucidate the complexities of DPL (for an overview, see Van Schoors et al., 2022; Zheng et al., 2022; Vandewaetere & Clarebout, 2011).

Together, these reviews and meta-analyses reveal consistent themes and notable gaps. Personalized learning demonstrates potential across multiple outcomes, yet its success depends on the implementation method and context. Despite numerous systematic reviews within PL research, significant methodological limitations persist. Key constraints include limited time frames (Xie et al., 2019; Shemshack & Spector, 2020), selective database coverage (Xie et al., 2019), and the omission of gray literature (Zheng et al., 2022). Most reviews do not adequately assess the quality of the studies they include. Only Shemshack and Spector (2020) and Major et al. (2021) fulfill this essential requirement outlined in PRISMA (Page et al., 2021). Earlier meta-analyses also exhibit limitations. Major et al. (2021) concentrated exclusively on low-income countries through 15 articles, while Zheng et al. (2022) neglected quality appraisal and gray literature. Both overlooked adaptive learning's role and non-cognitive factors. From a theoretical perspective, Bernacki et al. (2021) highlighted that utilizing theoretical frameworks can enhance personalized learning's effectiveness, advocating for theory-guided design approaches. Lin et al. (2024) corroborated this by illustrating how interest theory and cognitive load theory offer crucial theoretical insights for comprehending the effects of personalized learning. These gaps underscore the necessity for thorough search strategies, stringent quality assessment processes, theory-based approaches, and a broader examination of both cognitive and non-cognitive outcomes in DPL research. Consequently, the next chapter elaborates on Bernacki et al.'s (2021) personalized learning design framework to establish a robust theoretical foundation for this dissertation.

Table 2: Summary of personalized learning reviews and meta-analyses published in the past five years (adapted from Lin et al., 2024)

Review/Meta-analysis	Terms Used	Search Strategy	Foci/Frameworks
Xie et al. (2019) in *Computers & Education*	Adaptive learning, personalized learning	Journal articles from 2007 to 2017 in Web of Science	Learning support, system parameters, learning outcomes
Martin et al. (2020) in *Educational Technology Research and Development*	Adaptive learning	Journal articles from 2009 to 2018 in Education Research Complete and ERIC	Learner model, content model, instructional model
Plass and Pawar (2020) in *Journal of Research on Technology in Education*	Adaptivity for learning, personalization	N/A	Cognitive, motivational, affective, sociocultural variables
Shemshack and Spector (2020) in *Smart Learning Environments*	Personalized learning, adaptive learning, individualized instruction, customized learning	Journal articles from 2010 to 2020 in Scopus, Science Direct, EBSCOhost, IEEE Xplore, JSTOR, Web of Science, Google Scholar	Personalized learning terms
Zhang et al. (2020a, 2020b) in *Educational Research Review* and *Journal of Research on Technology in Education*	Personalized learning	Journal articles from 2006 and 2017 in ERIC, OmniFile Full-Text Select, Academic Search Complete, Web of Science	Various technologies facilitating personalized learning and contextual factors of its implementation
Bernacki et al. (2021) in *Educational Psychology Review*	Personalized learning, adaptivity	Journal articles, dissertations, and conference papers from 2010 to 2018 in ERIC, PsychInfo, and IEEE Xplore	Learner characteristics, design elements, outcomes

Review/Meta-analysis	Terms Used	Search Strategy	Foci/Frameworks
Li and Wong (2021) in *Interactive Learning Environments*	Personalized learning	Journal articles from 2001 to 2018 in Scopus	Features and trends
Major et al. (2021) in *British Journal of Educational Technology*	Personalized learning, personalized adaptive learning	Journal articles and reports from 2007 to 2020 in ERIC, Directory of Open Access Journals, ProQuest, Scopus, Web of Science	Technology-supported learning for school-aged learners in low- and middle-income countries
Van Schoors et al. (2021) in *British Journal of Educational Technology*	Digital personalized learning, adaptive learning	Journal articles from 1995 to 2020 in ERIC and Web of Science	Time, target, method, source, and context in primary and secondary education
Zhang et al. (2022) in *Journal of Computer Assisted Learning*	Personalized learning	Journal articles from 2006 and 2020 in ERIC, OmniFile Full-Text Select, Academic Search Complete, Web of Science	Instructional design
Zheng et al. (2022) in *Education and Information Technologies*	Personalized learning, individualized learning, customized learning	Journal articles from 2001 and 2020 in Web of Science, Scopus, and ERIC	Personalized learning facilitated by technologies
Zhong (2022) in *Interactive Learning Environments*	Personalized learning, adaptive learning	Journal articles and conference papers from 2001 and 2020 in IEEE Xplore, Scopus, ERIC, Web of Science, JSTOR, ProQuest	Learning content structuring, learning material sequencing, and readiness support in higher education
Alamri et al. (2023) in *Tech Trends*	Personalized learning, adaptive learning, individualized instruction	Journal articles from databases (e.g., IEEE Xplore, Scopus) published from 2000–2020	Technology-supported models that adapt to learners' needs based on cognitive, motivational, and contextual factors
Komalawardhana & Panjaburee (2023) in *Journal of Computers in Education*	Personalized learning, technology-enhanced personalized learning, adaptive learning	106 studies (2010–2022) sourced from Scopus database; focused on journal articles only	Parameters like personalized learning content, learning paths, diagnosis, suggestions, and platforms. Addressed trends, development, and effectiveness in science education
Li and Wong (2023) in *Journal of Computing in Higher Education*	Personalized learning, personalization	Journal articles from 2011 to 2020 in Scopus	STEM domain
Lin et al. (2024) in *Educational Psychology Review*	Personalized learning	Journal articles, dissertations, and conference papers from as early as possible to 2024 in Academic Search Ultimate, ERIC, PsychInfo, OpenDissertations, ACM Digital Library, IEEE Xplore	Personalized learning by interest—tailored to learners' contexts and familiar interests
Hooshyar et al. (2024) in *Computers & Education*	Personalized learning, adaptive learning, technology-enhanced learning	47 studies included, spanning from early 2000s to 2023, using databases such as WOS, Scopus, ScienceDirect, and Google Scholar (gray literature included)	Investigated personalized learning impact on cognitive and non-cognitive outcomes, moderating factors (e.g., research settings, delivery mode, learner models), and instructional strategies using meta-analysis and association rule mining

2.3 Personalized Learning Design Process Framework

Bernacki and colleagues (2021) highlight the complexity and multifaceted nature of personalized learning as both a research topic and an educational approach. The authors emphasize that personalized learning involves considering a wide array of learner characteristics, including "prior knowledge, motivations, goals, beliefs, interests, skills, experience, and culture" (p. 1676). These factors should inform the design of instructional experiences that are responsive to individual learner needs and promote engagement and performance.

The multidimensional nature of PL has attracted researchers from diverse disciplines, resulting in a large and disparate body of research. This diversity, while valuable for its breadth of perspectives, presents challenges in synthesizing findings and developing a coherent understanding of personalized learning's effects and best practices. The authors argue that this complexity requires a systematic approach to studying and implementing personalized learning.

The personalized learning design process proposed by Walkington and Bernacki (2020) offers a robust framework (Figure 2) for addressing the complexities of PL research. This model provides a structured approach to conceptualizing PL, grounding it in classical instructional design while incorporating personalization elements. The framework considers how learner characteristics inform adaptations to the learning environment, how these adaptations change from a base mode of instruction (i.e., non-personalized), and how these changes are intended to achieve specific outcomes.

It surpasses previous reviews, such as Xie et al. (2019), by offering a more comprehensive and theoretically grounded examination of PL designs. The framework aligns with established process models in design decision-making (Beese, 2019; Reigeluth et al., 2015), providing a precise structure for describing and analyzing PL approaches. Adopting this framework aids researchers in distinguishing between different types of personalized instruction instead of grouping them under one term in PL research (Bernacki et al., 2021). This is particularly important given the diverse implementations and definitions of personalized learning across educational settings and research disciplines. This comprehensive framework not only provides structural clarity for research but is also grounded in fundamental theoretical principles that guide its implementation.

At its core, the model integrates two fundamental concepts: the overarching principles of learning theories and the specific propositions of PL and adaptivity theories (Walkington & Bernacki, 2020). Learning theories typically posit that the attributes of learners affect their engagement in learning environments and the outcomes achieved (Figure 2; above broken line). Building upon this foundation (Figure 2; under broken line), “theories of PL and adaptivity propose that information about a learner, derived from data that is available or choices they make, can be used to adapt features of the learning environment to enhance learning outcomes” (Walkington & Bernacki, 2020, p. 240).

The process begins with an appraisal of one or more learner characteristics (Figure 2, bottom left). These characteristics may include cognitive factors such as prior knowledge, motivational aspects like interests or goals, or other individual differences. This information serves as the basis for adapting the learning environment.

Next, the model describes how the learning environment (e.g., the *Luca* office in Paper 2 or the IVR environment *InGo* in Paper 3) adapts to these learner characteristics (Figure 2, bottom middle). This adaptation represents a departure from a base mode of instruction and can manifest in various ways, such as adjusting content difficulty, scaffolding (e.g., digital learning prompts in Papers 1 and 2), altering the presentation of information, or modifying the sequence of learning activities.

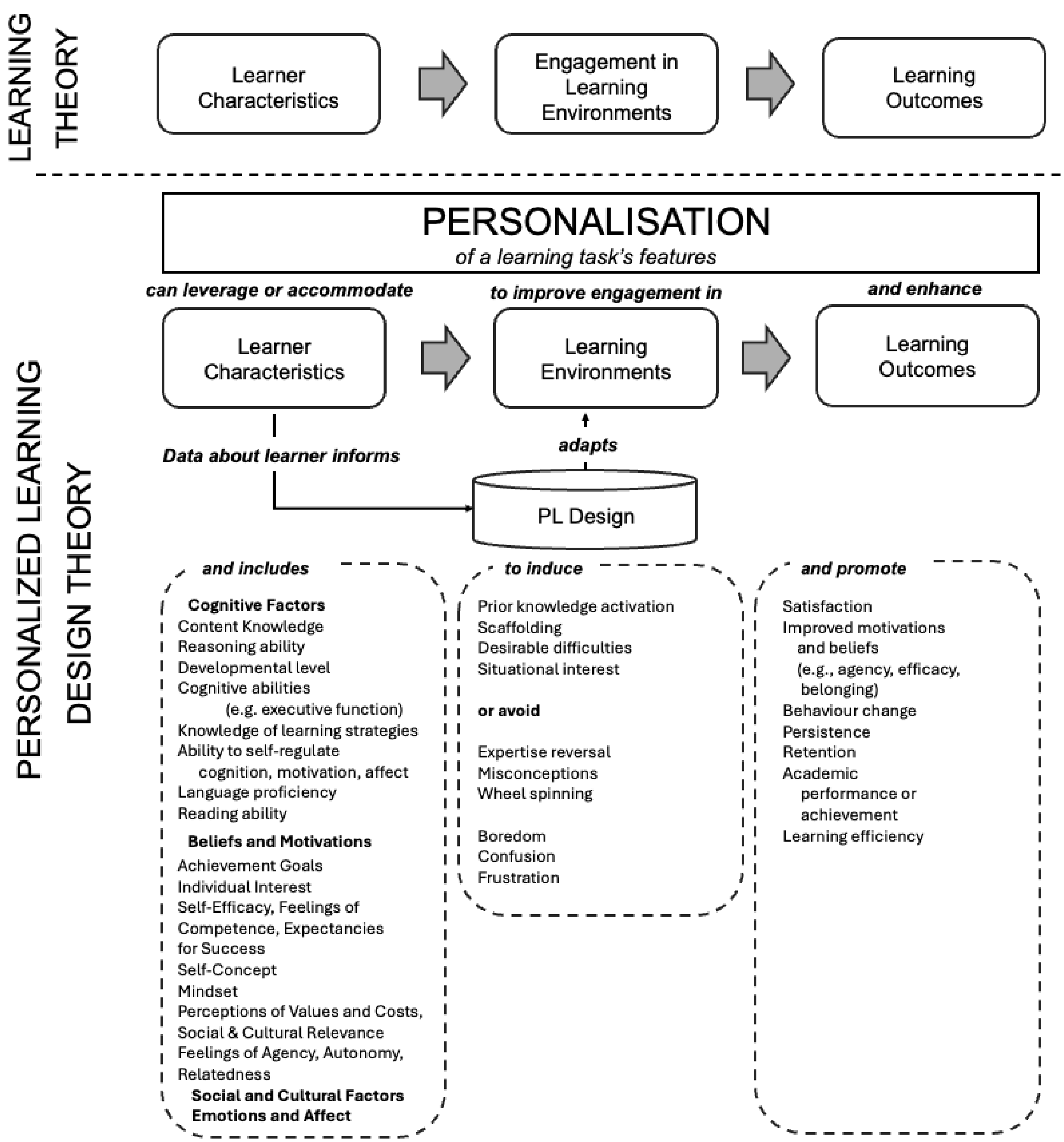
Finally, the model emphasizes that these adaptations are made to achieve specific learning outcomes (Figure 2, bottom right). These outcomes may include improved academic performance, increased engagement, enhanced motivation, or other desired learning results.

Importantly, the model illustrates this process as cyclical and iterative. As learners interact with the (digital) personalized environment, new information about their characteristics and performance is gathered, which can then inform further adaptations to the learning experience.

The model also acknowledges the complexity of personalized learning by recognizing that multiple learner characteristics may inform multiple adaptations, which in turn may target multiple outcomes. These multiple interconnected relationships reflect the intricate nature of personalized learning designs in theory and practice (Bernacki et al., 2021; Walkington & Bernacki, 2020; Xie et al., 2019).

Figure 2: Personalized learning design process model

(adapted from Bernacki et al., 2021; Walkington & Bernacki, 2020)



2.4 Learning Theories Relevant to Personalized Learning

Several established learning theories are particularly relevant to informing the design and implementation of personalized learning approaches (see Figure 2). Table 3 presents a concise synthesis of a sampling of learning theories that offer theoretical foundations for personalized learning approaches. While not exhaustive, this compilation represents an illustrative set of relevant theoretical frameworks that can inform the conceptualization and implementation of PL approaches (see Bernacki et al., 2021; Walkington & Bernacki, 2020). These theories serve as exemplars of the diverse conceptual foundations upon which personalized learning initiatives can be constructed and evaluated.

As Bernacki et al. (2021) note, these theories span cognitive, metacognitive, motivational, and affective domains. In the cognitive realm, theories such as mastery learning (Block & Burns, 1976), cognitive load theory (Sweller, 2011), and cognitive theory of multimedia learning (Mayer, 2001) provide frameworks for adapting instruction based on learners' prior knowledge and ongoing performance. The expertise reversal effect (Kalyuga, 2007) highlights the importance of tailoring support to learners' evolving knowledge levels. Metacognitive theories of self-regulated learning (Zimmerman & Schunk, 2011) emphasize how personalization can leverage and develop learners' metacognitive knowledge and skills.

Motivational theories, including achievement goal theory (Elliot, 1999), interest development theory (Hidi & Renninger, 2006), self-efficacy theory (Bandura, 1986), expectancy-value theory (Eccles & Wigfield, 2020), and self-determination theory (Deci & Ryan, 2000) offer insights into how personalization can enhance learner motivation through accommodating individual goals, interests, efficacy beliefs, and requirements for independence, competence, and relatedness. Control-value theory (Pekrun & Perry, 2014) elucidates the impact of personalization on learners' emotions in the affective domain during academic tasks.

Building upon the overview of learning theories relevant to personalized learning (Table 3), the following sections delve into a more detailed examination of specific theoretical frameworks that are particularly pertinent to this dissertation. Across all three papers, two cognitive theories—the Cognitive Load Theory (Sweller, 2011) and the Cognitive Theory of Multimedia Learning (Mayer, 2001)—serve as central conceptual pillars. These theories provide robust explanations for understanding how personalized learning interventions may influence cognitive processing and learning outcomes in digital learning environments.

The third paper expands the theoretical framework to incorporate motivational perspectives (Deci & Ryan, 2000), and specifically Interest Theory (Hidi & Renninger, 2006). This inclusion allows for a more comprehensive analysis of how personalization affects not only cognitive processes but also learners' affective and motivational states.

Table 3 Learning theories relevant to personalized learning (adapted from Bernacki et al., 2021)

Learning Theory	Central Thesis	Key Learner Characteristics	Focal Outcomes
(Meta)cognitive theories			
Mastery learning (Block & Burns, 1976)	Learners' current knowledge should inform selection of next tasks; feedback and support should be timely, specific	Prior knowledge and in-task performance	In-task performance, skill mastery, learning efficiency
Expertise reversal (Kalyuga, 2007)	Support benefits learners with low prior knowledge, undermines those with high	Prior knowledge	In-task performance
Cognitive load (Sweller, 2011)	Capacity is limited; extraneous load should be reduced to afford germane processing	Working memory capacity	Attention, performance
Cognitive theory of multimedia learning (Mayer, 2001)	Extends cognitive load theory principles to multimedia learning environments	Working memory capacity	Key principles have direct applications in technology-enhanced learning
Self-regulated learning (Zimmerman & Schunk, 2011)	Learners bring prior knowledge, skill, goals, and agency; can plan and enact strategies, monitor and adapt learning	Metacognitive knowledge of learning skills, prior knowledge, goals, motivation	Goal attainment, motivation, persistence, academic performance
Motivation theories			
Achievement goals (Elliot, 1999)	Learners may aim to improve/avoid decrease in mastery, performance	Achievement goals	Strategy use, persistence, achievement
Interest development (Hidi & Renninger, 2006)	Learners bring interests that are triggered and maintained by task, mature and change over time	Individual interests	Engagement, persistence, knowledge development, achievement
Self-efficacy (Bandura, 1986)	The belief that a learner can succeed in learning affects engagement, success	Prior personal, vicarious experiences and success in tasks	Engagement, persistence, achievement
Expectancy value (Eccles & Wigfield, 2020)	Learners appraise tasks to determine expectations, values and costs	Expectancy for success; utility, intrinsic, attainment value; effort, opportunity and psychological cost	Satisfaction, persistence, academic achievement
Self-determination (Deci & Ryan, 2000)	Learners are autonomous and motivated by choice; they thrive when they feel competent and that they belong	Ability to choose, affinity informing feelings of relatedness, self-efficacy	Satisfaction, persistence, academic achievement

Learning Theory	Central Thesis	Key Learner Characteristics	Focal Outcomes
Affect-related theories			
Control value (Pekrun & Perry, 2014)	Learners' appraisals of control and values arouse achievement emotions during learning, which influence engagement and outcome emotions	Emergent experiences of enjoyment, frustration, boredom during learning	Outcome emotions (joy, hope, pride, anxiety, shame, anger) related to success/failure

2.4.1 Cognitive Load Theory

Developed by John Sweller (1988), Cognitive Load Theory (CLT) provides a crucial framework for understanding the cognitive processes that underlie learning, especially in technology-enhanced environments (Sweller, 2020). CLT posits that optimal learning occurs when the cognitive load associated with the learning task does not exceed the learner's working memory capacity (Chen et al., 2023).

New information requires processing in working memory prior to long-term retention. Although long-term memory possesses vast storage capabilities, working memory is constrained in processing capacity. Consequently, tasks that surpass a learner's cognitive processing limits may negatively affect their learning (Chen et al., 2023; Sweller, 1988).

This dissertation adopts Kalyuga's (2011) perspective on CLT, which categorizes cognitive load into intrinsic (ICL) and extraneous (ECL). ICL stems from task complexity and expertise level (e.g., previous knowledge) and is essential for learning gains, while ECL, caused by inefficient instructional design, detracts from learning capacity. Increased ECL negatively affects learning and should be as low as possible (Sweller, 2020). This is particularly significant in digital learning settings, where learners are exposed to various information resources.

Applying CLT in digital learning settings necessitates strategies for managing various cognitive loads. For example, appropriately timed and well-designed scaffolds are crucial to preventing an increase in ECL, which could otherwise result in learner frustration and hinder motivation and the learning process (Chen et al., 2023; Thillmann et al., 2009; Hawlitschek & Joeckel, 2017). Moreover, the impact of ICL is particularly significant for novices and weaker learners, as they are more susceptible to cognitive overload in complex digital environments due to their limited prior knowledge. This underscores the need for instructional support that aligns with learners' knowledge levels to facilitate effective learning experiences (Wang & Lajoie, 2023).

2.4.2 Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning (CTML), introduced by Richard Mayer (2001), extends CLT principles to digital learning environments. CTML provides a comprehensive framework for grasping how people learn from multimedia content, highlighting the importance of designing educational materials that fit the human cognitive structure (Mayer, 2021; Mayer, 2024)

These foundational concepts of CTML (Mayer, 2001; Mayer, 2021) are built upon three key assumptions. First is the dual-channel assumption, which suggests separate processing of visual and auditory data. Second is the limited capacity assumption, which indicates restrictions on information processing within each channel. Third is the active processing assumption, which asserts that meaningful learning necessitates cognitive engagement.

Building on these assumptions, CTML establishes that working memory has a limited capacity, impacted by three cognitive demands: extraneous, essential, and generative processing (see Mayer, 2024). Extraneous processing refers to cognitive activities that do not aid the instructional objective, influenced by poor instructional design that introduces irrelevant information. Essential processing requires cognitive effort to hold information in working memory, depending on how complex the material is for the learner. Generative processing pertains to the cognitive efforts directed at comprehending incoming information, influenced by the learner's motivation to engage. Each processing type competes for the limited cognitive capacity, whereby extraneous processing detracts from essential and generative processing, and essential processing similarly limits generative processing capacity. The three cognitive demands lead to three objectives: reduce extraneous processing, regulate essential processing, and encourage generative processing (Mayer, 2024; Parong & Mayer, 2018).

To address these cognitive demands effectively, CTML research has identified 15 principles of multimedia instructional design, with effect sizes ranging from .10 to 1.35 (see Mayer, 2024 for an overview).

In the following, the focus is on the key principles (see Mayer, 2021; Makransky et al., 2021) that guided the research on the three papers, emphasizing the reduction of extraneous processing and management of generative processing:

1. Multimedia Principle: Learning is enhanced when words and images are presented together instead of words alone. This principle underscores the importance of leveraging multiple modalities in digital learning environments.
2. Spatial Contiguity Principle: Learning is improved when related words and images are displayed close to each other instead of being spaced far apart. In digital interfaces, this principle guides the spatial arrangement of text and visual elements.
3. Temporal Contiguity Principle: Learning is enhanced when related words and images are shown at the same time instead of one after the other. This principle informs the timing of multimedia presentations in digital learning materials.
4. Coherence Principle: Learning is improved when extraneous material is excluded. This principle is particularly relevant in digital environments where the temptation to include additional features or information can be high.
5. Redundancy principle: The inclusion of text cues alongside graphics and narration does not enhance learning. Individuals exhibit improved learning outcomes with graphics and narration alone in high-tempo situations.
6. Signaling Principle: Highlighting essential material enhances learning. In digital contexts, this can involve using visual cues, animations, or interactive elements to draw attention to key information.
7. Personalization principle: Individuals exhibit enhanced learning outcomes when verbal content is presented in a conversational tone instead of a formal tone.
8. Immersion principle: Immersive media, like the ones discussed in Paper 3, alone do not enhance learning. However, using effective instructional strategies within IVR or integrating these experiences into lessons can improve learning outcomes (see Makransky & Mayer, 2022)

While the first seven principles were mainly applied to designing and embedding digital learning prompts in paper 1 and 2, the immersion principle relates primarily to the virtual reality environment in paper 3.

2.4.3 Theories of Interest and Immersion

The transition from CTML to immersive learning environments represents a natural evolution in multimedia learning theory. The immersion principle integrates theoretical frameworks pertaining to interest (Renninger & Hidi, 2016), motivation (Ryan & Deci, 2000), and multimedia learning (Mayer, 2021) to elucidate the mechanisms through which augmented presence can enhance learning outcomes when effectively applied in a virtual reality design (Makransky & Mayer, 2022).

The primary benefit of immersive virtual reality (IVR) is its capacity to engage users in realistic settings, facilitating direct experiences of prospective work scenarios that are impractical to replicate in real life due to safety or cost concerns (Jensen & Konradsen, 2018). IVR utilizes head-mounted displays (HMD) and specialized controllers, distinguishing it from desktop VR using computer screens. IVR enhances sensory stimuli to create a more immersive experience, thereby increasing the user's feeling of presence within the virtual environment (Cummings & Bailenson, 2016; Makransky & Petersen, 2021). It is essential to differentiate between immersion, which pertains to the technical affordances of instructional technology, and presence, linked to the learner's personal perception. Specifically, immersion serves as an objective indicator of a system's capacity to create an engaging virtual environment while reducing awareness of real surroundings, while presence denotes the cognitive experience of 'being there' within the virtual simulation (Cummings & Bailenson, 2016; Makransky & Mayer, 2022; Slater, 1999; Slater & Wilbur, 1997).

As posited by Hidi and Renninger (2006), interest is divided into situational and individual interests. Situational interest reflects transient attention and emotion influenced by external stimuli (e.g., engaging visuals). Conversely, individual interest refers to a consistent preference for particular types of content (e.g., preferring videos to books). The four-phase model of interest development (Hidi & Renninger, 2016; Lin et al., 2024) posits that situational interest is initiated by stimuli capturing attention (phase 1). Subsequently, sustained affective responses and perceived value maintain this interest (phase 2). In phase 3, the individual actively seeks re-engagement opportunities (emerging individual interest). Ultimately, this interest solidifies into a consistent characteristic, referred to as individual interest (phase 4). In alignment with interest theory (Hidi & Renninger, 2016), it is posited that learners exert greater effort when they possess an intrinsic interest in the subject matter or when the educational context provokes situational interest. IVR environments can activate both interest forms, potentially augmenting the learner’s emotional and motivational states and fostering positive emotions, thereby facilitating enhanced learning outcomes (Schiefele, 2009; Parong & Mayer, 2008).

2.5 Personalized Learning and Prompting

Educators typically tailor their instructional strategies by offering additional assistance to students experiencing difficulties while simultaneously providing greater challenges to those excelling. Numerous educators perceive the fundamental aim of education as facilitating a personalized learning experience that aligns with each student's unique interests, prior knowledge, and skills at any moment (Holmes et al., 2018; Kerr, 2016).

To support this personalized approach, Wood, Bruner, and Ross first introduced and defined scaffolding in 1976. According to Wood et al. (1976), scaffolding is a systematic form of assistance that facilitates a learner's ability to achieve a specific objective or execute a task that would otherwise exceed their capabilities without external support. It is selectively administered at critical junctures during task execution and can be withdrawn once the learner achieves the task (Wood et al., 1976). A fundamental principle of scaffolding asserts that learners are expected to execute the task autonomously following the withdrawal of assistance (Lepper et al., 1997; Reiser, 2004). Given the diverse capabilities and expertise of learners, it is essential to offer tailored levels of assistance that can be adjusted as they progress (Umutlu & Gursoy, 2022; Clark & Hannafin, 2012; Lajoie, 2005).

Naturally, such individualized learning methodologies have captured the attention of instructional designers and educators long before the advent of digital learning technologies (Plass & Pawar, 2020). Nonetheless the automation features available in digital learning contexts have prompted a considerable shift from standardized approaches to personalized learning (Aleven et al., 2016). Specifically, innovations in digital education have allowed for the effective integration of scaffolding through what is referred to as 'prompts' within digital learning environments (Azevedo et al., 2005; Puntambekar & Hubscher, 2005).

Prompts are assistance or hints in the form of questions, suggestions and feedback presented during the learning process and promote the application of relevant processing strategies (Thomann et al., 2024; Wirth, 2009). As prompts require additional mental resources, they should not contain new information but instead aid in recalling and executing actions (Renkl & Scheiter, 2017; Bannert, 2009). They may appear in different formats, including instructional text, multimedia content, and interactive activities, and can be designed to achieve different learning goals. The level of personalization and explicitness of help can also vary among different prompts.

2.5.1 Theoretical Distinction of Prompt Features

The study of digital learning prompts is grounded in several theoretical frameworks that inform the understanding of learning processes, cognitive load (CLT; Sweller, 1988), and instructional design in digital environments (CTML; Mayer, 2021). This theoretical foundation provides the basis for investigating the effectiveness of prompts and their features in enhancing learning achievement.

Central to this research is a theoretical model that conceptualizes the interaction between learners and digital learning environments. This model, adapted from Clariana and Hooper (2012) and Mislevy and Riconscente (2005), comprises four interconnected components: the Expert Model, the Student Model, the Evidence Model, and the Instructional Model (Figure 3).

The Expert Model, derived from domain analysis, represents the ideal solution or knowledge state for a given task. It serves as the benchmark against which learner performance is evaluated.

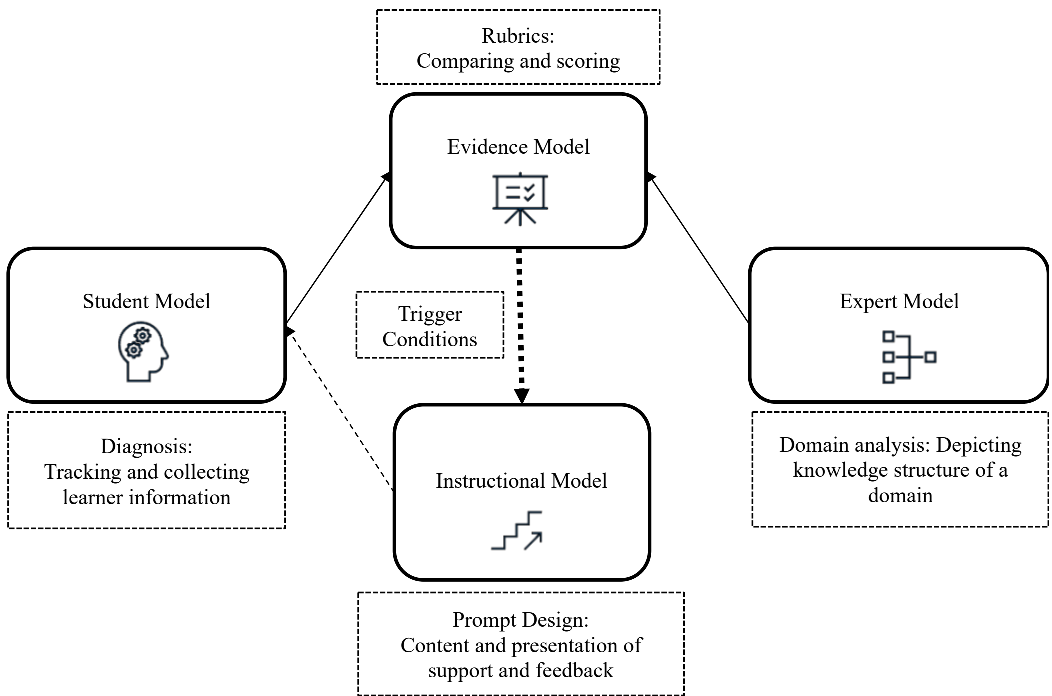
In contrast to the Expert Model, the Student Model captures the learner's problem-solving process and behaviors through continuous tracking within the digital environment. This model evolves as the learner interacts with the system, providing a dynamic representation of their current knowledge and learning progress.

The Evidence Model acts as the bridge between the Expert and Student Models. It employs scoring rubrics to detect discrepancies between the learner's current and desired knowledge states. This comparison process is crucial for identifying areas where the learner may need additional support or guidance.

Finally, the Instructional Model leverages the information from the Evidence Model to offer appropriate didactic support through prompts. These prompts are designed to address gaps in the learner's knowledge or skills, providing tailored assistance to guide them toward their desired learning outcomes.

This theoretical model underscores the adaptive and personalized nature of effective digital learning environments. It highlights the potential for prompts to serve as a key mechanism for providing timely and relevant support tailored to each learner's needs and progress. Prompts vary in design attributes, such as content, media, and triggering conditions within a learning context. Digital learning environments provide adaptability in prompting based on learner-specific progress and requirements, influenced by the environment's complexity. Prompting occurs in digital learning environments such as traditional e-learning systems, intelligent tutoring systems (ITS), learning simulation systems, educational games, virtual reality (VR), and augmented reality (AR) experiences[1].

Figure 3: Core model for assessing the effectiveness of prompts (copied from Paper 1)



[1] Following the definitions provided by Niegemann and Weinberger (2020) for categorization: Traditional e-learning systems are described as comparatively simple computer-based learning platforms with a feedback function that are used to deliver educational content and resources. Intelligent Tutoring Systems (ITS) provide personalized instruction to learners and extends beyond the simple feedback mechanisms of traditional e-learning systems. Digital learning simulations are interactive computer-based systems that stimulate real-world environments, processes, or situations. Educational (learning) games also typically function in a simulation-based way and embed educational activities within highly engaging, game-like interactions. VR and AR experiences are both types of computer-based simulations, but they differ in how they present digital information to the learner. VR creates a fully immersive, computer-based world that replaces the learners' view of the real world, whereas AR overlaps digital information on top of the learners' view.

2.5.2 Prompt Feature Categories

Building on this foundational model, Paper 1 explores four key prompt feature categories that can influence their effectiveness: (1) cognitive type (What is prompted?), (2) grouping of learners (Who is prompted?), (3) trigger conditions for prompts (When does prompting occur?), and (4) formulation and presentation of prompts (How are prompts formulated and presented, and do they include audiovisual content?). Table 4 provides an overview of these categories and examples from the studies included in Paper 1.

The "What" dimension focuses on the cognitive level of prompts, which can facilitate cognitive, meta-cognitive, and non-cognitive learning activities. In accordance with the classifications delineated by Azevedo et al. (2005), cognitive prompts aid in problem-solving and information processing, while meta-cognitive prompts enhance self-monitoring and goal setting, fostering learner engagement and self-awareness. Non-cognitive prompts aim to boost learner motivation through praise and guidance but do not directly influence knowledge acquisition.

The "Who" of prompts pertains to the selection of learners receiving prompts and the conditions of prompt adaptivity relative to learner quantity. This dimension is classified into three subcategories: “all learners,” “group of learners,” and “individual learners.” The “all learners” category encompasses prompts uniformly crafted for the entire learner population. The “group of learners” category involves prompts that vary among groups based on learner characteristics. For instance, prompts may differ based on learners' gender self-identification or responses to learning tasks. Additionally, group classifications can utilize prior assessments and questionnaires for differentiation in prompts. The “individual learner” category consists of prompts tailored to the specific behavior or characteristics of an individual learner. These individualized adaptive prompts are specifically designed for a unique learner.

The "When" dimension addresses the timing of prompt delivery, categorized as either time- or action-based (see Wirth, 2009). Time-based prompts appear at predetermined intervals (Robertson et al., 2015) and should be strategically timed to minimize cognitive load (Thillmann et al., 2009). Action-based prompts respond to learner behaviors, including navigation patterns and performance metrics (Bernacki et al., 2021). Additionally, they may be triggered by learner-initiated actions, such as clicking a prompt button (self-selected).

In addition to Paper 2, recent investigations, including those by Liu et al. (2024), Munshi et al. (2023), and Van der Graaf et al. (2023), have developed and integrated prompts into digital learning platforms utilizing trace (log) data and performance metrics. Trace data consists of digital logs reflecting learners' interactions to infer learning processes (Hadwin et al., 2007). It encompasses navigational logs, keystrokes, mouse movements, response behavior (clicking), and eye gaze points (Bernacki, 2019; Azevedo et al., 2013). A key advantage of trace data lies in its unobtrusive measurement of student activities (Winne, 2010). While earlier prompt studies primarily relied on navigation logs (Bannert et al., 2015; Müller & Seufert, 2018; Pieger & Bannert, 2018), these logs alone proved inadequate for reliably assessing students' ongoing learning processes due to their limited granularity (Järvelä & Bannert, 2021). In response to this limitation, recent research has established analytics-based approaches that leverage various forms of trace data to gain a deeper understanding of the learning processes involved in prompt development (Liu et al., 2024; Li et al., 2023; Munshi et al., 2023; Van der Graaf et al., 2023). The personalized prompt design of *Luca* office (Paper 2), which will be detailed in the next chapter, exemplifies this evolution. Additionally, Artificial Intelligence (AI) has recently surfaced as a powerful tool in educational prompting, offering personalized and adaptive learning experiences. For instance, Lim et al. (2024) developed analytics-based personalized prompts for essay writing, facilitated by a rule-based AI system, on learners' processes using log data (e.g., navigation logs, mouse movements). This analytics-driven approach with rule-based AI systems furnishes personalized support tailored to individual learners' distinct educational requirements. Nonetheless, such analytics-based prompts remain relatively rare, necessitating further investigation to refine and evaluate them.

Under the "How" category, prompts are classified as either "generic" or "directed" (see Davis, 2003). Generic prompts offer broad applicability across learning contexts, while directed prompts provide context-specific guidance characteristics (Davis, 2003; Ifenthaler, 2012).

Regarding presentation format, prompts can be text-only or multimedia-enriched. Multimedia integration aligns with situated learning principles and anchored instruction approach (Young & Kulikowich, 1992; Cognition and Technology Group at Vanderbilt, 1997; Langone, 1998; Oliver & Herrington, 2000). In this context, Mayer's (2021) cognitive theory of multimedia learning offers critical perspectives for enhancing multimedia efficacy (see Section 2.4.2).

These prompt feature categories are not isolated constructs but interconnected aspects of instructional design in digital learning environments. Their effectiveness is likely moderated by various factors, including learner characteristics, learning domain, and the specific digital learning environment in which they are implemented. For a thorough overview of these prompt feature categories and their effectiveness on learning outcomes, please refer to pages 4-8 (Paper 1) and section 5.1 of the discussion.

Table 4: Overview of prompt features (copied from Paper 1)

	Categories	Subcategories	Definition	Example
What	Cognitive	none	Prompts that support and promote learners' cognitive processes.	One of your answers is incorrect. Evaluate the expressions below. Write each response as an integer or as a fraction. (ID 75)
	Meta-cognitive	none	Prompts that activate the monitoring and control of learners' meta-cognitive activities.	How can I best organize the structure of the learning contents? (ID 85)
	Non-cognitive	none	Prompts that seek to increase learner motivation by praising the student or providing guidance on regulating their motivation.	The e-module was designed for all kinds of learners with different skills and backgrounds. Please feel assured that with appropriate effort, all participants should be able to complete the e-module successfully. (ID 70)
Who	All Learners	none	Prompts that are available for all learners and have the same content.	Think about the personal relevance of the learning material. (ID 41)
	Group of Learners	Grouping based on learner's activities. Grouping based on previous test/questionnaire: • Prior knowledge • Learner personality	Prompts that are only available for a specific group of learners.	Dear participant, we would like to advise you to look at the interactive visualization a little bit longer. (ID 76)
	Individual Learner	• Self-created prompts	Prompts which are based on a individuals learner's behavior or self-created prompts.	I plan my next learning steps. I reflect on my learning strategy. I make sure to cover all learning goals (ID 99)
When	Time-based	• During the learning sequence • Previous to the learning sequence • After the learning sequence	Prompts that appear after a certain amount of time	Before you start to learn, you should first prepare yourself. (ID 95)
	Action-based	• Self-selected • Previous Activity (navigation decision, viewed content) • Incorrect/Correct Solution	Prompts that are based on the learners' activities	Hi, here I am again! You have been busy for a while, but you still have not taken any notes. (ID 174)
How	Generic	none	Prompts that are worded in general terms and can be used for a variety of digital learning content.	Which are the main points in your opinion? (ID 51)
	Directed	none	Prompts that are content-specific and contextualized.	Why do you calculate the total acceptable outcomes by multiplying? (ID 47)

2.6 Personalized Learning in Vocational Education

The upcoming chapters examine the computer-based office simulation *Luca* (Paper 2) and the immersive Virtual Reality environment *InGo* (Paper 3), particularly in relation to VET. For an overview of the importance and context of personalized learning in VET, refer to Section 1.1.

2.6.1 Computer-based Learning Simulations

Contemporary workplaces necessitate advanced competencies such as problem-solving (World Economic Forum, 2023; Rausch et al., 2021; Binkley et al., 2012). The term "problem-solving" denotes the method of identifying a solution to a problem with inadequate prior knowledge (Ludwig et al., 2024). This challenge is particularly relevant given the increasing diversity of information at the workplace (and in VET), which necessitates the classification and critical evaluation of sources to navigate current challenges (Schoor et al., 2020, 2021). Computer-based simulations are instrumental in cultivating these essential 21st-century skills (Funke et al., 2018; Rausch et al., 2017). Two primary characteristics define computer-based simulations. They facilitate real-world scenarios that foster genuine learning and create a credible learning environment (Braunstein et al., 2022). Additionally, they incorporate computer-based assessments, enhancing the analysis and support of problem-solving behaviors. During interactions within these environments, learners' activities are recorded as time-stamped trace data in individual log files (see Hadwin et al., 2007).

This log data enables just-in-time personalization of prompts to better suit learners. Adaptive environments, such as the web-based office simulation *Luca*, offer more complex and engaging experiences and promote self-regulated learning more than less immersive alternatives (Zimmerman & Moylan, 2009; Moos & Bonde, 2016). Integrating personalized prompts improves the learning experience by delivering customized support to individuals with diverse proficiency levels (see Schumacher & Ifenthaler, 2021; Ifenthaler, 2012). Furthermore, they serve as a safety net, fostering an interactive learning environment where individuals take an active role in working through the task scenarios while correcting their errors as needed (Mead et al., 2019). Building upon these principles of personalized learning support, the following chapter presents the personalized prompt design within *Luca* office developed in Paper 2 and its theoretical background.

2.6.2 Personalized Prompt Design based on the 4C/ID Model

The Luca (Learning by Using Competence Assessment) is a web-based office simulation designed as a browser platform that serves dual purposes as both a learning and assessment environment. It includes essential office applications such as an email client, file management system, PDF viewer, spreadsheet program, calculator, notepad, and a basic enterprise resource planning (ERP) system. *Luca* facilitates the development of intricate scenario-based tasks via a graphical interface. Its replicated office tool interfaces are designed to stimulate authentic problem-solving actions. For an extensive examination of *Luca's* structure and functionalities, refer to Paper 2, Ludwig et al. (2024) and Rausch et al. (2021).

The didactic design of *Luca* office is based on approaches such as self-regulated learning (Zimmerman & Schunk, 2011) and scenario-based learning (Errington, 2010, 2011). Self-regulated learning (SRL) involves active processes where learners manage their cognitive, emotional, and behavioral processes to meet individual learning objectives (Pintrich, 2000; Zimmerman & Schunk, 2011). Building upon constructivist ideas, the design additionally incorporates connectivism learning principles. Connectivism posits that learning occurs within dynamic networks comprising both human and non-human elements. These networks may include real networks (e.g., colleagues and supervisors), as well as technological and informational resources (Rausch et al., 2021). Students frequently encounter difficulties in applying SLR (Miller & Bernacki, 2019). Research has demonstrated that learners require structured support to engage effectively in self-regulatory behaviors in computer-based learning environments (Moos & Bonde, 2016; Sonnenberg & Bannert, 2016). In this context, prompting has emerged as a crucial instructional strategy that enhances both learner motivation and learning regulation. Extensive research has investigated prompting design and implementation, particularly to facilitate SRL and the specific activities to be prompted (Schumacher & Ifenthaler, 2021; Ifenthaler, 2012; Bannert, 2009; Wirth, 2009). Although SRL provides valuable context for understanding the learning process, it is not the central focus of this dissertation. Instead, this research specifically examines learning achievement within prompted learning environments, investigating how different prompting approaches influence learning outcomes.

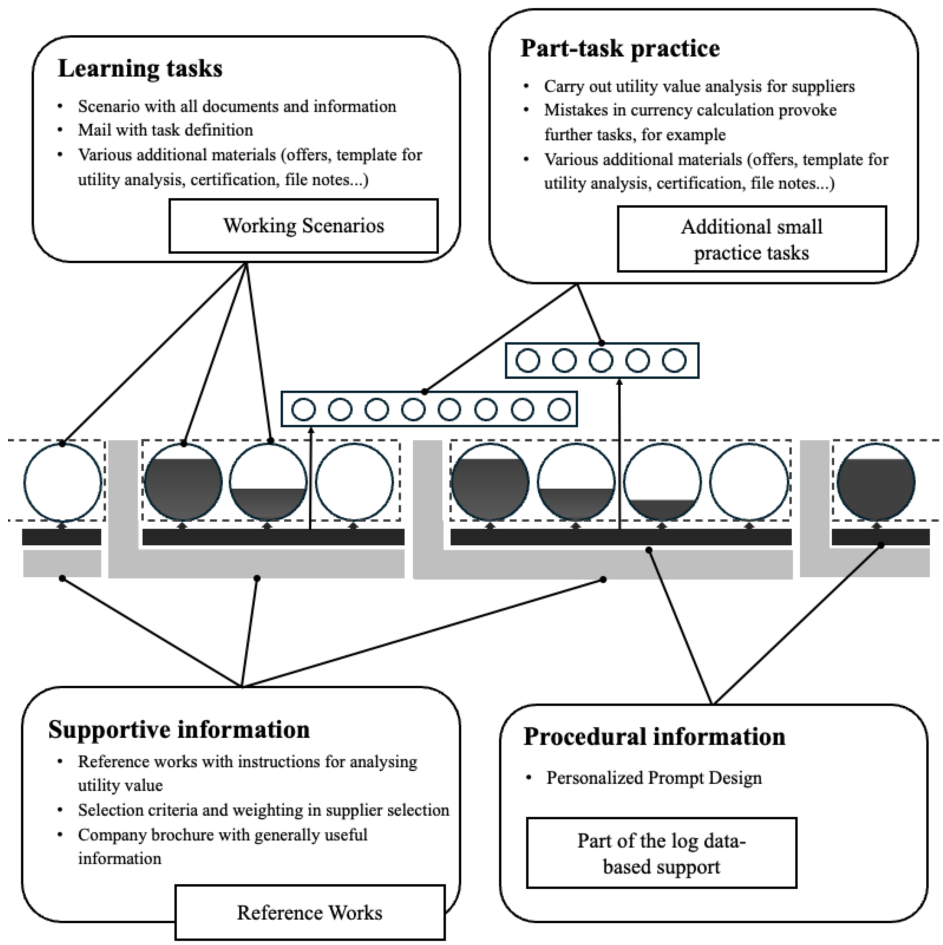
The design also incorporates scenario-based learning principles (Errington, 2010, 2011), which align fundamentally with the objectives and methodological requirements outlined in vocational curricula (Rausch et al., 2021). This alignment necessitates translating authentic occupational tasks into realistic work scenarios, which can be structured at varying levels of complexity to match learner competence in problem-solving (Jonassen, 2000). In this context, problems are divided into well-defined and ill-defined categories. Well-defined problems illustrate explicit processes and a restricted set of resolution operators. In contrast, ill-defined problems present numerous potential solutions and are often deemed more complex (Funke, 2003) During the design-based research process of Paper 2 (see Section 3.2.2), participants first underwent a 15-minute onboarding session to become acquainted with a fictional company, the simulation's features, and the operation of prompts. They then engaged in a 55-minute complex work scenario involving supplier selection for a bicycle manufacturer. The task required participants to analyze four supplier offers based on price, quality, and delivery time using provided documents and a spreadsheet. *Luca* office presented a mix of relevant and irrelevant information without task structuring. Consequently, the task exemplified an ill-defined problem characterized by increased complexity and ambiguity (Funke, 2003). Thus, it demanded improved self-regulation and problem-solving skills (Ludwig et al., 2024). The empirical testing of the personalized prompt design for the Luca office simulation is ongoing and will be detailed in a separate paper distinct from this dissertation.

To systematically implement these learning principles within *Luca* office, particular attention was paid to the four-component instructional design (4C/ID) model (van Merriënboer, 2013, 2019; van Merriënboer & Kirschner, 2018). The model consists of four interrelated components:

(1) Learning Tasks: These form the core of the model, presenting learners with authentic, whole-task experiences that integrate knowledge, skills, and attitudes.
(2) Supportive Information: This component assists learners in performing non-routine aspects of tasks by providing domain models and structured approaches to problem-solving.
(3) Part-task Practice: This provides additional rehearsal for selected routine elements requiring high levels of automaticity.
(4) Procedural Information: This offers just-in-time guidance for task aspects, facilitating rule automation.

These theoretical components of the 4C/ID model (see Figure 4) are systematically implemented in *Luca* office as follows (Paper 2; Rausch et al., 2021): (1) Learning Tasks: Complex authentic work scenarios “supplier selection”, (2) Supportive Information: Reference works or other learning materials containing general subject knowledge, domain-specific models, and subject heuristics, (3) Part-Task Practice: Short practice tasks in individual articles in the reference works and learning materials or short tests between scenarios, (4) Procedural Information: Automated assistance or ad-hoc help 'just in time' (digital learning prompts).

Figure 4: Schematic overview of the 4C/ID Model regarding Luca office (adapted from van Merriënboer et al., 2002)



The just-in-time prompt design based on log data was incorporated in line with the principles of the 4C/ID approach (van Merriënboer & Kirschner, 2018). Ultimately, the provision of complexity-reducing aids is intended to counteract excessive demands (scaffolding) and, at the same time, achieve increasing independence in task processing (fading), as envisaged in the cognitive apprenticeship approach (Collins et al., 1989). The prompt design was informed by the theoretical framework outlined in the prompt feature chapter based on Paper 1. Overall, 18 cognitive, non-cognitive, and metacognitive learning prompts were embedded in the final work scenario (see Table 5 for an overview).

A rule-based system analyzed learners' entries (log data) in the spreadsheet template to generate personalized cognitive prompts that aligned with their documented learning processes. Rule-based systems operate on a framework of conditional statements established by domain experts, typically structured as 'IF [condition] THEN [action]' (Lim et al., 2024). To implement this system, domain experts first conducted a comprehensive analysis to identify all plausible input values and potential errors. Then, trigger conditions for prompts were established for each solution-relevant cell in the spreadsheet. For instance, when learners input an incorrect exchange rate (IF) during reference price calculations, the system automatically generates a personalized prompt (THEN) identifying the potential error source (Table 5, No. 1). The system also incorporated non-cognitive prompts designed to enhance learning motivation. For example, when the task suggests considering additional selection criteria, learners who independently introduce supplementary factors (such as environmental compatibility or ethical considerations) receive reinforcing feedback (Table 5, No. 2). Additionally, metacognitive prompts (Table 5, No. 3) were implemented to direct learners' attention to relevant information that remained unexamined according to the log data.

Following the recommendations outlined in Paper 1, the amount of unnecessary text in prompts was reduced to manage the working memory demands and cognitive load faced by learners when they received the prompts. Each prompt suggested specific actions with each serving a distinct purpose. Special attention was given to the social embedding and personalization of prompts including personal addressing using usernames. Braunstein and colleagues (2022) note that this aligns with the "social interaction level within the taxonomy of social embedding" (p. 5). In *Luca*, social interaction is facilitated during task processing by allowing learners to engage with actual individuals or fictional characters within the simulation, such as colleague Aylin or their supervisor. Furthermore, the study utilized both generic and directed prompts depending on the learning objectives of the target learners, as well as their prior knowledge levels. To ensure effective engagement with the prompts, the target group was familiarized with their use and underwent additional onboarding training focused on utilizing *Luca* office and the learning prompts.

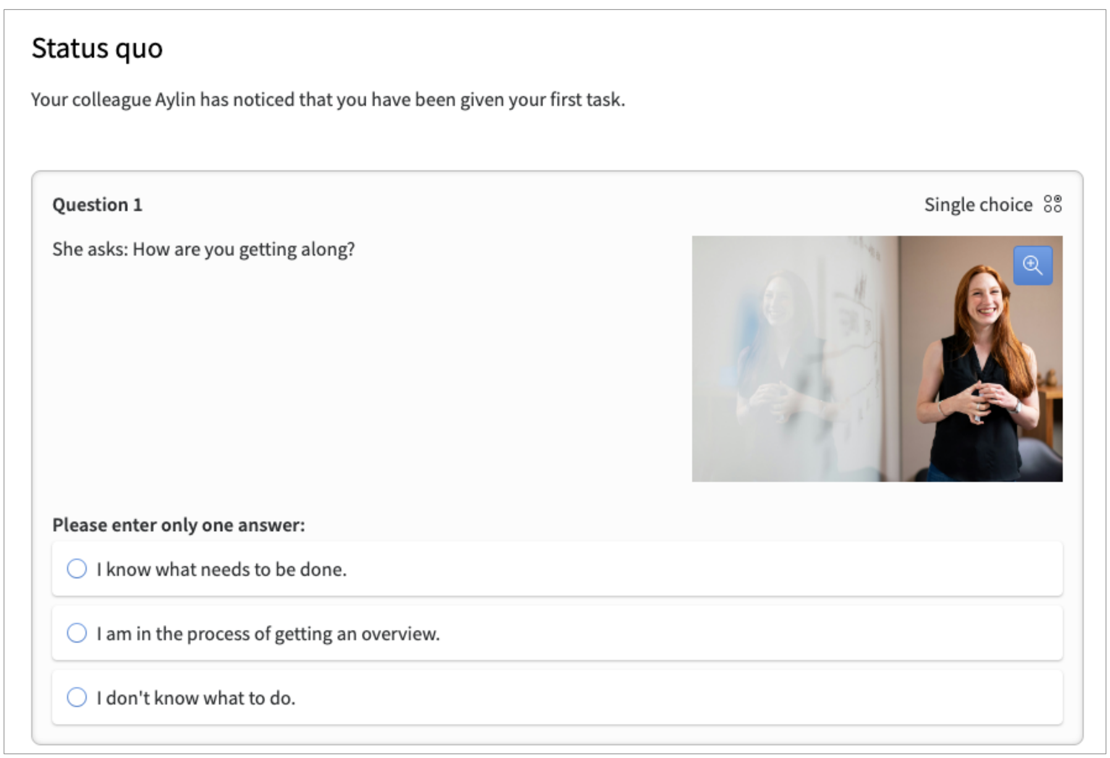
Due to the technical requirements of *Luca* office, the time for the appearance of each prompt also had to be specified. This can be described as a combination of action and time-based prompts. Numerous studies indicate that while scaffolds can enhance learning, poorly designed ones frequently disrupt the learning process. This disruption leads to learners' frustration and a negative user experience (Munshi et al., 2023; Van der Graaf et al., 2023; Álvarez et al., 2022; Shih et al., 2010). Consequently, rather than an immediate prompt window appearing in the *Luca* office at the designated time, an unread envelope icon was displayed in the interface's taskbar to capture learners' attention at the appropriate moment (e.g., minute 5). This icon served as a reminder of an unread email notification, effectively initiating the prompt while reducing disruptions. Moreover, all previous prompts could be revisited during the learning task.

Table 5: Task and prompt design overview (translated from German; Paper 2)

No.	Solution Step in the Work Scenario	Prompt Type	Trigger Condition	Prompt Presentation	Prompt Content
1	Learners calculate the subscription price and enter it in the cell provided in the spreadsheet template.	cognitive	wrong value in cell L14 *or* missing value in cell L14 (after 37 minutes)	E-Mail Prompt	Hello (form of address), did you pay attention to the current exchange rate when converting the currency? You will find a table on exchange rates in the reference book.
2	The learners independently add further selection criteria to the spreadsheet template for the utility analysis.	non-cognitive	Text input in cells B17 to B19 (after 40 minutes)	E-Mail Prompt	Hello (form of address), You have recognized well that it could be useful to include other selection criteria for the utility analysis. Keep up the good work!
3	Learners identify necessary information to determine order values.	meta-cognitive	Not opening a relevant file in the ERP system. (after 20 minutes)	E-Mail Prompt	Hello (form of address), have you had a chance to look at the supplier Jinshu Gongsi's file note and consider it in your selection?
4	Learners get an overview of the task requirements and the documents needed to resolve them.	meta-cognitive	Depending on the answer choice of the event query (answer option 3)	E-Mail Prompt	Hello (form of address), before I start the supplier selection process, I always take a moment to make notes. In particular, the preparation of a utility analysis requires various steps. You will find information on this in the supplier selection reference book.

Building upon this personalized prompting system, *Luca* office additionally incorporates what is called 'Events' or embedded experience sampling (EES) (refer to EES, Rausch et al., 2019). EES are short interruptions in the participants' problem-solving tasks, prompting them to respond to brief closed questions regarding their current experiences. These events are embedded in the storyline of the scenario, simulating typical office settings while reflecting an interdisciplinary approach and advocating for theory-driven design techniques (Rausch et al., 2024). Moreover, these Events can trigger personalized prompts based on learner responses, which may include assessments, test questions, or preferences presented as brief queries. For example, when a colleague asks about the task's progress (Figure 5), the system's response varies according to the learner's chosen answer. Selecting the first option elicits a non-cognitive prompt featuring praise. Choosing the second option, which reflects a need for an overview, generates a metacognitive prompt aimed at aiding task planning. This prompt is intentionally formulated in an unspecific and concise manner to avoid restricting learners' choice of problem-solving strategies. If learners select the third option, expressing feelings of overwhelm, the system issues a comprehensive metacognitive prompt that provides specific recommendations for further action (Table 5, No. 4).

Figure 5: Event as a query for generating personalized prompts in Luca office (translated from German)



2.6.3 Immersive Virtual Reality Simulations

> *“Students are not only active but also actors: they co-construct the virtual space.”* (Dillenbourg et al., 2002, p. 1)

This co-construction principle fundamentally shapes how learners interact with and benefit from Immersive Virtual Reality (IVR) simulations. These environments transcend traditional educational boundaries by adapting to individual preferences and unique learning behaviors, enabling learners to actively shape their educational experiences through dynamic interaction with the virtual space (Dillenbourg et al., 2002).

Moreover, the application of IVR technology, while predominantly concentrated in technical and medical education (Radianti et al., 2020), shows remarkable potential for VET. The distinctive characteristics of IVR, particularly its immersive capabilities, align exceptionally well with VET's core objectives. These include the development of action-oriented skills, the acquisition of domain-specific knowledge, and their practical application in novel work scenarios (Buchner & Mulders, 2020; Conrad et al., 2022; Zinn, 2019). Nevertheless, recent studies (Liu et al., 2024; Conrad et al., 2022; Hellriegel & Cubela, 2018) emphasize a notable gap in research regarding the systematic evaluation of IVR's effectiveness in cultivating domain-specific skills.

To address this gap, Paper 3 examines the effectiveness of IVR in warehouse logistics through the *InGo* simulation developed by the Fraunhofer Institute for Material Flow and Logistics (see Schlüter & Kretschmer, 2020 for an overview). According to Won et al.'s (2023) framework, InGo represents a medium to high-quality IVR environment, incorporating sophisticated sensory features, intuitive control mechanisms with real-time feedback, cohesive narrative structures that enhance learning objectives, and simulated social interactions (Braunstein et al., 2022). Table 6 presents a comprehensive analysis of *InGo’s* features.

Importantly, the connection between immersion and learning outcomes requires careful consideration through the perspective of Makransky and Petersen's (2021) cognitive-affective model for immersive learning (CAMIL). While enhanced immersion can significantly boost learner engagement and motivation, it simultaneously presents challenges related to cognitive load management and self-regulation. This aligns with Mayer's (2021) immersion principle, suggesting that immersion by itself does not ensure better learning outcomes. The CAMIL framework helps us understand how various cognitive and emotional factors influence the link between immersive experiences and learning effectiveness, providing crucial insights, for designing optimal learning environments (Obourdin et al., 2024).

Table 6: Categorization of the integration of design features in IVR environments (copied from Paper 3)

Design features	Description	Subcategories	Integration levels (Low – Medium – High)
Sensory	Representational fidelity. The presented virtual environment is representationally sound, so learners can feel that the virtual objects and places are authentic or real.	Visual	Medium (Medium resolution graphics. Oculus Quest (2 and 3 Gen.)
		Audio	High (Effective immersive sound effects for location, direction, and conversation with virtual agents, such as the truck driver and supervisor.)
		Haptic	High (Realistic force feedback and additional sensory responses, e.g., picking up phone and scanning packages.)
Actional	Intuitive interface design. The actions in a virtual environment feel natural and intuitive for learners to feel they are making real changes in the environment.	Interactivity	Medium to high (High interactivity and intuitive user control to feel the interactions are natural and rule-bound. However, walking is done by teleporting by clicking on the joystick.
		Embodied movement	High (Whole body movement is essential to carry out the learning tasks, e.g., walking, picking up things, scanning barcodes, etc.)
Narrative	Engaging content and task. The content and tasks are relevant and meaningful for learners to feel emotionally and intellectually engaged.	Roles	Medium (Clear role with an avatar to make consequential decisions, e.g., act as a logistic employee in a warehouse.)
		Contexts and storylines	Medium (Skillfully crafted storyline in a relevant context that appeals to learners' experiences, e.g., overarching decision-driven narrative guided by the truck driver Ingo.)
		Challenges and achievement	Low (Completion of a task as a one-time experience.)
Social	Constructive support. The learners and learning are supported through social interactions.	Social interactions	Medium No mediated social interactions from peers and teachers. However, learners receive socially framed instructions in each step via a tutoring system and tailored feedback on their learning performance at the simulation's end, which corresponds to the medium level of Social Reaction (level 3) in the taxonomy of authentic social embedding (Braunstein et al., 2022).

Recent literature reviews also highlight that IVR's personalization possibilities, while promising for designing effective instructional methods, remain significantly underutilized (Marougkas et al., 2023; Zahabi & Razak, 2020). Incorporating personalization features (besides co-constructing the virtual space by navigating through the environment) within IVR environments offers valuable capabilities. It can automatically regulate complexity, decrease cognitive load, and maintain learner focus and achievement through timely delivered prompts (see Obourdin et al., 2024). However, multiple reviews have repeatedly noted a lack of empirical research integrating personalized learning elements like prompts or tailored content in IVR environments (see Marougkas et al., 2023; Fraulini et al., 2024; Zahabi & Razak, 2020).

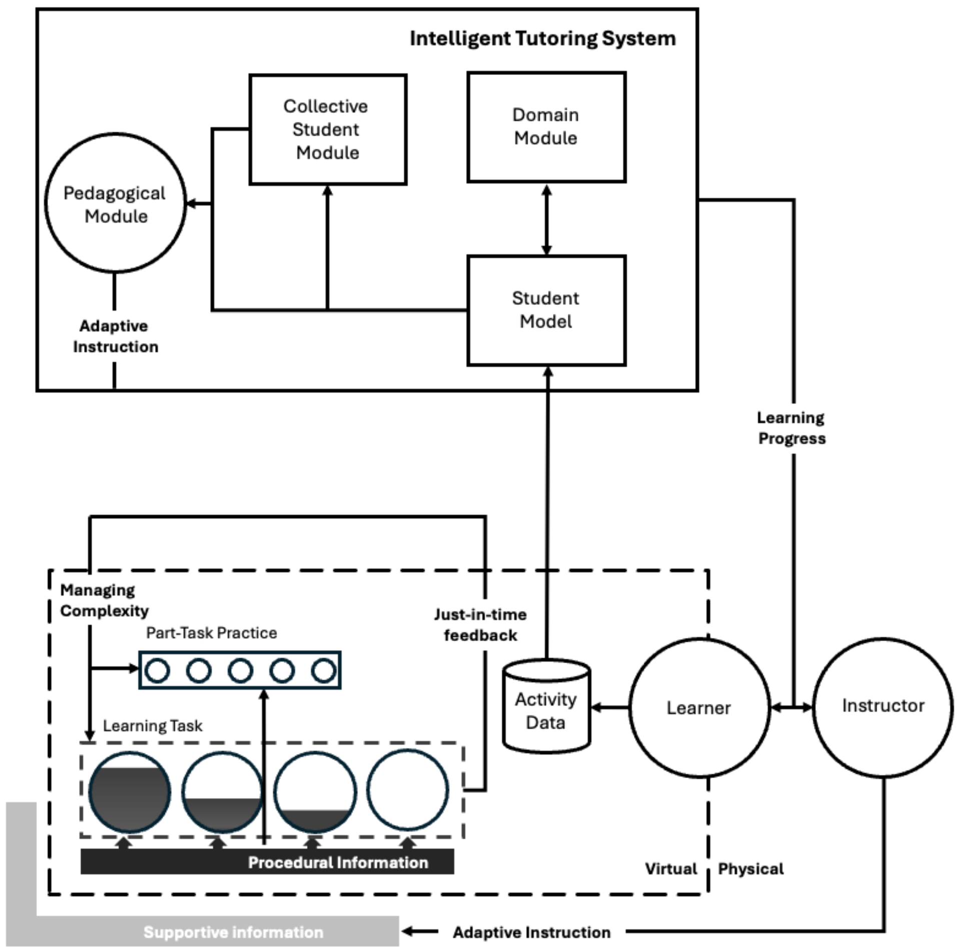
The *InGo* environment exemplifies both the potential and limitations of current IVR implementations in educational contexts. While it offers a personalized feedback dashboard displaying individual performance data post-task completion, it lacks real-time personalized learning prompts during task execution. For instance, prompts were only given when the learner overlooked important steps necessary for completing a task (e.g., forgetting to bring the mobile phone to scan the incoming goods). This raises important questions about how to effectively implement personalized features to ease the severity of IVR learning experiences and meet individual needs. Despite the extensive literature on IVR intervention design (see Mulders et al., 2020), there is a notable deficiency in guidance on the integration of personalized instruction within IVR to maximize these capabilities (Obourdin et al., 2024).

2.6.4 Personalized Prompting based on the FAIRI Framework in IVR

To address this gap in current research, this dissertation adapted Obourdin et al.'s (2024) FAIRI (Four-Component based, Adaptive, Immersive, Realistic & Intelligent) framework, and modified the intelligent tutoring system (ITS) with the intelligent prompt model components from Paper 1 (see also Section 2.5.1).

Figure 6 illustrates the instructional design principles, serving as a foundation for a personalized learning environment that employs IVR-enhanced instruction and incorporates the FAIRI framework. FAIRI, anchored in the "(F) 4C/ID model, maximizes IVR's (A) adaptive and (I) immersive affordances by emphasizing (R) realistic learning tasks within the virtual environment for efficacy whilst supporting and manipulating this learning environment with an (I) intelligent tutoring system" (Obourdin et al., 2024, p. 6). The name hints at a fairy's talent for tricking humans by producing lifelike illusions that satisfy their wishes. Similarly, the FAIRI framework offers a virtual environment tailored to the needs of each learner (Obourdin et al., 2024).

Figure 6: FAIRI (Four-Component, Adaptive, Immersive, Realistic, Intelligent) instructional design model (based on Obourdin et al., 2024)



The FAIRI model is predicated on two core design tenets. First, it restricts IVR applications to learning tasks that gain from enhanced realism, and second utilizing IVR's personalization capabilities through an intelligent tutoring system. ITS utilize artificial intelligence to provide personalized instruction, seeking to replicate the efficacy of human tutors (Sedlmeier, 2001). The model's dual-environment structure addresses cognitive load management and learner distraction by systematically integrating 4C/ID components (Section 2.6.2), with ITS components (see the prompt model in Section 2.5.1). Notably, Mulders (2022) emphasizes the importance of using the 4C/ID model to develop IVR learning environments that improve competence acquisition in VET.

Additionally, the FAIRI model is visualized through two interconnected environments - virtual and physical - that work together to optimize learning experiences.

The figure demonstrates how learning tasks are situated within the virtual environment (4C/ID model, bottom left) and presented as authentic, whole-task experiences with varying complexity levels. These tasks are supported by procedural information, which provides just-in-time support (prompting) and part-task practice sessions for mastering routine aspects of complex skills. Supportive information (e.g., learning materials containing general subject knowledge or heuristics) is relocated to the physical realm. The transmission of supportive information is comparably effective, if not superior, in the physical realm and does not confer an additional learning advantage via immersive virtual reality.

The intelligent tutoring system, centrally positioned in the figure, plays a crucial role and can be modified by the prompt-model components (expert model, student model, evidence model, and instructional model) for the emphasis of this dissertation (see Section 2.5.1). The student model monitors learner performance (Activity Data), the expert (domain) model provides benchmarks, the collective student model aggregates individual student models and summarizes prior learners' data, the evidence model identifies gaps (shown in the prompt model only), and the instructional model (pedagogical module) triggers personalized prompts and adjusts the virtual environment's complexity to match individual learner needs. Figure 6 illustrates the integration of the pedagogical module (instructional model) with the learning task and procedural information, emphasizing the prompt model's role in customizing the immersive learning environment to address the specific needs of the learner.

The system creates a two-way flow of information through the ITS (prompt model components). This data helps instructors tailor their teaching methods to individual students while enabling students to review and reflect on their IVR performance. By connecting both students and instructors to the ITS, as shown in Figure 6, the framework bridges the virtual and physical learning spaces. This integration aims to enhance students' virtual learning experiences by providing customized support and guidance in their physical learning environment. Augmenting the IVR experience encompasses equipping learners for upcoming learning tasks and pre-training on prompting to mitigate extrinsic cognitive load while promoting “reflective practices by using the physical world to decouple from IVR” (Obourdin et al., 2024, p. 7). The instructor customizes their pedagogical approach for each individual learner by utilizing insights derived from the ITS components (prompt model). Concurrently, the student employs these insights to introspect their recent performances within the virtual learning environment.

The integration of the prompt model's components with FAIRI, though only theoretically elaborated in this dissertation's framework section and not implemented in Paper 3, directly addresses the previously identified gap in personalized IVR research. The framework faces three limitations. First, the impact of prompts on cognitive load remains insufficiently understood. Second, procedural support through prompts requires careful integration within virtual environments to maintain presence while providing adequate guidance (Obourdin et al., 2024). Third, individual differences in spatial ability, working memory capacity, and predisposition toward IVR absorption significantly affect how learners interact with these supportive elements (Makransky & Petersen, 2021). Despite these constraints, it provides a structured foundation for future research. While empirical validation lies beyond the scope of this theoretical integration, the framework serves as a foundational basis for personalized prompts in IVR-supported learning environments.

3 Methodological Approach

This dissertation employs a multi-method approach to investigate personalized learning in vocational education through three distinct studies. Each study utilizes different methodological strategies to address specific aspects of digital tools in educational contexts.

3.1 Measurement Operationalization

3.1.1 Objective and Subjective Learning Achievement

In the meta-analysis of digital learning prompts in Paper 1, the main outcome variable, learning achievement, was defined using post-test scores from the studies examined. These scores were gathered from evaluations carried out following the learning interventions that included prompts. When studies reported various learning outcomes, a clear differentiation was made between declarative knowledge and procedural knowledge. Procedural knowledge pertains to methods of task execution, while declarative knowledge involves factual information and conceptual understanding (Anderson & Krathwohl, 2001; Greeno & Gelman, 1984).

The third paper operationalized several key constructs to investigate the effectiveness of IVR in VET. The measurement approach encompassed both objective and subjective measures, focusing on declarative knowledge acquisition. Objective declarative knowledge acquisition was measured through a pre- and post-test design. The pre-test included three items related to the warehouse logistics curriculum, comprising two open-ended questions and one assignment. The maximum score on the pre-test was 7.5 points, with partial credit available for each item. The post-test, designed as a parallel test to mitigate memorization effects (see Blumberg, 1981), included five items consisting of four multiple-choice questions (one point each) and one assignment item (partially correct answers accepted), resulting in a maximum total of 7.5 points. The tests, despite lacking psychometric reliability due to time limitations, effectively mirrored the content of the *InGo* simulation, functioning as a formative assessment of declarative knowledge acquisition (see Coltman et al., 2008). It is noteworthy that while the study primarily measured declarative knowledge, the IVR environment *InGo* was designed to teach a simplified goods acceptance process, which inherently involves procedural knowledge. Procedural knowledge, which refers to the understanding of action execution, was included in the learning experience yet not explicitly measured in the immediate post-test. Subjective knowledge acquisition was assessed using a single measure item adapted from Lee et al. (2010), capturing participants' perceptions of their learning progress.

3.1.2 Mood, Motivation, and Immersion

In the IVR study of Paper 3, the PANAS Scale (Mackinnon et al., 1999) was utilized to assess both positive and negative mood features. The positive mood scale ($\alpha = 0.78$) comprised items on inspiration, attentiveness, excitement, and relaxation, whereas the negative mood scale ($\alpha = 0.87$) evaluated emotions such as anxiety, annoyance, and nervousness. Intrinsic motivation was evaluated using a modified version of the "Short Intrinsic Motivation Scale" (Wilde et al., 2009). This scale comprised four dimensions: Interest and Enjoyment ($\alpha = 0.85$), Perceived Competence ($\alpha = 0.83$), Perceived Choice ($\alpha = 0.75$), and Pressure/Tension ($\alpha = 0.60$). The immersion scale ($\alpha = 0.89$) was based on Georgiou and Kyza (2017), assessing participants' level of engagement and absorption in the learning environment. For the IVR group, additional measures from the "Unified UX Questionnaire" (Tcha-Tokey et al., 2016) were employed, including positive and negative emotions, experience consequences (motion sickness), engagement, flow experience, presence, and overall judgment of the IVR experience. These scales demonstrated satisfactory reliability (α

= 0.718–0.908). Prior Experience (α = 0.82) and Demographics measures in the pre-questionnaire included age and gender, previous experiences with VR technology, and familiarity with incoming goods processes (Shou and Olney, 2021). All questionnaire items, except for the immersion scale, used a 5-point Likert scale ranging from "Strongly disagree" to "Strongly agree." The immersion scale employed a 7-point Likert scale.

3.2 Testing Methods

3.2.1 Paper 1 (Effectiveness of Digital Learning Prompts)

The first paper employed an abductive research approach, combining inductive theory development with deductive hypothesis testing. The research proceeded in two phases with distinct objectives:

Research Question 1 (Systematic Review): Which types of prompts are distinguished throughout the literature?

Research Question 2 (Meta-Analysis): What is the overall effect of prompts on learning achievement, and how is the effectiveness moderated by prompt features and study demographics?

To address these questions, the first phase involved conducting a systematic review (manuscript 1) to inductively develop a theoretical framework that categorizes prompt features across four dimensions: cognitive type ("What"), learner grouping ("Who"), timing ("When"), and formulation (“How”). The systematic review followed a rigorous protocol to ensure comprehensive coverage and minimize bias. The review process adhered to PRISMA guidelines (Page et al., 2021) and employed a snowballing technique to identify further relevant studies from the references of included papers. Studies were evaluated against the What Works Clearinghouse standards (WWC, 2022) to ensure high methodological quality. A detailed codebook was developed to categorize the prompts in the included studies. This codebook extended existing categories with nuanced subcategories to allow for a more fine-grained analysis of prompt features. Two independent raters coded all studies, achieving high inter-rater reliability (Cohen's kappa = 0.89), with any discrepancies resolved through discussion (Cohen, 1960).

Building upon this framework, hypotheses about prompt effectiveness were deductively formulated and tested through meta-analysis using random-effects models (manuscript 2). The meta-analysis aimed to synthesize the findings from the included studies to estimate the overall effect of digital learning prompts on learning achievement and explore potential moderating factors. Cohen's d served as the primary measure of effect size to synthesize findings. For studies with multiple experimental groups compared to a single control group, the "aggregate" function in R calculated study-wide mean effect sizes (Viechtbauer, 2010). A random effects model using restricted maximum likelihood method computed the weighted mean effect size, utilizing the R packages metafor and dmetar (Viechtbauer, 2010).

Prior to conducting the moderator analyses, a correlation analysis was performed to examine the relationships between potential moderator variables. This step was crucial in identifying potential collinearity issues and informing the selection of appropriate control variables for subsequent analyses (Borenstein et al., 2021). Subsequently, a series of moderator analyses were conducted using a meta-regression approach based on mixed effects models to examine the impact of different prompt features and study characteristics. This allowed for simultaneously considering multiple potential moderators while controlling for confounding variables (Higgins et al., 2019). Separate meta-regressions were performed for studies with multiple treatment groups to assess the specific effects of each prompt feature. In cases where the number of studies in a specific category was too limited (e.g., two studies) for the categorical moderators, those categories were omitted from the moderator analysis to ensure robust results.

Publication bias was evaluated utilizing funnel plots, Egger's regression test (Egger et al., 1997), and a likelihood ratio test using the Vevea and Hedges weight-function model (1995). The trim-and-fill method (Duval & Tweedie, 2000) was utilized to adjust effect size estimates for potential publication bias. Additionally, a sensitivity analysis using the "influence" function (Viechtbauer, 2010) was performed to identify any studies disproportionately affecting the overall effect size (Borenstein et al., 2021).

3.2.2 Paper 2 (Development of a Personalized Prompt Design)

The second paper adopts a conceptual-inductive approach to analyze the *Luca* office simulation, with a particular focus on its personalized prompt design. This research utilized design-based research (DBR), a systematic and adaptable approach designed to enhance educational methods through continuous evaluation, design, development, and implementation (Wang & Hannafin, 2005; Collins, 1992). DBR is a method used to iteratively design, test, and refine educational innovations through systematic analysis of the implementation in real-world contexts (Zheng, 2015). This approach was chosen for its capacity to address practical challenges in educational environments while contributing to both theory and practice.

Building on this foundation, the study followed Wang's (2020) DBR reporting framework, which includes: identifying problems, creating experiments, facilitating and assessing designs, detailing findings, and proposing ideas for subsequent phases. The development and testing of *Luca* office and the respective prompt design followed four phases, adhering to the iterative nature of DBR. Table 7 provides comprehensive information about the participants, research tools, and timeline for each phase.

Table 7: Design-based research information in each phase (adapted from Wang, 2020)

	Participants	**Research tools**	**Period**	**Purpose**
Phase 0	DBR Team (3 research scientists)	Literature Review and Meta-analysis (Paper 1)	07.2021 - 10.2021 02.2022 - 10.2022	Preparing stage
Phase 1	DBR Team	None	10.2021 - 11.2021	Prompt development and first testing
Phase 2	DBR Team, 4 research assistants, 2 teachers	(open-ended) questionnaire; Observation	11.2022 – 12.2021	Redefining prompt design and testing
Phase 3	647 VET students	(open-ended) questionnaire; log-file analysis; analysis of problem-solving competence	12.2021 – 05.2022	Evaluate the complexity of the working scenario and effectiveness of the prompt design
Phase 4	222 VET students	(open-ended) questionnaire; log-file analysis; analysis of problem-solving competence	10.2022 – 01.2023	Evaluate the complexity of the working scenario and effectiveness of the prompt design

In the preparation phase (Phase 0), a systematic review and meta-analysis (Paper 1) were conducted to analyze the prompt features in various technology-enhanced learning environments and review the related learning theories (e.g., CLT, CTML) and strategies for addressing the possible learning problems. These observations highlighted a lack of focus on prompt design, with studies often providing only prompt examples. They did not cover detailed design and implementation strategies. Paper 2 aims to fill this gap by providing detailed strategies for creating personalized prompts.

Building on these findings, in Phase 1, digital learning prompts were implemented to align with the "supplier selection" working scenario in *Luca* office. The prompts were created and tested (refer to Section 2.3.2) by three research scientists, following the rules of thumb outlined in Paper 1 (p. 23) and incorporating trigger conditions (see Appendix Table A1).

During Phase 2, four research assistants conducted initial testing of the working scenario, examining the prompts for both content and timing. Subsequently, two experienced subject teachers reviewed the working scenario and provided feedback on the prompt design. Considering the feedback on the scenario's complexities, the errors encountered, and suggestions about the current prompts, adjustments were made for greater clarity, trigger conditions were made timelier, and their overall number was decreased (see Appendix Table A2; redefined or added sentences marked red).

Following these refinements, in Phase 3, the personalized prompt design was tested with 647 vocational education students in the first wave of data collection. Participants were randomly assigned to either the experimental group, which received prompts (n = 318), or the control group, which did not receive prompts (n = 329). The discrepancy in sample size resulted from data loss attributed to technical issues. This study used the DBR methodology, incorporating diverse measurements, tools, and data collection sources (refer to Rausch et al., 2021 for an overview). It gathered both quantitative and qualitative data regarding students' problem-solving processes during the learning task. Additionally, a questionnaire featuring open-ended questions was used to collect qualitative insights about the working scenario and prompt design. This was the most comprehensive redefinition of the prompt design. As a result, the DBR team determined that combining certain prompts and providing more explicit support would be beneficial in addressing the challenges with instructions in *Luca* office (see Appendix Table A3). For instance, the event for the intermediate state and the associated prompts have been removed (Appendix, Table A2, No. 1). Instead, an event query with a prompt for a common error (work efficiency) was added to better support learners (Appendix, Table A3, No. 1). Individual prompts that received little attention in phase 2 or were triggered together were also merged (Appendix, Table A3, No. 2).

During Phase 4, the prompt design underwent another round of testing with 222 vocational education students as part of the second data collection wave. Participants were randomly allocated to either the experimental group, which received prompts (n = 109), or the control group (n = 113). Both quantitative and qualitative data were collected similarly to Phase 3. The DBR process was stopped after Phase 4 results confirmed the design. These results demonstrated that the revised prompt design significantly enhanced students' problem-solving skills compared to the control group that did not receive any prompts. Nevertheless, feedback from learners indicated that the prompt design is not yet optimal and has room for improvement in future iterations.

To contextualize these findings within the broader research agenda, Paper 2 focuses on the development of the prompt design, while the empirical testing of the prompt design (Phases 3 and 4) is currently taking place and will be published in another paper. These insights were taken up in the discussion section of this dissertation as an outlook.

3.2.3 Paper 3 (Effectiveness of Immersive Virtual Reality)

The third paper investigates the effectiveness of immersive virtual reality in VET, specifically focusing on the domain of warehouse logistics. To address the first research question, a randomized controlled trial design was employed, comparing IVR-based learning to traditional paper-based methods. Participants were randomly assigned to either the experimental (IVR) group or the control (paper-based) group. Both groups completed pre-tests to assess prior domain-specific knowledge, followed by their respective interventions, and then post-tests to measure knowledge acquisition. Before conducting the main analyses, assumptions of homogeneity of variance (Levene's test) and normality (Skewness and Kurtosis) were tested. Independent sample t-tests were used to compare the pre-test and post-test scores between the two groups. Additionally, perceived knowledge gains were measured through a questionnaire, analyzing the differences between groups using t-tests. A power analysis was performed to evaluate the statistical power of their study. Missing data were evaluated for being Missing Completely at Random (MCAR) via Little's test. Complete Case Analysis (CCA) was used to handle missing data, ensuring the robustness of the findings.

For the second research question, the relationship between objective and subjective knowledge acquisition was examined. Pearson correlation coefficients were separately calculated for both groups to assess the strength of this relationship. Then Fisher's z-transformation was applied to these coefficients, enabling more robust statistical inference and facilitating comparisons across samples. The q-statistic was used to represent the difference between Fisher's z-transformed correlations, allowing for a comparison of the strength of these relationships between the IVR and paper-based groups.

To examine the third research question, the study employed detailed surveys based on measurement instruments developed by Tcha-Tokey et al. (2016) and the framework proposed by Fokides and Antonopoulos (2024). Descriptive data were compared, and independent sample t-tests were conducted to compare the scores on these scales between the experimental and control groups for most measures. However, Welch's t-tests were performed for scales where the assumption of homogeneity of variances was violated. To ensure the robustness of their findings, the researchers applied a Holm-Bonferroni correction to control for multiple comparisons across all tests and questionnaire scales.

4 Paper Publications

Before presenting Sections 4.1 through 4.3, which include the original published papers, this section provides a brief overview of their contents (as summarized in Table 1). While Papers 1 and 2 are shown in their published form, Paper 3 is included as the initial manuscript submitted in March 2024, which is currently undergoing major revisions.

As previously discussed (Figure 1), Paper 1 lays the theoretical and empirical groundwork for digital learning prompts across various domains and settings, pinpointing key features of prompts that enhance learning outcomes. This understanding guides the methodologies used in the following studies on personalization. Paper 2 utilizes these insights by applying a design-based research approach to create and implement a personalized prompt design within the *Luca* office simulation for VET. Paper 3 extends this research into immersive virtual reality, comparing an adaptive IVR environment with traditional learning methods in vocational logistics education, examining both cognitive and affective learning outcomes, and validating the effectiveness of personalized learning approaches across increasingly immersive learning environments.

Paper 1 (manuscripts 1 and 2) employs a systematic literature review and meta-analysis. It analyzes 68 experimental studies using random-effects models and meta-regression to examine the effectiveness of digital learning prompts. The study uses Cohen's d for effect size calculation and conducts moderator analyses on various prompt features and study characteristics. The study's primary concern is understanding the effectiveness of digital learning prompts in educational settings and identifying the factors that affect their impact on learning achievement.

Paper 2 uses a conceptual-inductive approach. It analyzes the *Luca* office simulation's functions and provides a concrete application example of a personalized prompt design. The study demonstrates how log data can be leveraged for adaptive prompting strategies, connecting theoretical insights with practical applications.

Paper 3 employs a randomized controlled trial involving 72 vocational students. It utilizes a pre-test/post-test design to compare IVR-based learning with traditional paper-based methodologies. The research uses t-tests, correlation analyses, and questionnaires to assess knowledge acquisition, perceived learning, motivation, and immersion. The primary concern of this study is evaluating the effectiveness of IVR as a learning tool in vocational education and comparing it to traditional methods in terms of knowledge acquisition and motivational impact.

4.1 Paper 1 (Manuscripts 1 & 2): Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement

Thomann, H., & Deutscher, V. (2025). Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement. *Educational Research Review*, 47, 100686. https://doi.org/10.1016/j.edurev.2025.100686

Educational Research Review 47 (2025) 100686

Contents lists available at ScienceDirect

Educational Research Review

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

Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement

Herbert Thomann *, Viola Deutscher

Chair of Business Education – Digital Vocational Learning, Georg-August University Göttingen, Germany

A B S T R A C T

We present results from meta-analyses of 68 experimental studies on digital learning prompts, examining what features constitute an effective prompt for learning achievement. First, a systematic review reveals the different features of prompts used across various studies. Second, a quantitative meta-analysis using a random-effects model shows that digital prompts significantly enhance learning achievement ($d = .394$), though adjustment for publication bias yielded a more conservative estimate ($d = .220$). Their effectiveness is largely moderated by prompt design features. Based on meta-regression models, we find the highest efficacy for action-based prompts (rule-based AI) ($d = .465$), prompts designed for specific learner groups ($d = .513$), and combinations of generic and directed prompts ($d = .571$). Regional differences were pronounced, with studies from East Asia showing substantially larger effects than European settings. The effectiveness of learning prompts is further moderated by the learning domain and target group. Our findings reveal that prompts are not a universal solution but require thoughtful implementation. We recommend implementing action-based prompts triggered by learner behavior, using log data to tailor prompts to expertise levels. Designers should keep prompts concise to minimize cognitive load. We advise combining generic and directed prompts based on learning goals. Finally, it is essential to ensure learners are familiar with prompt use. These evidence-based guidelines can help optimize digital learning prompts to support diverse learner needs.

1. Introduction

1.1. Relevance and rise of digital learning prompts

Teachers usually personalize their classroom teaching by giving additional support to struggling learners while further challenging those making good progress. Many even consider the ultimate goal of education to provide each student with the kind of learning experience they need at any given time based on individual interest, prior knowledge, and capabilities (Holmes et al., 2018). In this context, scaffolding refers to teaching aids and strategies provided by educators to help learners accomplish tasks they couldn't manage alone (Reiser & Tabak, 2014). Of course, such personalized learning approaches have been of great interest to learning designers and pedagogues since long before the emergence of digital learning (Plass & Pawar, 2020). However, automation options in digital learning environments have facilitated a significant paradigm shift from one-size-fits-all to personalization (Aleven et al., 2016). In particular, advancements in digital learning have enabled the seamless incorporation of scaffolding in the form of so-called 'prompts' within digital environments (Puntambekar & Hubscher, 2005). Prompts are hints in the form of questions, suggestions, and feedback that appear during digital learning and are intended to promote the application of relevant processing strategies (e.g., Quintana et al., 2004, pp. 337–386; Wirth, 2009). Prompts are designed to activate processing strategies that learners already possess

* Corresponding author. Georg-August University Göttingen Chair of Business Education – Digital Vocational Learning Platz der Göttinger Sieben, 5 37073, Göttingen, Germany.

E-mail address: herbert.thomann@uni-goettingen.de (H. Thomann).

https://doi.org/10.1016/j.edurev.2025.100686
Received 28 March 2024; Received in revised form 21 March 2025; Accepted 21 April 2025
Available online 22 April 2025

but may not spontaneously apply during learning or may apply inadequately (Renkl & Scheiter, 2017). They can take various forms, such as instructional text, multimedia content, and interactive activities, and can be designed to achieve different learning goals. The level of personalization and explicitness of help provided can also vary among different types of prompts.

Despite their growing implementation, there remains a critical gap in our understanding of prompt effectiveness. This gap is particularly concerning as educational institutions increasingly invest resources in digital learning environments without sufficient empirical guidance on optimal prompt design. By providing tailored assistance across different proficiency levels, prompts have the potential to markedly improve the educational experience and accommodate diverse learner needs (Schumacher & Ifenthaler, 2021; Ifenthaler, 2012; Davis, 2003), yet this potential cannot be fully realized without more systematic knowledge.

While numerous studies have examined digital learning prompts, there is a notable lack of systematic understanding regarding which prompt features are most effective and under what conditions (Renkl & Scheiter, 2017). Existing research tends to focus on individual implementations rather than providing comprehensive analyses of prompt design principles that could guide future developments. This fragmentation in the literature has led to inconsistent implementation practices and missed opportunities for optimizing learning outcomes through evidence-based prompt design.

The scarcity of research on creating and effectively utilizing personalized prompts in digital learning environments represents another significant limitation in current knowledge (Guo, 2022; Zheng et al., 2022). Both earlier and recent studies primarily used standardized prompts, where the content remained identical for each learner, and the design was mostly based on insights drawn from existing literature rather than learner-specific characteristics (Lim et al., 2024; Li et al., 2023; Guo, 2022; Bannert et al., 2009). This approach fails to leverage the full potential of digital environments to provide truly adaptive support.

Research so far on prompt effectiveness has been inconclusive. The effectiveness of prompts in digital learning environments represents a complex phenomenon influenced by multiple interacting factors. While prompts have been widely implemented as digital scaffolding tools to enhance learning, their impact varies considerably depending on their timing, presentation form, specificity, and personalization. This complexity explains the seemingly contradictory findings across studies, where prompts are highly effective (e.g., Karakostas & Demetriadis, 2011), ineffective (e.g., Bartholomé & Bromme, 2009), or even detrimental to learning (e.g., Cavanagh et al., 2016). Rather than considering prompts as universally beneficial tools, they should be understood as contextually dependent interventions whose effectiveness is moderated by factors such as the learning domain, learner characteristics, target group, the digital learning environment, and their design specifications.

Given these inconsistent findings and knowledge gaps, this study aims to systematically review prompts, their specific features, and their effect on learning achievement. In the first step, a systematic review was conducted to provide the theoretical basis for the subsequent meta-analysis. Here, the following research question was raised: (1) Which types of prompts are distinguished throughout the literature? In the second step, a meta-analysis aimed to investigate the effectiveness of educational prompts on learning achievement. Here, we raised two research questions: (2) What is the overall effect of prompts on learning achievement? (3) How is the effectiveness moderated by prompt features and study demographics?

By addressing these questions, our study presents the first systematic review and meta-analysis of digital learning prompts with such a fine-grained analysis of prompt features across all learning domains. Previous research has notably absented this level of detailed examination. Our meta-analysis not only synthesizes existing knowledge but also identifies specific prompt features and contexts that optimize learning achievement, thereby addressing a critical need for more systematic understanding in this evolving field.

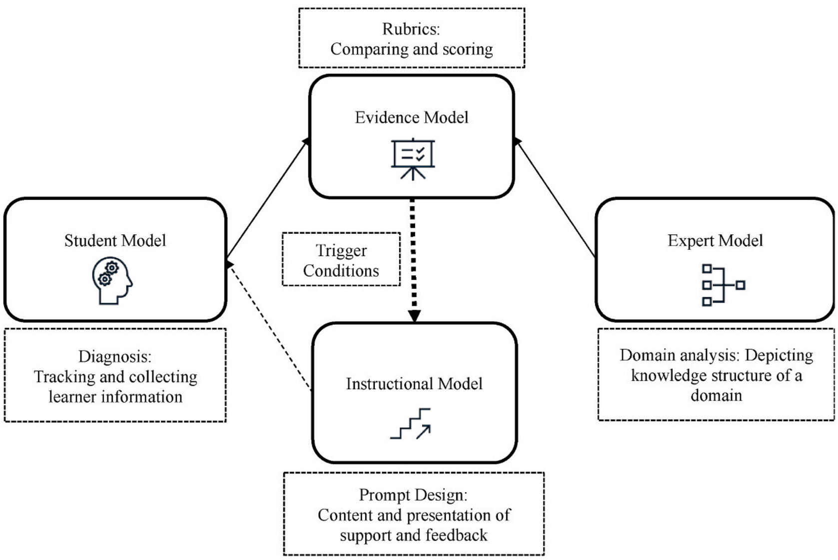


Fig. 1. Core model for assessing the effectiveness of prompts.

2. Literature review

2.1. Theoretical distinction of prompt features

To understand and classify learning prompts and their effectiveness, a theoretical model based on Clariana and Hooper (2012) and Mislevy and Riconscente (2005) is presented in Fig. 1. In our model, learning prompts work through the interaction of four interconnected components in adaptive learning environments: (1) the "Expert Model", outlining correct task solutions as a result of a domain analysis; (2) the "Student Model", detailing the learner's problem-solving process and behaviors through tracking; (3) the "Evidence Model", detecting deviations between the Student and Expert Models using scoring rubrics; and (4) the "Instructional Model", offering didactic support in the form of prompts to address these deviations. Prompts can defer regarding several design characteristics, including their content, media, and triggering conditions in a given learning environment. Digital learning environments offer particular flexibility in triggering prompts based on the learner's progress, behavior, and needs, depending on the complexity of the learning environment. Examples of digital learning environments where prompting occurs include traditional e-learning systems, intelligent tutoring systems (ITS), and learning simulation systems, such as educational games, virtual reality (VR), and augmented reality (AR) experiences.

While some studies have compared different types of prompts (e.g., Davis, 2003; Ifenthaler, 2012; Wirth, 2009), few have examined their effectiveness in promoting learning achievement (Renkl et al., 2015). Moreover, the design and presentation of prompts within specific learning contexts warrants closer examination. Thus, we examined 4 key prompt feature categories: (1) cognitive type and type of learning achievement (What is prompted?), (2) grouping of learners (Who is prompted?), (3) trigger conditions for prompts (When does prompting occur?), and (4) formulation and presentation of prompts (How are prompts formulated and presented, and do they include audiovisual content?). Table 1 provides an overview of these categories, along with examples from the studies included in our analysis. The systematic review section can be considered the theoretical foundation for the subsequent quantitative meta-analysis.

Table 1
Overview of prompt features.

	Categories	Subcategories	Definition	Example
What	Cognitive	none	Prompts that support and promote learners' cognitive processes.	One of your answers is incorrect. Evaluate the expressions below. Write each response as an integer or as a fraction. (ID 75)
	Meta-cognitive	none	Prompts that activate the monitoring and control of learners' meta-cognitive activities.	How can I best organize the structure of the learning contents? (ID 85)
	Non-cognitive	none	Prompts that seek to increase learner motivation by praising the student or providing guidance on regulating their motivation.	The e-module was designed for all kinds of learners with different skills and backgrounds. Please feel assured that with appropriate effort, all participants should be able to complete the e-module successfully. (ID 70)
Who	All Learners	none	Prompts that are available for all learners and have the same content.	Think about the personal relevance of the learning material. (ID 41)
	Group of Learners	Grouping based on learner's activities. Grouping based on previous test/ questionnaire: Prior knowledge Learner personality	Prompts that are only available for a specific group of learners.	Dear participant, we would like to advise you to look at the interactive visualization a little bit longer. (ID 76)
	Individual Learner	Self-created prompts	Prompts which are based on a individuals learner's behavior or self-created prompts.	I plan my next learning steps. I reflect on my learning strategy. I make sure to cover all learning goals (ID 99)
When	Time-based	During the learning sequence Previous to the learning sequence After the learning sequence	Prompts that appear after a certain amount of time	Before you start to learn, you should first prepare yourself. (ID 95)
	Action-based	Self-selected Previous Activity (navigation decision, viewed content) Incorrect/Correct Solution	Prompts that are based on the learners' activities	Hi, here I am again! You have been busy for a while, but you still have not taken any notes. (ID 174)
How	Generic	none	Prompts that are worded in general terms and can be used for a variety of digital learning content.	Which are the main points in your opinion? (ID 51)
	Directed	none	Prompts that are content-specific and contextualized.	Why do you calculate the total acceptable outcomes by multiplying? (ID 47)

2.2. Hypotheses regarding the effectiveness of digital prompts on learning achievement

2.2.1. Category "what"

The content of prompts ("What") can promote the execution of cognitive, meta-cognitive, and non-cognitive learning activities. Following the classifications by Azevedo et al. (2005), cognitive prompts are designed to support problem-solving and information processing, while meta-cognitive prompts aim to activate self-monitoring, goal setting, planning, and evaluation, encouraging active learner engagement and self-awareness. Non-cognitive prompts seek to increase learner motivation by praising the student or providing guidance on regulating their motivation; but they are not directly related to the acquisition of knowledge or cognitive processes. In the last years, various studies have examined the effectiveness of cognitive and meta-cognitive prompts, whereas only very few studies have done so for non-cognitive (motivational) prompts (e.g., Backhaus et al., 2017; Chen, 2014; Daumiller & Dresel, 2019). Despite this empirical gap, research conducted by Daumiller and Dresel (2019) revealed the effectiveness of non-cognitive prompts on both short-term and long-term learning achievement. Based on the limited empirical evidence of these prompt types, we anticipate that cognitive, meta-cognitive, and non-cognitive prompts would similarly affect learning achievement.

When considering the impact on specific types of knowledge acquisition, the empirical evidence is more mixed and context dependent. Procedural knowledge and declarative knowledge may be affected differently by prompts. Procedural knowledge is knowledge about how to do something, while declarative knowledge is knowledge about facts and information (Greeno et al., 1984). Research has shown inconsistent findings regarding the effectiveness of prompts for different types of knowledge acquisition. While some studies demonstrate that prompts can be effective for both declarative and procedural knowledge (e.g., Kauffman, 2004), others have shown no effect for either type (see Gentner & Seufert, 2020). Furthermore, studies examining multiple learning outcomes have revealed differential effects, with some showing effectiveness for declarative knowledge but not for procedural knowledge (see Berthold et al., 2011). These varying results suggest that the effectiveness of prompts may depend on the type of knowledge being acquired. Thus, we hypothesize that there will be differences in prompt effectiveness between declarative and procedural knowledge acquisition types.

2.2.2. Category "who"

The "Who" of prompts refers to the selection of learners who receive a prompt and describes the conditions of prompt adaptivity in relation to the number of learners. This aspect is divided into three sub-categories: "all learners," "group of learners," and "individual learners." The category "all learners" includes prompts available and formulated identically for all learners. The category "group of learners" includes prompts differing across groups based on learner characteristics or learner activity. For example, learners may receive different types of prompts depending on their self-classification as male, female, or diverse or how they respond to a learning task. Furthermore, the classification of groups with different prompts can also be conducted by prior tests and questionnaires. The category "individual learner" contains prompts based on the individual learner's behavior or known characteristics. These individual adaptive prompts are designed for one specific learner.

Both theoretical foundations and empirical studies inform our hypotheses in this category. Theoretically, the presumed degree of learning *effectiveness* increases as the prompts are more personalized toward the learner's characteristics and behavior. Thus, prompts for groups of learners or individual learners should be more effective in enhancing learning achievement across diverse educational settings. This theoretical expectation is supported by recent meta-analyses conducted by Guo (2022) and Zheng et al. (2022), which emphasize the importance of catering prompts to learners' ongoing progress and providing tailored support based on the ongoing diagnosis of their learning. However, it should be noted that prompts developed only for groups of individuals reach fewer learners than prompts designed for all learners and, therefore, are less efficient regarding their benefit-cost ratio. As prompts become more personalized, the efficiency of a single designed prompt in impacting all learners may decrease. Nevertheless, we assume that more individualized prompts are more effective in promoting learning achievement.

This hypothesis is grounded in both theory and emerging empirical evidence. Theoretically, individualized prompts enable learners to navigate the complexities of a digital learning environment while receiving tailored support when encountering challenges. This safety net allows for an active learning experience where learners can work through task scenarios, correct errors, and receive guidance toward the desired learning path, as proposed by Mead et al. (2019). Empirically, previous research has shown that prompts often lack adequate personalization to meet the ongoing learning needs of students (Bannert & Mengelkamp, 2013; Lallé, Conati, Azevedo, Mudrick, & Taub, 2017), suggesting untapped potential for improvement. Adaptive prompt designs are still uncommon (e.g., prompts based on previous tests or learner characteristics), and more research is needed to develop and test personalized prompts that can provide tailored guidance (Bernacki et al., 2021; Guo, 2022; Zheng, 2016).

The category "Who" can also refer to the target group's ability to handle and work with prompts. Zheng et al. (2022) and Cai et al. (2022) have posited that technology-enabled personalized learning holds the potential to cater to the diverse educational needs of students across all levels of education and regions. Hence, we anticipate that prompts will be equally effective for all learners, regardless of their educational level or region (Europe, East Asia, and North America). Major et al. (2021) and Cai et al. (2022) conclusively demonstrated the effectiveness of personalized digital learning across learning domains, so we anticipate prompts will enhance learning achievement in diverse domains. Additionally, immersive learning environments, such as simulations and ITS, offer more complex and engaging experiences and promote self-regulated learning more than traditional (less immersive) digital learning environments (see Zimmerman and Moylan, 2009; Moos & Bonde, 2016). Prompts can help struggling learners, especially in such complex environments. Thus, we anticipate that prompts will be more effective in immersive learning environments than in traditional (less immersive) digital learning environments, though we acknowledge that empirical evidence specific to this comparison is still emerging.

2.2.3. Category "when"

The category "When" differentiates between two categories: "time-based" and "action-based" prompts, based on a framework by Wirth (2009). Both theoretical frameworks and empirical studies inform this categorization, though the latter are still developing for certain subcategories.

Time-based prompts appear in the digital learning environment after a certain amount of time, similar to how a kitchen timer works (Robertson et al., 2015). They can occur before, during, or after the learning sequence and should be timed to provide support when the learner needs it, thus avoiding additional cognitive processing (Thillmann et al., 2009). The theoretical foundation for understanding time-based prompts comes from "Cognitive Load Theory" (CLT), which posits that working memory's finite capacity necessitates efficient processing of new information before its storage in long-term memory, which possesses a nearly unlimited capacity (Chen et al., 2023; Sweller, 1988).

Kalyuga (2011) differentiates cognitive load into intrinsic (ICL), linked to task complexity and learner expertise (e.g., previous knowledge), and extraneous (ECL), resulting from poor instructional design. Increased ECL negatively affects learning and should be as low as possible (Sweller, 1988). This is especially relevant in digital learning environments, where learners are exposed to various information resources. According to the Cognitive-Affective Theory of Learning with Media (Moreno & Mayer, 2007), strategic regulation of learning resources is crucial to counteract cognitive overload.

Applying these theoretical frameworks to prompt timing, timely and well-designed prompts are essential in preventing further increase in ECL, which could otherwise lead to learner frustration and disrupt motivation and the flow of learning (Chen et al., 2023; Hawlitschek & Joeckel, 2017; Thillmann et al., 2009). Additionally, ICL's impact is especially critical for novices (and weaker learners), who face a higher risk of cognitive overload in complex learning situations due to their limited prior knowledge. This emphasizes the necessity of instructional support that matches learners' knowledge levels to support effective learning processes (Wang & Lajoie, 2023).

Action-based prompts, in contrast, are based on individual learners' behavior. These prompts occur based on previous learner activities such as navigation decisions, content viewed, and correctness of answers in the digital learning environment (see Bernacki et al., 2021). Furthermore, they can occur when learners click on a button that provides a prompt (self-selected). The theoretical basis for action-based prompts is their potential to provide more timely and personalized support. Empirical evidence for more personalized action-based prompts is emerging but promising. Recent studies, such as Munshi et al. (2023), Li et al. (2023), and Deutscher et al. (2022), embedded action-based prompts in a digital learning environment using log and performance data. This analytics-based approach provides tailored support to individual learners based on their specific learning needs and progress. These personalized prompts can be delivered in real-time, providing learners immediate feedback and support as they engage in the learning environment (e.g., rule-based AI prompts). However, such analytic-based prompts are still uncommon, and further research is needed to develop and test them. Combining theoretical principles of cognitive load and personalized support with the emerging empirical evidence, we anticipate action-based prompts will have a greater effect than time-based prompts, as they are personalized and presented timelier based on the learner's behavior.

2.2.4. Category "how"

In the category "How", prompts are classified as "generic" or "directed" following Davis' (2003) differentiation. Generic prompts are worded in general terms and applicable to various digital learning contexts. In contrast, directed prompts are content-specific and contextualized, guiding learners to a precisely desired action within a specific learning activity.

Empirical Research indicates the impact of prompt specificity varies for different learner groups (e.g., Belland et al., 2017; Davis, 2003; Ifenthaler, 2012; Renkl et al., 2015; Serge et al., 2013). According to Ifenthaler (2012), directed prompts effectively reminded learners of forgotten information and improved learning performance, making them more suitable for learners with lower levels of prior knowledge. In contrast, generic prompts were effective for learners with higher levels of prior knowledge and learning skills, as they can take advantage of the higher degree of autonomy provided by such prompts. This context-dependent effectiveness is further supported by Belland et al. (2017), who found no statistical significance between generic and directed prompts in their meta-analysis in STEM education. Although Zheng et al. (2022) identified the strongest effects when combining both prompt types, the variability in outcomes across different contexts suggests that the effects of individual prompt types may counterbalance each other in heterogeneous learning environments. Thus, based on the mixed empirical evidence across different contexts, we expect that there will be no difference in effect sizes for specific or generic prompts.

Prompts can also differ regarding their presentation form in text-only prompts (without response options) and prompts enriched with visual media formats such as audio, video, and images. The theoretical basis for media-enriched prompts comes from established educational frameworks. Enriching prompts with multimedia align with the principles of situated learning and anchored instruction approach by creating an engaging and immersive learning experience (e.g., Langone, 1998; Oliver & Herrington, 2000, pp. 178–191; Young & Kulikowich, 1992, pp. 1–21; Cognition and Technology Group at Vanderbilt, 1997).

The empirical evidence on multimedia-enriched prompts is mixed across different contexts. Previous research suggests multimedia-enriched prompts can improve learning compared to text-only prompts (Berthold & Renkl, 2009; Brünken et al., 2005; Peeck, 1993). In this context, Mayer's "cognitive theory of multimedia learning" (2021) provides valuable insights for optimizing multimedia support. For instance, it recommends aligning verbal and visual information, segmenting content, using cues for emphasis, and avoiding redundancy. Well-designed prompts based on Mayer's multimedia principles can enhance learner attention and mitigate cognitive load. Poorly designed prompts, on the other hand, can distract learners and impede learning.

This potential for both benefit and distraction may be linked to cognitive overload (Sweller, 1988; Thillmann et al., 2009). Media richness may require learners to filter information and disrupt their learning process, perceiving prompts as additional tasks outside the core learning environment. This illustrates the complexity of prompts' interaction with multimedia content. Consequently, based on theoretical principles and empirical evidence, we anticipate no differences in the effectiveness of media-enriched and text-only prompts.

2.3. The current meta-analysis

The subsequent meta-analysis section differed in several ways from previous studies including Lazonder and Harmsen (2016), Zheng (2016), Belland et al. (2017), Bisra et al. (2018), Chernikova et al. (2019), Jansen et al. (2019), Chernikova et al. (2020), Cai et al. (2022), Guo (2022) and Zheng et al. (2022). We focused solely on digital learning environments and a broad participant range, including higher education, high school, elementary, vocational students, and working adults, diverging from works like Guo (2022), which limited its scope to school settings, and Belland et al. (2017), Chernikova et al. (2019, 2020), Cai et al. (2022) focusing on specific learning domains or environments. It differs from Jansen et al. (2019), which concentrated on Self-Regulated Learning outcomes, and from Zheng (2016), Lazonder and Harmsen (2016), and Bisra et al. (2018), who reviewed various forms of instructional support alongside prompts. Moreover, while studies like Zheng (2016), Bisra et al. (2018), Jansen et al. (2019), and Guo (2022) were limited to meta-cognitive prompts, our scope encompasses a variety of cognitive prompt types. Thus, our meta-analysis addresses a gap in the literature by assessing prompt effects across cognitive types and design features, informing prompt optimization.

3. Method

3.1. Literature search

We searched the Web of Science and Social Science Citation Index (SSCI) databases, focusing on peer-reviewed articles published between 1999 and 2022 that met our criteria, as outlined in Table 2. This time frame was chosen based on Lin and Lehman's influential study (1999) highlighting the benefits of meta-cognitive prompts, and studies conducted before this period no longer meet current technical standards for digital prompts. Studies were evaluated against the What Works Clearinghouse (WWC, 2022) standards, focusing on those assessing the impact of prompts on learning achievement in digital environments, irrespective of subject area. So, we could analyze a broader spectrum and account for the fact that prompts with different learning contents can be found in digital learning environments. The search, conducted from July 2021 to December 2022, utilized 12 key terms related to digital personalized learning (personalized learning, adaptive learning, digital learning, e-learning) and prompts (prompt, hint, intervention, feedback, nudge, scaffold, assistance, guidance), combined in 32 ways to ensure comprehensive coverage and inclusivity of terminological variations.

3.2. Literature selection

Our literature search aimed for high sensitivity to identify all potentially eligible studies, ensuring broad coverage through our chosen search terms. Following PRISMA guidelines (Page et al., 2021), we identified 12,989 studies, including duplicates (Fig. 2), using broader keywords for prompts to capture a wide range of studies. Table 3 provides a detailed overview of the hits per database based on the selected term combinations. Screening based on titles and abstracts refined this to 160 potentially relevant studies, primarily excluding those without any prompts as defined by our search criteria or the focus on non-digital settings (Table 2). After removing duplicates ($n = 18$) and adding studies via snowballing ($n = 30$), 142 studies underwent full-text review. These were further narrowed down by excluding those without control groups ($n = 24$), only measuring affective learning outcomes ($n = 15$), lacking randomized allocation ($n = 10$), non-digital settings ($n = 9$), collaboration during the treatment ($n = 8$) and identical sample groups ($n = 2$).

Lastly, studies were categorized according to What Works Clearinghouse (WWC) standards into three groups: not meeting, meeting with reservations, and fully meeting the standards. After applying these criteria, six studies failing to meet WWC standards were excluded due to issues like baseline non-equivalence and high attrition. Ultimately, 68 high-quality studies were included in our meta-analysis, focusing on the effect of prompts in digital learning environments on learning achievement. All but one study achieved the highest WWC rating. For Leemkuil's and de Jong's (2012) study, which reported high attrition ($10.64 > 6.1$), a difference-in-difference adjustment (.08) was made to meet WWC standards with reservations.

3.3. Coding of the included studies

We developed a detailed codebook to categorize the prompts in the studies, extending existing categories with subcategories for nuanced analysis (Table 4). Studies featuring multiple types of prompts were classified at a higher personalization level to ensure the overall learning design related to higher personalization options. Two raters independently coded all studies, achieving a Cohen's kappa inter-rater reliability of .89, indicating reliability (Cohen, 1960). Differences were resolved through discussion. A detailed list of all included studies, their prompts and WWC coding is presented as an *Excel* spreadsheet (supplementary material).

3.4. Effect size extraction and analysis

After the coding stage, effect sizes were extracted from each study to calculate the overall effect of prompts and respective

Table 2
Inclusion/exclusion criteria.

Criteria	Inclusion	Exclusion
Publication date	1999–2022	Prior to 1999
Publication type	Peer-reviewed journals	Book chapters, dissertations, presentation papers
Language	Journal article was written in English	Other languages were not included
Prompt concept	Studies should include an experimental group using prompts only and a control group without receiving prompts	Studies with only one experimental group with prompts
Outcome measurement	Learning achievement should be measured via post-test	SRL activities and only qualitative study measurements (think-a-loud, eye-tracking etc.)
Subject area	All subject areas (psychology, mathematics etc.)	None
Setting	Educational settings and workplace settings	None
Medium	Prompts, learning sequence and tests were included in a digital learning environment	Traditional classroom lectures or studies in which prompts were presented in paper-pencil form

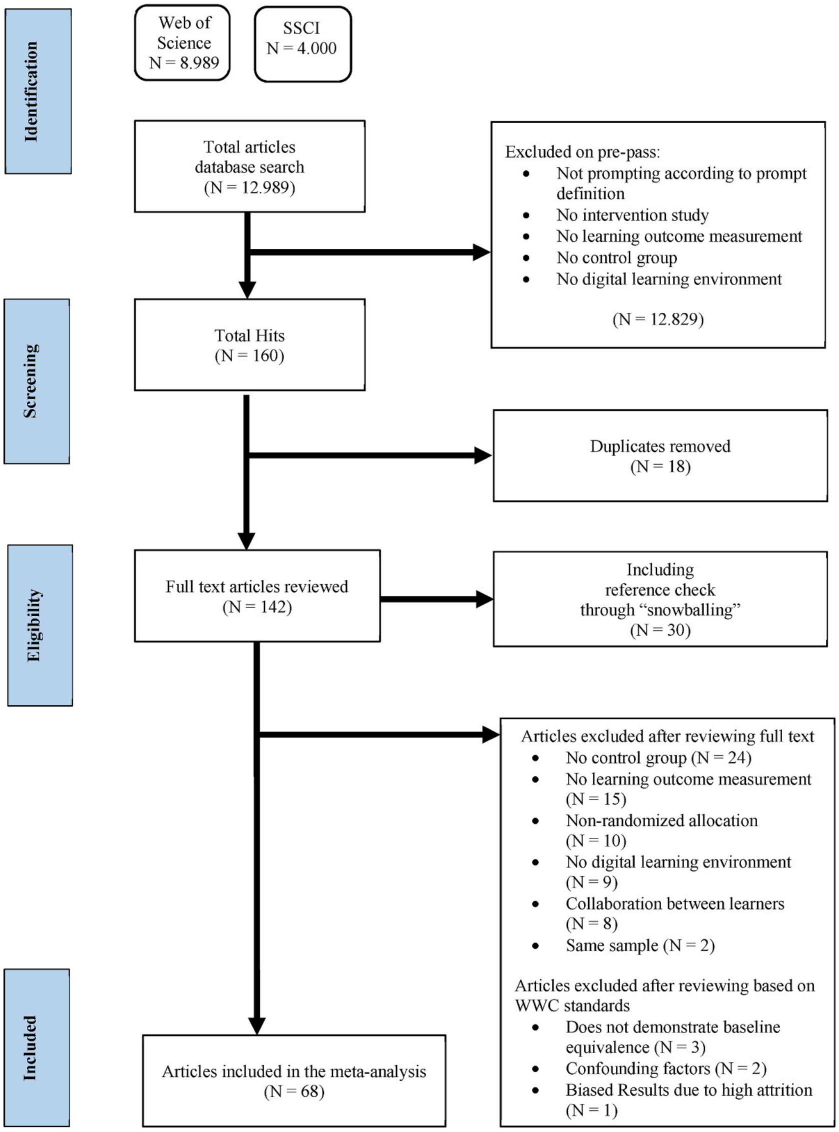


Fig. 2. Prisma flow diagram.

Table 3
Literature search overview per database and keywords.

Search Criteria		Databases		∑
Key Term 1	Key Term 2	Web of Science	Social Science Citation Index (SSCI)	
E-Learning	Prompt	301	39	340
	Intervention	2.532	716	3248
	Hint	70	69	139
	Nudge	5	17	22
	Feedback	2.809	242	3051
	Scaffold	143	51	194
	Assistance	222	109	331
	Guidance	505	143	648
Digital Learning	Prompt	60	94	154
	Intervention	355	660	1015
	Hint	22	17	39
	Nudge	2	3	5
	Feedback	567	491	1058
	Scaffold	83	15	98
	Assistance	65	19	84
	Guidance	169	24	193
Personalized Learning	Prompt	17	48	65
	Intervention	104	79	183
	Hint	6	9	15
	Nudge	2	4	6
	Feedback	198	130	328
	Scaffold	13	2	15
	Assistance	25	19	44
	Guidance	59	4	63
Adaptive Learning	Prompt	45	96	141
	Intervention	282	704	986
	Hint	29	26	55
	Nudge	0	7	7
	Feedback	209	137	346
	Scaffold	22	3	25
	Assistance	17	13	30
	Guidance	51	10	61
∑		**8.989**	**4000**	**12989**

Literature Search Overview per Database and Keywords.

moderator analyses.[1] The meta-analysis aimed to compare a prompt condition with a non-prompt control condition. When studies included multiple experimental groups versus a single control group, leading to data dependency, we used the "aggregate" function in R to calculate study-wide mean effect sizes (Viechtbauer, 2010). This function uses a weighted averaging method to consider the precision of each individual effect size by weighting each effect size by the inverse of its sampling variance. Furthermore, two effect sizes were extracted for studies with independent samples conducting two experiments (Kraiger et al., 2020; Sitzmann et al., 2009).

We constructed individual meta-analytic models to examine the impact of different prompt features and conducted moderator analyses using a meta-regression approach based on mixed effects models (Higgins et al., 2019; Viechtbauer, 2007). The mixed-effects model was used to fit a set of outcomes, and control moderators were specified based on a correlation matrix of all potential moderators (Borenstein et al., 2021). The model allowed for the simultaneous consideration of both within- and between-study variation, and the inclusion of moderators that affect the mean effect size across studies. By using this method, the impact of each prompt feature on the effect size could be examined while accounting for potential confounding variables. This approach is recommended in meta-analysis when multiple moderators are present and their effects need to be evaluated simultaneously (Higgins et al., 2019).

Where applicable, the individual effect sizes of multiple treatment studies are used in separate moderator analyses. For instance, we performed three separate meta-regressions in studies like Berthold et al. (2007), where different groups received cognitive, meta-cognitive, and a combination of both types. This approach allows us to assess the specific effects of each prompt feature while controlling for potential confounding variables. It also provides a more nuanced understanding of the factors that contribute to the overall effect size.

In cases where studies reported multiple learning outcomes (e.g., Müller & Seufert, 2018), we conducted additional moderator analyses to distinguish between "declarative knowledge" and "procedural knowledge", guided by our codebook categorization. This method offered a clearer understanding of the effects on different types of learning outcomes.

In this study, the R software packages metafor and dmetar (Viechtbauer, 2010) were employed to compute the weighted mean

[1] Cohen's d was used as a measure of effect size. Hattie's (2008) definition of effect size was adopted, with an effect size of .2 considered as small, .4 as medium, and .6 as large. The preferred method to calculate Cohen's d was to use the means and standard deviations of posttest scores reported in the study, but if only other statistics, such as F or t values were reported, those were used instead.

Table 4
Codebook for study and prompt categorization.

Study Characteristics and Statistics		
Study ID, Title, Reference, Journal Name Year, Country of Origin, Region Sample Size (Experimental & Control Group) and Mean Age Statistical Data (Means, Standard Deviations, Standard Error, Confidence Intervalls etc.) Number and Content of Prompts		
Learner Target Group		
Elementary School High School Higher Education Working Adults		
Learning Domain		
Social Science Life Science Physical Science Technological Science		
Digital Learning Environment		
Category	Description	Example
Traditional learning environment	Studies that employ comparatively simple digital learning environments with a feedback function.	Hypermedia learning environment (Study ID 63) E-Learning Journal Learning Management Systems Learning Activity Management System environment Multimedia-supported e-learning environment Web-based environment Computer Architecture Laboratory Digital Learning Environment Adaptive e-learning environment Online database programming environment Computerized dynamic assessment system
Intelligent Tutoring System (ITS)	Learners can choose between pedagogical agents who then support them with computer-based learning environment.	“MetaTutor” (ID 64) Intelligent Software Tutor Web-based interactive tutoring systems Adaptive tutoring system Intelligent tutorial system
Immersive Learning Environment	Includes educational games, simulation-based learning environments, VR/AR learning environment	Study ID 81 Study ID 57
Category “What” (Cognitive Level of Prompts)		
Cognitive Prompt	These prompts are designed to support learners’ information processing.	“Higher values in verbal comprehension tend to go with higher values in verbal fluency, and lower values in verbal comprehension tend to go with lower values in verbal fluency” (Study ID 11)
Meta-Cognitive Prompt	A variety of different terms are used in the literature for meta-cognitive prompts. To guarantee an unambiguous classification, the individual prompts of each study must be analyzed. These prompts aim to activate the monitoring and control of learners’ cognitive activities. This includes planning, goal setting, and evaluation of learning processes and outcomes.	“Before you start working with the information presented within this program you should prepare first: Get an overview over the learning material.” (Study ID 55); Self-Explanation Prompt (Study ID 57); Elaboration Prompt (Study ID 87)
Non-Cognitive Prompt	These prompts seek to increase learner motivation by praising the student or providing guidance for how to regulate their own motivation.	“The e-module was designed for all kinds of learners with different skills and backgrounds. You can improve your performance by remembering your goals for taking this e-module. Focus on these goals for taking this e-module and how you have successfully learned in the past. Please feel assured that with appropriate effort, all participants should be able to complete the e-module successfully. Your hard work will pay off!” (Study ID 70)
Mixed	This category includes possible combinations of the three cognitive types of prompts.	Cognitive, Meta-Cognitive and Non-Cognitive Cognitive and Meta-Cognitive
Form of Prompt Presentation		
Text-Only Prompts (without selections)	Includes all prompts in text-only form with no selection options for the learner’s response.	–
Audiovisual Prompts (without & with selections)	Includes all audiovisual prompts with and without selection options for the learner’s response.	–
Category “How”		

(*continued on next page*)

Table 4 (*continued*)

Generic Prompt	Includes prompts that are worded in general terms and can be used for a variety of digital learning content	Procedural Prompt Non-personalized speech Context-independent feedback
Directed Prompt	Includes prompts that are content-specific, contextualized and use personalized speech	Specific Prompt Directive Prompt Personalized speech Contextualized feedback Context-depended feedback
Generic & Directed Prompts	Includes prompts that are generic and directed (combination)	–
Category "Who"		
All Learners	Includes all prompts that are available for all learners and have the same content.	"Please justify your decision." (Study ID 5)
Group of Learners	Includes all prompts that are only available for a specific group of learners.	Based on previous test/questionnaire: Prior Knowledge (Study ID 57) Learner strategy & Meta-knowledge (Study ID 68)
Category "When"		
Time-based	Includes prompts that appear after a certain amount of time.	During the learning sequence (Study ID 2) Previous to the learning sequence (Study ID 87) After the learning sequence (Study ID 68)
Action-based	Includes prompts that are based on the individual behavior of the learner.	Based on previous activity: Self-selected - Clicking Button (Study ID 65) Incorrect/Correct Solution (Study ID 18) Navigation decision (Study ID 56)
Irrelevant		
Examples of studies that are excluded and therefore marked as "irrelevant": All studies that only use a learning environment without prompts. (Study 39) All studies that do not use a digital setting. (Study 4) All studies that only performed a log-file data analysis and did not use prompts. (Study 21) All studies that have only conducted a qualitative study. (Study 22) All studies that do not examine the effect of prompts on learning achievement. (Study 1)		

effect size (Cohen's d), perform moderator analyses, and analyze potential publication bias. The studies included in the meta-analysis were not completely identical due to differences in contextual characteristics, prompt features, and measurement instruments. Therefore, a random effects model using the restricted maximum likelihood method was used. Finally, a forest plot was created to depict the effect sizes and 95 % confidence interval from each study.

Given that the meta-analysis only included peer-reviewed journal articles, publication bias might have inflated the estimate of effect sizes. To address this issue, we employed several methods: funnel plots for visual bias detection, Egger's regression test for funnel plot asymmetry (Egger et al., 1997), and a likelihood ratio test using the Vevea and Hedges (1995) weight-function model for a more precise bias measurement. The results of all tests showed signs of possible publication bias, and therefore, the trim-and-fill method was applied to adjust the effect size estimates (Duval & Tweedie, 2000). Additionally, we performed a sensitivity analysis using the "influence" function (Viechtbauer, 2010) to identify any studies disproportionately affecting the overall effect size (Borenstein et al., 2021).

4. Results

4.1. Overall effect of prompts

The overall effect of prompts was estimated based on 68 peer-reviewed journal articles (70 effect sizes) published between 1999 and 2022. The average sample size of the included effect sizes was $N = 98$ (range 26–656) with an average age of 25 years. The analyses showed, using a random effects model, that educational prompts had positive, medium effects on learning achievement. The average weighted effect size was Cohen's $d = .394$ ($SE = .040$) with a 95 % confidence interval ranging from .31 to .48, favouring learners who were prompted ($p < .001$) compared to the control group without prompting. Fig. 3 presents the cumulative forest plot depicting each study's overall effect size and the effect sizes of all studies.

In addition, the heterogeneity tests for learning achievement were significant, with a moderate amount of variance in the effect sizes ($Q = 179.22$, $df = 69$, $I^2 = 62.67$). The studies included differing mainly in their learning environment, the learning content, and the type of prompts. Thus, the hypothesis of homogeneity was rejected, suggesting the necessity of moderator analyses to ascertain variables that might explain the variability.

4.2. Publication bias and sensitivity analysis

Examination of the funnel plot (Fig. 4) showed that the effect sizes were not distributed symmetrically around the mean effect size. Likewise, Egger's regression and the likelihood ratio test results suggested that the observed effect sizes for learning achievement

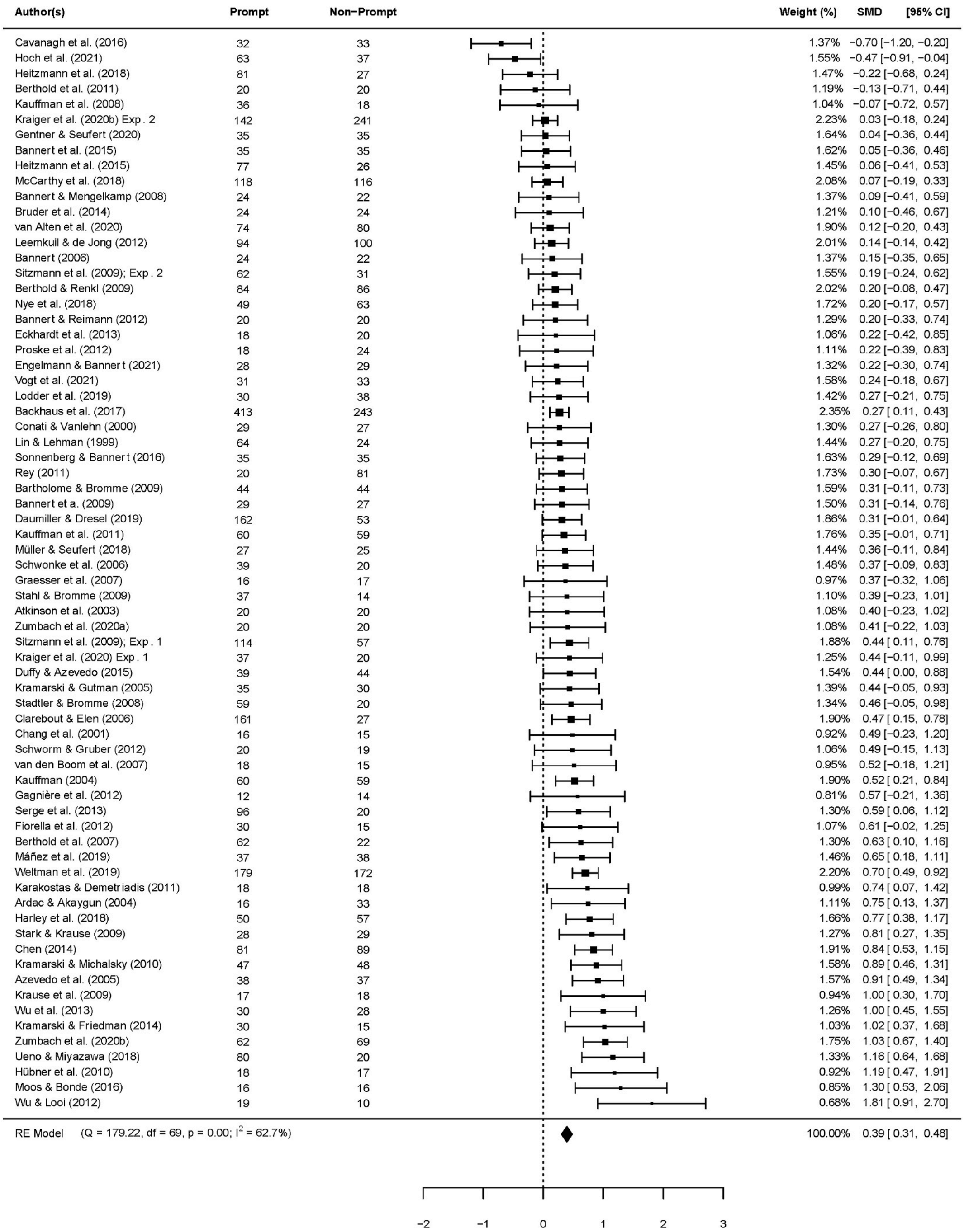


Fig. 3. Forest plot of the overall effect on learning achievement (Ardac & Akaygun, 2004; Atkinson, Renkl, & Merrill, 2003; Bannert, 2006; Bannert & Mengelkamp, 2008; Bannert & Reimann, 2012; Bruder, Blessing, & Wandke, 2014; Chang, Sung, & Chen, 2001; Clarebout & Elen, 2006; Conati & Vanlehn, 2000; Duffy & Azevedo, 2015; Eckhardt, Urhahne, Conrad, & Harms, 2013; Engelmann & Bannert, 2019; Fiorella, Vogel-Walcutt, & Fiore, 2012; Gagnière, Betrancourt, & Détienne, 2012; Graesser et al., 2007; Harley, Taub, Azevedo, & Bouchet, 2018; Heitzmann, Fischer, & Fischer, 2018; Heitzmann, Fischer, Kühne-Eversmann, & Fischer, 2015; Hoch, Scheiter, Stalbovs, & Gerjets, 2021; Hübner, Nückles, & Renkl, 2010; Kauffman, Ge, Xie, & Chen, 2008; Kauffman, Zhao, & Yang, 2011; Kramarski & Friedman, 2014; Kramarski & Gutman, 2005; Kramarski & Michalsky, 2010; Krause, Stark, & Mandl, 2009; Lodder, Heeren, & Jeuring, 2019; Máñez Sáez, Vidal-Abarca Gámez, & Martínez Giménez, 2019; McCarthy, Likens, Johnson, Guerrero, & McNamara, 2018; Niegemann & Weinberger, 2020; Nückles, Hübner, & Renkl, 2009; Nye et al., 2018; Polanin & Pigott, 2015; Proske, Narciss, & McNamara, 2012; Rey, 2011; Schmidt & Hunter, 2014; Schworm & Gruber, 2012; Stadtler & Bromme, 2008; Stahl & Bromme, 2009; Stark & Krause, 2009; Ueno & Miyazawa, 2018; van den Boom, Paas, & van Merriënboer, 2007; van Schoors, Elen, Raes, & Depaepe, 2021; Vogt, Babel, Hock, Baumann, & Seufert, 2021; Weltman, Timchenko, Sofios, Ayres, & Marcus, 2019; L. Wu & Looi, 2012; P. H. Wu, Hwang, & Tsai, 2013; Zumbach, Ortler, Deibl, & Moser, 2020; Zumbach, Rammerstorfer, & Deibl, 2020).

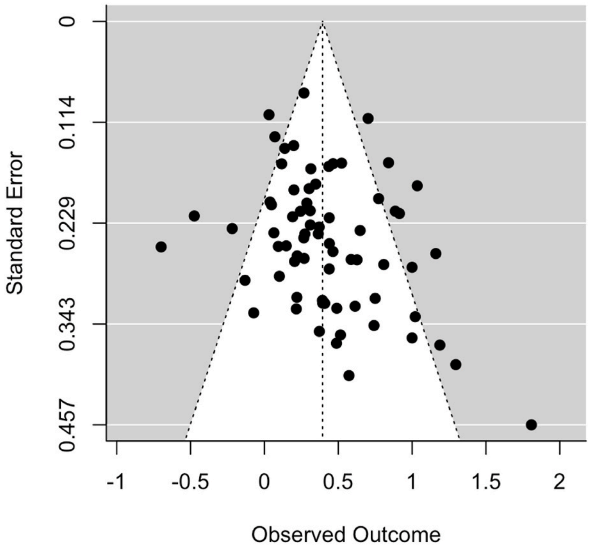


Fig. 4. Funnel plot of the overall effect on learning achievement.

might be skewed by publication bias..[2]

To assess possible publication bias, we used the trim-and-fill method (Duval & Tweedie, 2000) and found that 21 studies were missing on the left side of the distribution, indicating potential asymmetry. Next, the overall effect size was computed again by considering the 21 additional effect sizes, and an adjusted effect size of $d = .220$ ($p < .050$) was found. Thus, the observed direction for the overall effect in this meta-analysis is robust because the adjusted effect size remains statistically significant even after applying the trim-and-fill method to account for publication bias. The funnel plot with the additional studies is presented in Fig. 5. The leave-one-out analysis (Viechtbauer, 2010) was applied to check the extreme positive or negative effect sizes for the overall effect sizes. In this meta-analysis, no studies were determined to be outliers.

4.3. Moderator analysis

4.3.1. Correlation analysis

Given the moderate heterogeneity in the included studies, 12 mixed-effects models were conducted to shed light on possible moderators (Table 5). Correlations between the categories are shown in a correlation matrix (Fig. 6). The highest correlations were observed for the categories "Who" (all learners; group of learners), "What" (cognitive; meta-cognitive, etc.), and "How" (directed; generic; a combination of both types). However, only the categories "Who" and "How" showed statistically significant effects in individual meta-regression models and therefore were used as control variables for our moderator analyses (Table 5, model 4). If the number of studies for a specific category was too small (e.g., two studies) for the categorial moderators, we omitted this category in the moderator analysis (Higgins et al., 2019). This was the case for the categories "individual learners", the combination of action-based and time-based prompts, time-based prompts presented after the learning sequence and prompts based on results of previous tests or questionnaires.

4.3.2. Moderating effects for category "what"

No statistically detectable significance was observed in the mixed effects model for the cognitive types of prompts ($p = .106$). Thus, our hypothesis of similar effects of cognitive, meta-cognitive prompts and non-cognitive prompts can be supported. Furthermore, this category was not used as a control variable in the further course of our analysis. Each cognitive type of prompt was associated with a statistically detectable mean effect size (Table 5; model 1). Cognitive prompts were associated with a large effect size ($d = .592$), whereas the combination of all cognitive types was associated with a medium effect size ($d = .402$). Meta-cognitive prompts, and the combination of meta-cognitive and cognitive prompts were associated with slightly lower moderate effect sizes ($d = .371$; $d = .356$).

Model 5 (Table 5) presents statistics of moderator analyses for learning achievement. There was no statistically significant difference between the different types of learning achievement, namely procedural and declarative knowledge ($p = .326$). Consequently, our hypothesis stating significant differences among the various types of learning achievements is not supported. Procedural knowledge as a measurement of learning achievement was associated with a small and statistically not significant effect size ($d = .225$;

[2] The results of Egger's regression test for funnel plot asymmetry confirmed a potential publication bias, $Z = 2.42$, $p = .02$, $b = .06$ (*CI*: .22, .34). The likelihood ratio test using the weight-function model for the overall effect sizes was significant, indicating that the data basis of this meta-analysis might be skewed by publication bias. (X^2 (1) $= 4.56$, $p = .03$).

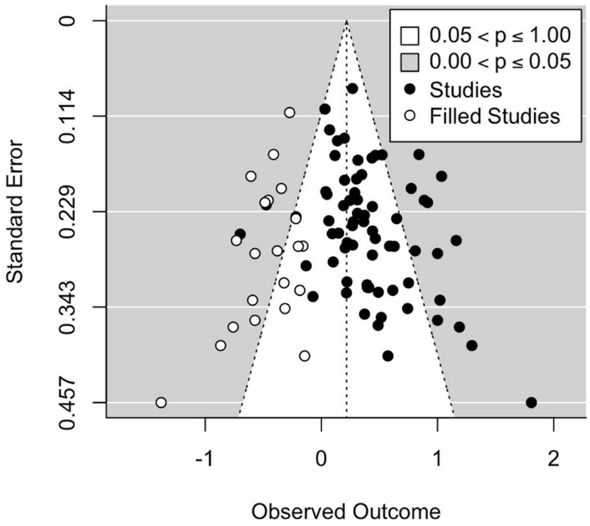


Fig. 5. Funnel plot (trim-and-fill) of the overall effect on learning achievement.

$p = .110$). Besides, declarative knowledge was associated with almost no effect and did also not reach statistical significance ($d = .055$; $p = .717$).

4.3.3. Moderating effects for category "who"

The "Who" of prompts, which refers to the selection of learners who receive a prompt, was analyzed in Model 2 (Table 5) and shows a statistically detectable difference among all categories ($p = .003$). Forty-four studies using prompts for all learners to enhance learning achievement had the smallest effect size ($d = .302$). Studies with group-based prompts had the largest effect size in this category ($d = .563$). This result shows that group-based prompts are more effective in improving learning achievement than prompts for all learners and supports our hypothesis.

Different learning environments do not impact learning achievement, as shown in Model 6 ($p = .080$). Most studies ($N = 52$) were conducted in traditional e-learning environments, which was correlated with a medium effect size ($d = .364$). ITS ($N = 7$) were associated only with a marginal higher effect size ($d = .384$) than traditional e-learning environments. The category immersive learning environments was also associated with a medium effect size ($d = .341$) and includes seven studies with simulation-based learning environments and two studies, each with educational games and VR/AR learning environments. These findings suggest that our hypothesis, which stated that prompts will be more effective in immersive learning environments than in traditional digital learning environments, is not supported.

Three main types of groups were selected, namely high school, higher education, and working adults (Table 5, model 7). The analysis revealed a significant difference between the three sample groups ($p = .017$). The effect size was largest for high school students ($d = .412$) followed by higher education students ($d = .320$). Only a small effect size ($d = .071$) was found for working adults, implying that prompts had almost no effect on learning achievement for them. However, the number of studies for working adults could have been too low ($N = 4$) to reach statistical significance for identified effects ($p = .696$). Overall, our hypothesis that the impact of prompts does not vary among the targeted learner groups was not substantiated.

The learning domain model (Table 5, model 8) has a statistically significant impact on learning success ($p = .029$). Specifically, the largest effect size was observed in the social science domain ($d = .479$), followed by technological science ($d = .432$). In contrast, smaller effect sizes were observed in the life science and physical science domains, with effect sizes of $d = .329$ and $d = .268$, respectively. These findings suggest that our initial hypothesis, which stated that prompts have no varying effects across learning domains, is not supported.

The research region (Table 5, model 9) was coded as North America, Europe, East Asia, and Mixed (e.g., Israel, New Zealand, and the Dominican Republic). The highest significant effect size was associated with studies conducted in East Asia ($d = .862$), while the smallest effect size was associated with Europe ($d = .278$). A moderate effect size was associated with studies from North America ($d = .356$) and studies from mixed regions ($d = .431$). We discovered a significant difference between the four categories ($p = .001$), which contradicts our hypothesis that the effects of prompting are consistent across all regions.

4.3.4. Moderating effects for category "when"

Model 11 (Table 5) presents statistics of moderator analyses for the trigger conditions of prompts and supports our hypothesis. A statistically significant difference was found between action-based and time-based prompts ($p = .008$). Action-based prompts ($d = .447$) are considerably more effective in improving learning achievement than time-based prompts ($d = .240$).

Furthermore, the moderator analysis (Table 5, model 12) for subcategories of action-based and time-based prompts also showed a statistically significant moderating effect ($p = .050$). Regarding time-based prompts, most studies implemented prompts during the

Table 5
Moderated Meta-Regression based on Mixed-Effects Models.

Model 1: What

Category	Type	N	d	SE	p-value
What					.106
	Cognitive (ref.)	16	.592	.080	<.001
	Meta-cognitive	43	.371	.048	<.001
	Cognitive & Meta-cognitive	11	.356	.097	<.001
	Combination	4	.402	.134	.003

Model 2: Who

Category	Type	N	d	SE	p-value
Who (control variable)					.003
	All learners (ref.)	44	.302	.053	<.001
	Group of leaners	24	.563	.069	<.001

Model 3: How

Category	Type	N	d	SE	p-value
How (control variable)					.047
	Directed (ref.)	30	.473	.068	<.001
	Generic	31	.289	.064	<.001
	Generic & Directed	9	.571	.121	<.001

Model 4: Who & How

Category	Type	N	d	SE	p-value
Overall				.24	
Who (control variable)	All learners (ref.)	48	.439	.080	<.001
	Group of leaners	25	.513	.071	<.001
How (control variable)	Directed (ref.)	30	omitted	omitted	omitted
	Generic	31	−.103	.092	.264
	Generic & Directed	9	.044	.123	.723

Model 5: Learning Achievement

Category	Type	N	d	SE	p-value
Overall					.326
Who (control variable)	All learners (ref.)	44	ommited	omitted	omitted
	Group of leaners	24	.201	.170	.235
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	.103	.160	.521
	Generic & Directed	9	.031	.338	.926
Learning Achievement	Procedural Knowledge (ref.)	16	.225	.140	.11
	Declarative Knowledge	22	.055	.153	.717

Model 6: Learning Environment

Category	Type	N	d	SE	p-value
Overall				.08	
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.201	.105	.056
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.095	.110	.386
	Generic & Directed	9	.059	.140	.674
Learning Environment	E-learning environment (ref.)	52	.364	.098	<.001
	ITS	7	.384	.145	.008
	Immersive learning environment	11	.341	.131	.009

Model 7: Target Group

Category	Type	N	d	SE	p-value
Overall					.017
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.214	.105	.041
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.029	.112	.794
	Generic & Directed	9	.055	.137	.69
Target Group	High School Students	21	.412	.097	<.001
	Higher Education Students	44	.320	.106	.003
	Working Adults	4	.071	.181	.696

Model 8: Learning Domain

(*continued on next page*)

Table 5 (*continued*)

Category	Type	N	d	SE	p-value
Overall					.029
Who (control variable)	All learners (ref.)	44	omitted	ommited	ommited
	Group of leaners	24	.180	.108	.094
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.161	.108	.136
	Generic & Directed	9	.030	.142	.835
Learning Domain	Life Science	13	.329	.117	.005
	Social Science	39	.479	.108	<.001
	Technological Science	5	.432	.180	.016
	Physical Science	13	.268	.116	.021
Model 9: Region					
Category	Type	N	d	SE	p-value
Overall				.001	
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.136	.099	.167
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.041	.098	.675
	Generic & Directed	9	.110	.131	.402
Region	East Asia	5	.862	.183	<.001
	North America	13	.356	.101	<.001
	Europe	31	.278	.092	.003
	Mixed	6	.431	.157	.006
Model 10: Form of Presentation					
Category	Type	N	d	SE	p-value
Overall					.038
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.190	.106	.072
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.095	.103	.359
	Generic & Directed	9	.051	.139	.712
Presentation Form	Text (with response option); Audiovisual	15	.327	.123	.008
	Text (without response option)	55	.377	.092	<.001
Model 11: When					
Category Type	N	d	SE	p-value	
Overall				.008	
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.109	.105	.302
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.060	.097	.537
	Generic & Directed	9	.010	.129	.9371
When	Action-based	48	.447	.094	<.001
	Time-based	25	.240	.096	.013
Model 12: When (Subcategory)					
Category	Type	N	d	SE	p-value
Overall				.05	
Who (control variable)	All learners (ref.)	44	ommited	ommited	ommited
	Group of leaners	24	.108	.139	.437
How (control variable)	Directed (ref.)	30	ommited	ommited	ommited
	Generic	31	−.072	.101	.475
	Generic & Directed	9	.004	.134	.978
Time-based	during the learning sequence	19	.223	.105	.033
	Previous to the learning sequence	6	.333	.162	.04
Action-based	Incorrect/Correct Solution	14	.465	.173	.007
	learner self-selected	6	.407	.172	.018
	previous activity (navigation decision)	28	.458	.100	<.001

learning sequence ($N = 19$), which was associated with a small effect size ($d = .233, p = .033$). Six studies implemented prompts before the learning sequence, which was associated with a slightly larger effect size ($d = .333$, $p = .040$). Only one study implemented prompts after the learning sequence and thus was not included in the moderator analysis and should not be interpreted. It was observed that the three types of action-based prompts (incorrect/correct solution; previous activity; learner self-selected) showed significant moderate effects ($d = .465$; $d = .458$; $d = .407$) in improving learning achievement.

4.3.5. *Moderating effects for category "how"*

Regarding the specificity of prompts (Table 5, model 3), there was a significant difference among the three categories ($p = .047$).

	What	How	Who	When	Presentation	Region	Target Group	Domain
How	**.548**							
Who	**.355**	**.395**						
When	.223	.172	**.359**					
Presentation	**.351**	.225	**.422**	.222				
Region	.180	.265	.213	.202	.245			
Target Group	.151	.257	.114	.229	.272	**.327**		
Domain	.270	.230	.254	.226	.193	**.379**	**.417**	
Environment	**.307**	.232	.154	.296	.091	**.333**	.111	.237

Fig. 6. Correlation Matrix for potential Moderators.

The effect size for the generic prompts ($d = .289$) was lower than for directed prompts ($d = .473$). However, the highest effect size ($d = .571$) was associated with both generic and directed prompts, which was implemented in nine studies. Thus, our hypothesis of no differences in the effectiveness of directed or generic prompts cannot be supported.

The effectiveness of prompts on learning achievement was found to be moderated by the presentation form, as shown by the statistically significant results of the moderator analysis in model 10 ($p = .038$). Most studies ($N = 55$) used text-only prompts without response options, which was correlated with the strongest significant moderate effect size ($d = .377$). Studies that used both text-only forms (with and without response options) and enriched their prompts with visual media formats such as audio, video, and pictures had a slightly smaller effect size ($d = .327$). Therefore, we reject our hypothesis that there are no differences in the effectiveness between media-enriched and text-only prompts.

5. Discussion

5.1. Overall effect

The results of this meta-analysis, which focused on peer-reviewed publications and rigorous WWC coding, show that learning prompts have moderate overall positive effects on learning achievement across a broad range of different domains and educational contexts ($d = .394$). These results are consistent with other meta-analyses that also showed moderate to large effects of prompts on learning achievement for specific learning domains such as medical and teacher education (Chernikova et al., 2019), STEM education (Belland et al., 2017), or cognitive types (Guo, 2022). However, this study's novel contribution is the discovery of a significant publication bias within the research field, leading to a more conservative effect size estimate ($d = .220$). This adjusted estimate suggests a significant but smaller overall positive effect than previously reported. It highlights the tendency in educational technology research to prioritize publishing positive findings while overlooking studies with limited or negative outcomes. As the first to rigorously quantify this bias in prompt research, we provide a more accurate effectiveness assessment, with the adjusted effect still representing meaningful educational impact comparable to established interventions.

While prompts in digital learning environments can significantly contribute to students' learning, caution is needed against over-euphoric expectations. The adjusted effect size estimation and many identified moderating variables in this meta-study should warn researchers and practitioners that only implementing prompts is no silver bullet for enhancing learning achievement. This study's main contribution lies in exploring moderator variables to gain insights into how prompts and study features lead to positive effects.

Methodologically, our rigorous application of WWC standards strengthens credibility by excluding weak studies and applying strict inclusion criteria. Combined with our comprehensive multi-database search strategy, this rigor provides a more complete picture than previous meta-analyses limited to specific domains or environments.

5.2. Category "what"

The lack of significance in the mixed effects model ($p = .106$) supports our hypothesis that prompts have consistent effects across different cognitive types. We hypothesize that this similarity occurs because all prompt types share a fundamental mechanism: they direct learners' attention to important aspects of the learning task. Whether fostering cognitive processing, meta-cognitive awareness, or motivation, each prompt type ultimately enhances attention allocation and information processing. While cognitive and meta-cognitive prompts have been extensively studied, non-cognitive (motivational) prompts have received limited attention, with only four studies investigating them combined with other prompt types. Moreover, task motivation likely affects prompt compliance and effectiveness, especially for learners interested in the task and eager to enhance learning achievement (Jansen et al., 2019). Exploring the influence of task motivation on prompt interaction, perception, and learning achievement holds promise for future research.

Regarding knowledge type acquisition, contrary to some previous findings (e.g., Berthold et al., 2011), we found no statistically significant evidence that prompts enhanced declarative knowledge acquisition ($d = .055$). For procedural knowledge, while the effect size was larger ($d = .225$), it still is considered small. Notably, our study addresses a significant gap in the literature, as previous meta-analyses on prompts have not examined their differential effects on declarative versus procedural knowledge acquisition. Thus, further exploration of prompt design and implementation is needed to optimize learning support for different knowledge types.

5.3. Categories “who” and “when”

Meta-analytic testing confirms the significant effectiveness of personalized prompts, particularly group-based ($d = .513$) and action-based ($d = .447$), compared to non-personalized prompts (models 4 and 11). These findings extend recent meta-analytical work by Guo (2022) and Zheng et al. (2022), who advocated for personalized prompts in digital learning environments. Our interpretation of this finding is that personalized prompts empower learners to navigate the complexities of a digital learning environment while receiving appropriate support when they encounter difficulties. In line with Mead et al. (2019), we argue that personalized prompts serve as a safety net, enabling the design of a learning experience where students take an active role in working through the task scenarios while correcting their errors as needed. In contrast, low prompt personalization may potentially lead to poor compliance by the learner and hence, missing benefits on learning achievement (Bannert & Mengelkamp, 2013; Lallé, Conati, Azevedo, Mudrick, & Taub, 2017; van Alten et al., 2020). A key advantage of personalized prompts, particularly action-based prompts, is their timely and targeted delivery, ensuring that only learners who genuinely need support receive them. This reduces the likelihood of learners perceiving prompts as unnecessary or disruptive (van Alten et al., 2020) and, in turn, increases compliance. In this respect, more research is needed to investigate the relationships between presentation time, the number of prompts, and perceived cognitive load based on previous knowledge (Sonnenberg & Bannert, 2016; Wang & Lajoie, 2023).

Our analysis revealed significant variations in prompt effectiveness across different learner groups ($p = .017$), with effectiveness decreasing from high school students ($d = .412$) to higher education students ($d = .320$) to working adults ($d = .071$). This finding contradicts both our hypothesis and previous meta-analytic work by Zheng et al. (2022, 2016), which indicated no significant differences in effect sizes across different sample groups. Most studies primarily focus on high school and higher education settings (94 %). Few have been conducted in other contexts like workplace learning ($N = 4$). This emphasis is due to the rapid growth of technology adoption and remote learning in higher education compared to other settings (Martin et al., 2020). The average participant age of 25 years indicates a focus on young learners, who appear to benefit more from prompting while developing metacognitive skills. Concerningly, two studies (Cavanagh et al., 2016; Kraiger et al., 2020) reported negative effects of prompts on learning achievement among older adults and workers (55–70 years), who may be less familiar with digital prompts. Considering these findings, caution should be exercised when implementing prompts in digital learning environments for older adults (Kraiger et al., 2020). Additional training on prompt usage should be provided for older adults. Further research is needed to determine if the positive effects of prompts generalize to older learners.

A notable finding in our meta-analysis was the significant regional variation in prompt effectiveness ($p = .001$). Studies conducted in East Asian contexts demonstrated substantially larger effects ($d = .862$) compared to those from European settings ($d = .278$). This finding contradicts Cai et al.'s (2022) meta-analytic conclusions, which reported no statistically significant moderating effect of sample region on prompt effectiveness. While the differences in scope between Cai et al.'s (2022) focus on game-based learning and our broader examination of digital learning environments may partially explain this discrepancy, the extent of regional differences deserves further consideration. These regional variations suggest that sociocultural factors may influence prompt effectiveness in ways not previously recognized in the literature. The higher learning achievements in East Asia, known for its technology affinity, could be due to learners' better handling of prompts, potentially leading to superior learning outcomes. Additionally, greater familiarity with technology-enhanced learning in these regions may lead to more productive engagement with prompts. This interpretation challenges the assumption of cultural universality in educational technology effects. However, these results are based on just five studies from East Asia. Additional research is necessary to further explore the relationship between the larger effect size estimates in the technical science domain and the East Asian region.

Our hypothesis that the effect of prompts would not vary across different learning domains is rejected by the statistically significant impact of the learning domain on learning achievement ($p = .029$). Social science and technical science domains demonstrated the most substantial effect sizes ($d = .479$ and $d = .432$), contradicting prior meta-analytic findings by Zheng et al. (2016, 2022) and Cai et al. (2022) that reported no significant differences among learning domains. These findings offer a tentative suggestion that learners in these domains might be more technologically adept and thus they navigate prompts more effectively to enhance learning achievement. Additionally, these domains might naturally incorporate more structured problem-solving approaches that align well with prompting. The distinct pedagogical practices within each discipline could also mediate how prompts influence learning processes. However, it is important to underline that these are merely suppositions at this stage and lack empirical support within current research. The inconsistency with previous meta-analyses suggests potential moderating variables at work that have not been fully identified. Consequently, it remains essential to conduct additional research to either confirm or dismiss this proposition.

Prompts were not more effective in more immersive learning environments ($d = .341$), such as educational games, learning simulations, or virtual reality, compared to traditional learning environments ($d = .364$) or ITS ($d = .384$). No significant differences were found among these digital learning environments ($p = .080$). This finding contradicts theoretical predictions that more immersive environments should enhance prompt effectiveness through greater engagement and self-regulated learning opportunities (see Zimmerman and Moylan, 2009; Moos & Bonde, 2016). This equivalence suggests prompt effectiveness may rely more on design features than on the complexity of the environment. Additionally, the advantages of immersive environments might be counterbalanced by a higher cognitive load. However, the limited number of studies in immersive environments (N = 11) constrains our analysis and necessitates further research to examine differences in their impact across varying immersive settings.

5.4. Category "how"

Our analysis revealed a statistically significant effect of prompt specificity, with the combination of generic and directed prompts showing the strongest impact (d = .571). This finding extends previous meta-analytical work by Zheng et al. (2016) by providing empirical evidence for the complementary nature of these approaches. The effectiveness of this combined approach can be explained through the distinct functions each prompt type serves. Directed prompts help learners recall forgotten information and improve performance, while generic prompts make learners aware of underperformance without providing specific guidance (Serge et al., 2013). Renkl et al. (2015) suggest that directed prompts can repair specific deficits, while generic prompts address broader gaps. When implemented together, these prompts appear to create synergistic effects by simultaneously addressing specific knowledge gaps and fostering broader self-regulatory skills. Contrary to our hypothesis and to Belland et al. (2017), who found no difference between individual prompt types but did not examine their combination, our findings suggest that combining generic and directed prompts yields significantly stronger effects than either type alone. This synergistic effect accommodates diverse learner needs, as skilled learners benefit from the high degree of autonomy of generic prompts, while those with limited prior knowledge may need the structured guidance of directed prompts (Davis, 2003; Ifenthaler, 2012).

Contrary to initial assumptions, there is a significant difference in the effectiveness of media-enriched and text-only learning prompts. Although previous studies demonstrated benefits of media-enriched prompts (Berthold & Renkl, 2009; Brünken et al., 2005; Peeck, 1993), our findings reveal that these prompts did not have a substantially greater impact on learning compared to text-only prompts (d = .377). Surprisingly, they showed a slightly reduced effect (d = .327). Nevertheless, both media-enriched and text-only prompts demonstrated moderate effectiveness overall. Cognitive load theory (Sweller, 1988) offers an explanation for this unexpected result. Multimedia elements may introduce extraneous processing requirements that compete with the primary learning task. Simple text prompts, by contrast, may integrate more seamlessly into the learning process without creating additional processing demands. While Mayer's "cognitive theory of multimedia learning" (2021) offers principles to enhance the efficacy of multimedia support, our analysis revealed that none of the included studies explicitly integrated these principles into the design of their multimedia prompts. In the studies included, excessive support through elaborate prompts may have resulted in "over prompting" (Rosenshine et al., 1996). This aligns with Moreno and Mayer's (2007) recommendation that prompts should contain as much media stimuli as necessary but as little as possible not to overwhelm the learner. The central principle emerging from our analysis is that prompts should support core learning tasks rather than adding extra tasks. However, more studies are needed to validate these observations, particularly regarding the specific effects of different media types (audio, visual, or combined).

5.5. Limitations and research outlook

This meta-analysis only identified the following moderating categories: prompt personalization features, learning achievements, learning environments and demographics of those studies included. Future research should explore additional potential moderators, like learner characteristics (e.g., prior knowledge, motivation, digital skills), to fully understand how to optimize prompts in digital learning. This analysis identified only three studies (see Schwonke et al., 2006) in which learners (after randomized allocation) received prompts based on previous knowledge tests and questionnaires.

Recent studies have integrated action-based prompts in digital learning environments using log and performance data (Deutscher et al., 2022; Li et al., 2023; Munshi et al., 2023). These analytics-driven approaches remain uncommon, though researchers like Lim et al. (2024) have made progress by developing analytics-based personalized prompts facilitated by a rule-based artificial intelligence (AI) system. Despite these advancements, a significant research gap persists. To our knowledge, AI has only supported the incorporation of prompts into learning environments, while truly AI-generated prompts have yet to be explored. This distinction represents a promising avenue for future research that could further enhance personalized learning experiences.

Moreover, a few moderators could not be analyzed due to missing information in the included studies, possibly explaining some variance in effect sizes. Key missing data included learning task characteristics such as duration and difficulty, as well as detailed descriptions of prompt designs beyond mere examples. This lack of comprehensive reporting impedes the possibility of a more detailed prompt categorization. Learners ' compliance with prompts was another potential moderator that could not be accounted for. Learners likely complied with the prompts to varying degrees, which would almost certainly influence the effects of the prompts on learning achievement. As the results of previous studies suggested (Davis et al., 2016; Jansen et al., 2020; Wong et al., 2021), future research should take this into account and report more rigorously on prompt compliance. To achieve this, researchers could employ think-aloud protocols, log data, and eye-tracking methodologies to assess compliance patterns and inform the development of personalized prompts designed to enhance learner engagement (Lallé, Conati, Azevedo, Mudrick, & Taub, 2017).

Another limitation is that the meta-analysis only focused on learning achievement outcomes, overlooking the potential influence of prompts on emotional and motivational aspects. Yet another promising avenue for future research could involve examining learning performance on immediate post-tests versus delayed post-tests. However, the scarcity of empirical studies on long-term effects, variability in delayed post-test formats, and timing differences make a comparison or analysis of little value at this moment. Another limitation is the variability and underspecification of test formats used across the primary studies. While the most common formats were a combination of multiple-choice and open-ended questions, many studies provided incomplete details on test characteristics, length, response types, and scoring rubrics. Few studies published the full assessment instruments. This heterogeneity and lack of documentation precluded moderator analyses comparing prompt effectiveness across test formats.

6. Conclusions for the practical design of learning prompts

Our findings reveal that prompt effectiveness depends on careful consideration of multiple factors, including personalization, timing, context, and presentation. The complexity and potential interactions of these factors necessitate a nuanced approach to prompt design that incorporates cost-effectiveness considerations across multiple dimensions.

Our findings emphasize that prompt effectiveness hinges on personalization, timing, context, and presentation, necessitating a carefully balanced design that optimizes learning impact and cost-effectiveness. More institutions are investing in digital learning through prompts, but the effectiveness varies greatly and improves significantly with increased personalization. However, more personalized prompts require higher investments. Group- and action-based prompts depend on sophisticated tracking and rule-based systems, whereas simple time-based prompts are easier to implement (Deutscher et al., 2022; Li et al., 2023).

A potential cost-saving strategy lies in text-only prompts ($d = .377$), which slightly outperform media-enriched prompts ($d = .327$), suggesting that expensive multimedia elements may not always be justified. Additionally, learner familiarity plays a crucial role. Suboptimal prompt use (Bannert et al., 2015; Moser et al., 2017) underscores the need for investments in user training to improve long-term cost-effectiveness.

Beyond cost considerations, technical constraints within existing learning environments may limit the feasibility of certain prompt designs, posing practical challenges to their implementation. These technical limitations may lead to pragmatic decisions that do not align with our theoretical findings but represent the best option given the practical restrictions for a prompt design. Indeed, not every prompt designer has access to a learning environment where various and more personalized types of prompts can be implemented.

Nevertheless, based on our empirical analyses, we feel confident enough to formulate five rules of thumb based on the current state of research (which, of course, may change and refine over time) that might be considered during the prompt design process and thereby possibly inspire contemporary research and practice on learning prompts.

Consider the following recommendations if they are technically feasible and align with the contents and goals of your learning environment.

- Implement action-based prompts that are administered based on an individual learners' behavior instead of a predefined timeline.
- Utilize log or performance data to tailor prompts to the expertise level of learners (prior knowledge), reducing intrinsic cognitive load.
- Keep prompts brief and avoid unnecessary stimuli, information, media, or requests for action to decrease extraneous cognitive load.
- Implement generic and directed prompts depending on the learning goals of the target learners and based on their previous knowledge.
- Ensure your target group is familiar with the use of digital learning environments and prompts or provide additional training.

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used Chat-GPT 3.5 in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interest

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.edurev.2025.100686.

Data availability

Data will be made available on request.

References

References marked with an asterisk indicate studies included in the meta-analysis

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*, 522–560.

* Ardac, D., & Akaygun, S. (2004). Effectiveness of multimedia-based instruction that emphasizes molecular representations on students' understanding of chemical change. *Journal of Research in Science Teaching, 41*(4), 317–337. https://doi.org/10.1002/tea.20005.

* Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774–783. https://doi.org/10.1037/0022-0663.95.4.774.

* Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science, 33*(5–6), 381–412. https://doi.org/10.1007/s11251-005-1273-8.

* Backhaus, J., Jeske, D., Poinstingl, H., & Koenig, S. (2017). Assessing efficiency of prompts based on learner characteristics. *Computers, 6*(1), 7. https://doi.org/10.3390/computers6010007.

* Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research, 35*(4), 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388.

* Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior, 25*(4), 829–835. https://doi.org/10.1016/j.chb.2008.07.002.

* Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?. *Metacognition and Learning, 3*(1), 39–58. https://doi.org/10.1007/s11409-007-9009-6.

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In *In International handbook of metacognition and learning technologies* (pp. 171–186). New York, NY: Springer New York.

* Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4.

* Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038.

* Bartholomé, T., & Bromme, R. (2009). Coherence formation when learning from text and pictures: What kind of support for whom?. *Journal of Educational Psychology, 101*(2), 282–293. https://doi.org/10.1037/a0014312.

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309–344.

Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose (s)? *Educational Psychology Review, 33*(4), 1675–1715. https://doi.org/10.3102/0034654316670999

* Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007.

* Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology, 101*(1), 70–87. https://doi.org/10.1037/a0013247.

* Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior, 27*(1), 69–75. https://doi.org/10.1016/j.chb.2010.05.025.

Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing self-explanation: A meta-analysis. *Educational Psychology Review, 30*(3), 703–725. https://doi.org/10.1007/s10648-018-9434-x

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

* Bruder, C., Blessing, L., & Wandke, H. (2014). Adaptive training interfaces for less-experienced, elderly users of electronic devices. *Behaviour & Information Technology, 33*(1), 4–15. https://doi.org/10.1080/0144929X.2013.833649.

Brünken, R., Seufert, T., & Zander, S. (2005). Fostering coherence formation in learning with multiple representations. *Zeitschrift für Padagogische Psychologie, 19*(1/2), 61–75.

Cai, Z., Mao, P., Wang, D., He, J., Chen, X., & Fan, X. (2022). Effects of scaffolding in digital game-based learning on student's achievement: A three-level meta-analysis. *Educational Psychology Review, 34*(2), 537–574. https://doi.org/10.1007/s10648-021-09655-0

* Cavanagh, T. M., Kraiger, K., & Peters, J. (2016). *Cognitive prompts fail to moderate the impact of stereotype threat on older adults' training performance* (Vol. 11).

* Chang, K. E., Sung, Y. T., & Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid: Learning through computer-based concept mapping. *Journal of Computer Assisted Learning, 17*(1), 21–33. https://doi.org/10.1111/j.1365-2729.2001.00156.x.

* Chen, C.-H. (2014). An adaptive scaffolding e-learning system for middle school students' physics learning. *Australasian Journal of Educational Technology, 30*(3). https://doi.org/10.14742/ajet.430.

Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*(2), 63.

Chernikova, O., Heitzmann, N., Fink, M. C., Timothy, V., Seidel, T., & Fischer, F. (2019). Facilitating diagnostic competences in higher education—a meta-analysis in medical and teacher education. *Educational Psychology Review, 32*(1), 157–196. https://doi.org/10.1007/s10648-019-09492-2

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research, 90*(4), 499–541. https://doi.org/10.3102/0034654320933544

* Clarebout, G., & Elen, J. (2006). Open learning environments and the impact of a pedagogical agent. *Journal of Educational Computing Research, 35*(3), 211–226. https://doi.org/10.2190/3UL1-4756-H837-2704.

Clariana, R. B., & Hooper, S. (2012). Adaptive evaluation systems. In N. M. Seel (Ed.), *Encyclopaedia of the sciences of learning* (pp. 104–106). Springer.

Cognition and Technology Group at Vanderbilt. (1997). *The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

* Conati, C., & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*(4), 389–415.

* Daumiller, M., & Dresel, M. (2019). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education, 87*(1), 161–176. https://doi.org/10.1080/00220973.2018.1448744.

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences, 12*(1), 91–142. https://doi.org/10.1207/S15327809JLS1201_4

Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G.-J. (2016). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-regulated learning strategies at scale. In K. Verbert, M. Sharples, & T. Klobučar (Eds.), *Adaptive and adaptable learning* (Vol. 9891, pp. 57–71). Springer International Publishing. https://doi.org/10.1007/978-3-319-45153-4_5.

Deutscher, V., Seifried, J., Rausch, A., Thomann, H., & Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen. In K.-H. Gerholz, P. Schlottmann, P. Slepcevic-Zach, & M. Stock (Eds.), *Digital Literacy in der beruflichen Lehrer: innenbildung: Didaktik, Empirie und Innovation* (pp. 107–121). Bielefeld: wbv Media.

* Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior, 52*, 338–348. https://doi.org/10.1016/j.chb.2015.05.041.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463.

* Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional support for learning with computer simulations?. *Instructional Science, 41*(1), 105–124. https://doi.org/10.1007/s11251-012-9220-y.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

* Engelmann, K., & Bannert, M. (2019). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction, 72*, Article 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002.

* Fiorella, L., Vogel-Walcutt, J. J., & Fiore, S. (2012). Differential impact of two types of metacognitive prompting provided during simulation-based training. *Computers in Human Behavior, 28*(2), 696–702. https://doi.org/10.1016/j.chb.2011.11.017.

* Gagnière, L., Betrancourt, M., & Détienne, F. (2012). When metacognitive prompts help information search in collaborative setting. *European Review of Applied Psychology, 62*(2), 73–81. https://doi.org/10.1016/j.erap.2011.12.005.

* Gentner, N., & Seufert, T. (2020). The double-edged interactions of prompts and self-efficacy. *Metacognition and Learning, 15*(2), 261–289. https://doi.org/10.1007/s11409-020-09227-7.

* Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*(2–3), 89–105. https://doi.org/10.1007/s11409-007-9013-x.

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16*(1), 94–143.

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12650

* Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2018). Let's set up some subgoals: Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. *IEEE Transactions on Learning Technologies, 11*(1), 54–66. https://doi.org/10.1109/TLT.2017.2756629.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge. https://www.routledge.com/Visible-Learning-A-Synthesis-of-Over-800-Meta-Analyses-Relating-to-Achievement/Hattie/p/book/9780415476188.

Hawlitschek, A., & Joeckel, S. (2017). Increasing the effectiveness of digital educational games: The effects of a learning instruction on students' learning, motivation and cognitive load. *Computers in Human Behavior, 72*, 79–86.

* Heitzmann, N., Fischer, F., & Fischer, M. R. (2018). Worked examples with errors: When self-explanation prompts hinder learning of teachers diagnostic competences on problem-based learning. *Instructional Science, 46*(2), 245–271. https://doi.org/10.1007/s11251-017-9432-2.

* Heitzmann, N., Fischer, F., Kühne-Eversmann, L., & Fischer, M. R. (2015). Enhancing diagnostic competence with self-explanation prompts and adaptable feedback. *Medical Education, 49*(10), 993–1003. https://doi.org/10.1111/medu.12778.

Higgins, J. P., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470712184.fmatter

* Hoch, E., Scheiter, K., Stalbovs, K., & Gerjets, P. (2021). The intention was good: How promoting strategy use does not improve multimedia learning for secondary students. *British Journal of Educational Psychology, 91*(4), Article e12417. https://doi.org/10.1111/bjep.12417.

Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). *Technology-enhanced personalised learning: Untangling the evidence*. Stuttgart: Robert Bosch Stiftung GmbH. http://www.studie-personlisiertes-lernen.de/en.

* Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society, 15*(1).

Jansen, R. S., van Leeuwen, A., Janssen, J., Conijn, R., & Kester, L. (2020). Supporting learners' self-regulated learning in massive open online courses. *Computers & Education, 146*, Article 103771. https://doi.org/10.1016/j.compedu.2019.103771

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review, 28*, Article 100292. https://doi.org/10.1016/j.edurev.2019.100292

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19. https://doi.org/10.1007/s10648-010-9150-7

* Karakostas, A., & Demetriadis, S. (2011). Enhancing collaborative learning through dynamic forms of support: The impact of an adaptive domain-specific support strategy: Enhancing collaborative learning. *Journal of Computer Assisted Learning, 27*(3), 243–258. https://doi.org/10.1111/j.1365-2729.2010.00388.x.

* Kauffman, D. F. (2004). Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of Educational Computing Research, 30*(1–2), 139–161. https://doi.org/10.2190/AX2D-Y9VM-V7PX-0TAD.

* Kauffman, D. F., Ge, X., Xie, K., & Chen, C.-H. (2008). Prompting in web-based environments: Supporting self-monitoring and problem solving skills in college students. *Journal of Educational Computing Research, 38*(2), 115–137. https://doi.org/10.2190/EC.38.2.a.

* Kauffman, D. F., Zhao, R., & Yang, Y.-S. (2011). Effects of online note taking formats and self-monitoring prompts on learning from online text: Using technology to enhance self-regulated learning. *Contemporary Educational Psychology, 36*(4), 313–322. https://doi.org/10.1016/j.cedpsych.2011.04.001.

* Kraiger, K., Cavanagh, T. M., & Willis, C. M. G. (2020). Why do cognitive prompts hurt learning in older adults?. *International Journal of Training and Development, 24*(1), 40–56. https://doi.org/10.1111/ijtd.12169.

* Kramarski, B., & Friedman, S. (2014). Solicited versus unsolicited metacognitive prompts for fostering mathematical problem solving using multimedia. *Journal of Educational Computing Research, 50*(3), 285–314. https://doi.org/10.2190/EC.50.3.a.

* Kramarski, B., & Gutman, M. (2005). How can self-regulated learning be supported in mathematical E-learning environments?: Self-regulated learning in mathematical E-learning. *Journal of Computer Assisted Learning, 22*(1), 24–33. https://doi.org/10.1111/j.1365-2729.2006.00157.x.

* Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction, 20*(5), 434–447. https://doi.org/10.1016/j.learninstruc.2009.05.003.

* Krause, U.-M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158–170. https://doi.org/10.1016/j.learninstruc.2008.03.003.

Lallé, S., Conati, C., Azevedo, R., Mudrick, N., Taub, M., Hu, X., Barnes, T., & Hershkovitz, A. (2017). *On the influence on learning of student compliance with prompts fostering self-regulated learning* (pp. 120–127). China: International Educational Data Mining Society.

Langone, J. (1998). The effects of technology-enhanced anchored instruction and situated learning on preservice teachers in a special education methods course: An exploratory study. *Journal of Developmental and Physical Disabilities, 10*, 35–54. https://doi.org/10.1023/A:1022809500853

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718. https://doi.org/10.3102/0034654315627366

* Leemkuil, H., & de Jong, T. (2012). Adaptive advice in learning with a computer-based knowledge management simulation game. *The Academy of Management Learning and Education, 11*(4), 653–665. https://doi.org/10.5465/amle.2010.0141.

Li, T., Fan, Y., Tan, Y., Wang, Y., Singh, S., Li, X., Raković, M., van der Graaf, J., Lim, L., Yang, B., Molenaar, I., Bannert, M., Moore, J., Swiecki, Z., Tsai, Y.-S., Shaffer, D. W., & Gašević, D. (2023). Analytics of self-regulated learning scaffolding: Effects on learning processes. *Frontiers in Psychology, 14*, Article 1206696.

Lim, L., Bannert, M., Van Der Graaf, J., Fan, Y., Rakovic, M., Singh, S., Molenaar, I., & Gašević, D. (2024). How do students learn with real-time personalized scaffolds? *British Journal of Educational Technology, 55*(4), 1309–1327. https://doi.org/10.1111/bjet.13414

* Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*(7), 837–858. https://doi.org/10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U.

* Lodder, J., Heeren, B., & Jeuring, J. (2019). A comparison of elaborated and restricted feedback in LogEx, a tool for teaching rewriting logical formulae. *Journal of Computer Assisted Learning, 35*(5), 620–632. https://doi.org/10.1111/jcal.12365.
Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low-and middle-income countries: A meta-analysis. *British Journal of Educational Technology, 52*(5), 1935–1964.
* Máñez Sáez, I., Vidal-Abarca Gámez, E., & Martínez Giménez, T. (2019). Does computer-based elaborated feedback influence the students' question-answering process?. *Electronic Journal of Research in Educational Psychology, 17*(47), 81. https://doi.org/10.25115/ejrep.v17i47.2156.
Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research & Development, 68*(4), 1903–1929. https://doi.org/10.1007/s11423-020-09793-2
Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition, 10*(2), 229–240.
* McCarthy, K. S., Likens, A. D., Johnson, A. M., Guerrero, T. A., & McNamara, D. S. (2018). Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system. *International Journal of Artificial Intelligence in Education, 28*(3), 420–438. https://doi.org/10.1007/s40593-018-0164-5.
Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education, 67*(2), 131–142. https://doi.org/10.1080/10899995.2019.1565285
Mislevy, R. J., & Riconscente, M. M. (2005). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing, & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Erlbaum.
* Moos, D. C., & Bonde, C. (2016). Flipping the classroom: Embedding self-regulated learning prompts in videos. *Technology, Knowledge and Learning, 21*(2), 225–242. https://doi.org/10.1007/s10758-015-9269-1.
Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review, 19*, 309–326.
Moser, S., Zumbach, J., & Deibl, I. (2017). The effect of metacognitive training and prompting on learning success in simulation-based physics learning. *Science Education, 101*(6), 944–967.
* Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction, 58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011.
Munshi, A., Biswas, G., Baker, R., Ocumpaugh, J., Hutt, S., & Paquette, L. (2023). Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. *Journal of Computer Assisted Learning, 39*(2), 351–368.
Niegemann, H., & Weinberger, A. (2020). *Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen*. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-54368-9
Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259–271. https://doi.org/10.1016/j.learninstruc.2008.05.002
* Nye, B. D., Pavlik, P. I., Windsor, A., Olney, A. M., Hajeer, M., & Hu, X. (2018). SKOPE-IT (Shareable Knowledge Objects as Portable Intelligent Tutors): Overlaying natural language tutoring on an adaptive learning system for mathematics. *International Journal of STEM Education, 5*(1), 12. https://doi.org/10.1186/s40594-018-0109-4.
Oliver, R., & Herrington, J. (2000). Using situated learning as a design strategy for Web-based learning. *Instructional and cognitive impacts of web-based education*. IGI Global.
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*, n71. https://doi.org/10.1136/bmj.n71
Peeck, J. (1993). Increasing picture effects in learning from illustrated texts. *Learning and Instruction, 3*, 227–238.
Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education, 52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943
Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods, 6*(1), 63–73.
* Proske, A., Narciss, S., & McNamara, D. S. (2012). Computer-based scaffolding to facilitate students' development of expertise in academic writing: Writing Development by computer-based scaffolding. *Journal of Research in Reading, 35*(2), 136–152. https://doi.org/10.1111/j.1467-9817.2010.01450.x.
Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1–12.
Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The journal of the learning sciences*. Psychology Press.
Reiser, B. J., & Tabak, I. (2014). Scaffolding. In *The cambridge handbook of the learning sciences* (2nd ed., pp. 44–62). Cambridge University Press.
Renkl, A., & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. *Educational Psychology Review, 29*(3), 599–621. https://doi.org/10.1007/s10648-015-9340-4
Renkl, A., Skuballa, I. T., Schwonke, R., Harr, N., & Leber, J. (2015). The effects of rapid assessments and adaptive restudy prompts in multimedia learning. *Journal of Educational Technology & Society, 18*(4), 185–198.
* Rey, G. D. (2011). Time advice and learning questions in computer simulations. *Australasian Journal of Educational Technology, 27*(3). https://doi.org/10.14742/ajet.951.
Robertson, K., Rosasco, C., Feuz, K., Schmitter-Edgecombe, M., & Cook, D. (2015). Prompting technologies: A comparison of time-based and context-aware transition-based prompting. *Technology and Health Care, 23*(6), 745–756. https://doi.org/10.3233/THC-151033
Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*(2), 181–221.
Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications.
Schumacher, C., & Ifenthaler, D. (2021). Investigating prompts for supporting students' self-regulation–A remaining challenge for learning analytics approaches? *The Internet and higher education, 49*, Article 100791.
* Schwonke, R., Hauser, S., Nückles, M., & Renkl, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior, 22*(1), 77–92. https://doi.org/10.1016/j.chb.2005.01.002.
* Schworm, S., & Gruber, H. (2012). e-Learning in universities: Supporting help-seeking processes by instructional prompts: Supporting help-seeking in e-learning environments. *British Journal of Educational Technology, 43*(2), 272–281. https://doi.org/10.1111/j.1467-8535.2011.01176.x.
* Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior, 29*(3), 1150–1158. https://doi.org/10.1016/j.chb.2012.10.007.
* Sitzmann, T., Bell, B. S., Kraiger, K., & Kanar, A. M. (2009). A multilevel analysis of the effect of prompting self-regulation in technology-delivered instruction. *Personnel Psychology, 62*(4), 697–734. https://doi.org/10.1111/j.1744-6570.2009.01155.x.
* Sonnenberg, C., & Bannert, M. (2016). Evaluating the impact of instructional support using data mining and process mining: A micro-level analysis of the effectiveness of metacognitive prompts. *Journal of Educational Data Mining, 8*(2), 51–83.
* Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer tool met.a.ware on the web search of laypersons. *Computers in Human Behavior, 24*(3), 716–737. https://doi.org/10.1016/j.chb.2007.01.023.
* Stahl, E., & Bromme, R. (2009). Not everybody needs help to seek help: Surprising effects of metacognitive instructions to foster help-seeking in an online learning environment. *Computers & Education, 53*(4), 1020–1028. https://doi.org/10.1016/j.compedu.2008.10.004.
* Stark, R., & Krause, U.-M. (2009). Effects of reflection prompts on learning outcomes and learning behaviour in statistics education. *Learning Environments Research, 12*(3), 209–223. https://doi.org/10.1007/s10984-009-9063-x.
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Thillmann, H., Künsting, J., Wirth, J., & Leutner, D. (2009). Is it merely a question of "what" to prompt or also "when" to prompt?: The role of point of presentation time of prompts in self-regulated learning. *Zeitschrift für Padagogische Psychologie, 23*(2), 105–115. https://doi.org/10.1024/1010-0652.23.2.105

* Ueno, M., & Miyazawa, Y. (2018). IRT-based adaptive hints to scaffold learning in programming. *IEEE Transactions on Learning Technologies, 11*(4), 415–428. https://doi.org/10.1109/TLT.2017.2741960.

* van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior, 108*, Article 106318. https://doi.org/10.1016/j.chb.2020.106318.

* van den Boom, G., Paas, F., & van Merriënboer, J. J. G. (2007). Effects of elicited reflections combined with tutor or peer feedback on self-regulated learning and learning outcomes. *Learning and Instruction, 17*(5), 532–548. https://doi.org/10.1016/j.learninstruc.2007.09.003.

van Schoors, R., Elen, J., Raes, A., & Depaepe, F. (2021). An overview of 25 years of research on digital personalised learning in primary and secondary education: A systematic review of conceptual and methodological trends. *British Journal of Educational Technology bjet.* , Article 13148. https://doi.org/10.1111/bjet.13148

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60*(3), 419–435. https://doi.org/10.1007/BF02294384

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine, 26*(1), 37–52.

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the metafor package. *Journal of Statistical Software, 36*(3). https://doi.org/10.18637/jss.v036.i03

* Vogt, A., Babel, F., Hock, P., Baumann, M., & Seufert, T. (2021). Prompting in-depth learning in immersive virtual reality: Impact of an elaboration prompt on developing a mental model. *Computers & Education, 171*, Article 104235. https://doi.org/10.1016/j.compedu.2021.104235.

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review, 35*(3), 69.

* Weltman, H. R., Timchenko, V., Sofios, H. E., Ayres, P., & Marcus, N. (2019). Evaluation of an adaptive tutorial supporting the teaching of mathematics. *European Journal of Engineering Education, 44*(5), 787–804. https://doi.org/10.1080/03043797.2018.1513993.

What Works Clearinghouse. (2022). What works Clearinghouse standards handbook (version 5.0). *National center for education evaluation and regional assistance*. Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/handbooks.

Wirth, J. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Padagogische Psychologie, 23*(2), 91–94. https://doi.org/10.1024/1010-0652.23.2.91

Wong, J., Baars, M., de Koning, B. B., & Paas, F. (2021). Examining the use of prompts to facilitate self-regulated learning in Massive Open Online Courses. *Computers in Human Behavior, 115*, Article 106596. https://doi.org/10.1016/j.chb.2020.106596

* Wu, P. H., Hwang, G. J., & Tsai, W. H. (2013). An expert system-based context-aware ubiquitous learning approach for conducting science learning activities. *Journal of Educational Technology & Society, 16*(4), 217–230.

* Wu, L., & Looi, C. K. (2012). Agent prompts: Scaffolding for productive reflection in an intelligent learning environment. *Journal of Educational Technology & Society, 15*(1), 339–353.

Young, M. F., & Kulikowich, J. M. (1992). *Anchored instruction and anchored assessment: An ecological approach to measuring situated learning*. San Francisco, CA: Annual Meeting of the American Educational Research Association. Retrieved from https://files.eric.ed.gov/fulltext/ED354269.pdf.

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review, 17*(2), 187–202. https://doi.org/10.1007/s12564-016-9426-9

Zheng, L., Long, M., Zhong, L., & Gyasi, J. F. (2022). The effectiveness of technology-facilitated personalized learning on learning achievements and learning perceptions: A meta-analysis. *Education and Information Technologies, 27*(8), 11807–11830.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (pp. 299–315). Routledge.

* Zumbach, J., Ortler, C., Deibl, I., & Moser, S. (2020). Using prompts to scaffold metacognition in case-based problem solving within the domain of attribution theory. *Journal of Problem-Based Learning, 7*(1), 21–31. https://doi.org/10.24313/jpbl.2020.00206.

* Zumbach, J., Rammerstorfer, L., & Deibl, I. (2020). Cognitive and metacognitive support in learning with a serious game about demographic change. *Computers in Human Behavior, 103*, 120–129. https://doi.org/10.1016/j.chb.2019.09.026.

4.2 Paper 2: Die LUCA Office Simulation in der Lehrerinnen-und Lehrerbildung-didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen

Deutscher, V., Seifried, J., Rausch, A., Thomann, H. und Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – Didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen. In *Digital Literacy in der beruflichen Lehrer:innenbildung: Didaktik, Empirie und Innovation* (Vol. 68, pp. 107-121). Bielefeld: wbv.

Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – Didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen

Viola Deutscher, Jürgen Seifried, Andreas Rausch, Herbert Thomann & Anke Braunstein

Abstract

Vor dem Hintergrund einer u. a. durch die Digitalisierung bedingten Verschiebung von Kompetenzanforderungen an Lernende bei gleichzeitig wachsenden digitalen Möglichkeiten an beruflichen Schulen muss nicht nur von einer neuen digitalen Realität beruflicher Lernprozesse, sondern auch von einer neuen Realität beruflicher Lehrprozesse ausgegangen werden. Wie ein digital gestützter Unterricht für die kaufmännische Bildung aussehen kann und welche professionellen Kompetenzen von Lehrkräften hierfür relevant sein könnten, wird am Beispiel der an der Universität Mannheim entwickelten Bürosimulation LUCA[1] erörtert. Der Beitrag geht konzeptionell-induktiv anhand der LUCA-Funktionen eines konkreten Anwendungsbeispiels sowie Modellen digitaler Lehrkompetenzen der Frage nach, welche Unterrichtskompetenzen diesbezüglich bei Lehrkräften erforderlich sind. Im Ergebnis werden spezifische Aspekte digitaler Unterrichtskompetenz identifiziert, die für die Anwendung virtueller Lernsimulationen, wie der LUCA-Bürosimulation, hilfreich sind.

Schlagworte: E-Learning, Lernen mit Simulationen, digitale Unterrichtskompetenz, Lehrerbildung

Against the backdrop of a shift in competence requirements for learners caused by digitisation and new digital possibilities at vocational schools, we must assume not only a new digital reality of vocational learning processes, but also a new reality of vocational teaching processes. Using the example of the LUCA Office Simulation developed at the University of Mannheim, we discuss how digitally supported teaching can be designed for commercial education. In addition, we examine which professional competences of teachers are relevant for the design of such learning environments. For this purpose, we initially present features of the LUCA office simulation. We then give an example of digital instruction in LUCA and conceptually explore

1 Die Entwicklung der Bürosimulation wurde aus Mitteln des Bundesministeriums für Bildung und Forschung (Förderkennzeichen: 21AP008A) im Rahmen der ASCOT+-Initiative gefördert (siehe https://www.ascot-vet.net). Weiterführende Hinweise zu LUCA finden sich auf der Projekt-Website unter https://luca-office.de/.

which teaching competences are necessary for teachers to design digital learning by drawing on models of digital teaching competences. As a result, specific aspects of digital teaching competences are identified that are helpful for the application of virtual learning simulations, such as the LUCA office simulation.

Keywords: E-Learning, Simulation-Based Learning, Digital Teaching Competence, Teacher Education

1 Ausgangslage

Im Zuge der Digitalisierung haben sich sowohl die Arbeitswelt als auch die Arbeits- und Lerngewohnheiten von Jugendlichen und jungen Erwachsenen grundlegend verändert. Im Hinblick auf die digitale Transformation am Arbeitsplatz lässt sich festhalten, dass Routinetätigkeiten immer häufiger durch digital vernetzte Systeme gestützt bzw. ersetzt werden und Kommunikationsprozesse zunehmend digital ablaufen. Große Bedeutung kommt daher der Bewältigung komplexerer Aufgabenstellungen in digitalen Netzwerkstrukturen zu (vgl. Frey & Osborne 2017; Seeber, Weber, Geiser u. a. 2019). Es kann also von einer Verschiebung von Kompetenzanforderungen ausgegangen werden (vgl. hierzu die „Skill-Shift-Debatte“, vgl. Bughin, Hazan, Lund u. a. 2018). Parallel dazu haben sich Lernmöglichkeiten und -gewohnheiten grundlegend verändert und das Angebot digitaler Lehr-Lern-Tools an beruflichen Schulen wächst. Zudem führen die technologischen Möglichkeiten der räumlichen und zeitlichen Entgrenzung von Unterricht zu einer Zunahme von synchronen und asynchronen Fernlernangeboten, die insbesondere durch die Corona-Pandemie angetrieben wurden.

Vor dem skizzierten Hintergrund sind digitale Lernumgebungen und didaktische Ansätze weiterzuentwickeln. Eine für den mathematisch-naturwissenschaftlichen Unterricht vorgelegte Meta-Analyse (vgl. Hillmayr, Ziernwald, Reinhold u. a. 2020) verweist beispielsweise auf positive Effekte digitaler Tools bei variierenden Effektstärken. Im Schnitt zeigen sich ein mittlerer Effekt ($g = 0{,}65$) auf die Leistung und große Moderationseffekte auf die Wirksamkeit durch Lehrkräfteschulungen ($g = 0{,}84$). Von zentraler Bedeutung für die Wirksamkeit des Einsatzes digitaler Tools ist dabei die *professionelle Kompetenz* von Lehrkräften. Im Zentrum stehen hier die Kompetenzen bezüglich der Planung und Durchführung von digital unterstütztem Unterricht (unterrichtliche Kompetenzen in digitalen Lehr-Lern-Settings). *Digitale Unterrichtskompetenz* kann dabei aufbauend auf Definitionen digitalen Lernens (z. B. Wheeler 2012) sowie unter Rückgriff auf ein holistisch-prozessuales Unterrichtsverständnis verstanden werden, als die *Planung, Durchführung und Kontrolle technologisch gestützter oder virtueller Lehr-Lern-Settings in schulischen Kontexten*. Eine aktuelle Untersuchung des Bundesverbands für Lehrkräfte an Beruflichen Schulen (BVLB) auf Basis von Lehrkräftebefragungen verweist darauf, dass entsprechende Kompetenzen der Lehrkräfte durchaus vorhanden sind (vgl. Gerholz, Schlottmann, Faßhauer u. a. 2022). Allerdings zeigt eine Charakterisierung der Unterrichtspraxis während der Corona-Pandemie an kaufmännischen Schulen in Baden-Württemberg auf Basis des SAMR-Modells (vgl. Puentedura 2014), dass Lehrkräfte digitale Tools in erster Linie substitutiv zur Distribution bestehender Unterrichtsmaterialien nutzen. Lernangebote, die einen zusätzlichen Mehrwert i. S. einer Transformation bieten, sind dagegen seltener zu finden (vgl. Mayer, Gentner & Seifried im Druck).

Im vorliegenden Beitrag wird exemplarisch anhand der Funktionen der LUCA Office Simulation erörtert, welche Fähigkeiten Lehrkräfte zur Nutzung transformativer digitaler Unterrichts-Konzepte benötigen. Zunächst werden in Kapitel 2 zentrale Funktionen der LUCA Office Simulation erläutert. Kapitel 3 enthält didaktische Design-Empfehlungen für den unterrichtlichen Einsatz von LUCA. Eine konkrete Umsetzung für den kaufmännischen Bereich wird in Kapitel 4 gezeigt. Im Anschluss werden die professionellen (digitalen) Kompetenzen von Lehrkräften diskutiert (Kapitel 5), die schließlich mit Blick auf die unterrichtliche Nutzung virtueller Lernsimulationen konkretisiert werden (Kapitel 6).

2 Funktionen der Bürosimulation LUCA

Die Bürosimulation LUCA ist eine browserbasierte Lehr-Lern-Umgebung, die im Rahmen des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts „Problemlöseanalytik in Bürosimulationen" (PSA-Sim) entwickelt wurde (vgl. Rausch, Deutscher, Seifried u. a. 2021). LUCA ermöglicht es Lehrenden, authentische, adaptive Arbeitsszenarien für Lernende bereitzustellen sowie die Lernprozesse zu begleiten. Das LUCA Office stellt Lernenden typische Werkzeuge eines kaufmännischen PC-Arbeitsplatzes zur Verfügung, in denen die Arbeitsszenarien bearbeitet werden. Adaptivität wird über eine logdatenbasierte Echtzeitanalyse der Problemlöseprozesse (Problem Solving Analytics; PSA) ermöglicht, die von Lehrenden auch ohne spezifische IT-Kompetenzen konfiguriert werden kann. LUCA läuft als betriebssystemunabhängiger Online-Dienst und setzt auf Seiten der Nutzenden lediglich eine Internetverbindung und einen aktuellen Internetbrowser voraus. Abbildung 1 gibt einen Überblick über die wichtigsten Funktionen der LUCA Komponenten.

Das LUCA Office bietet Lernenden Softwarewerkzeuge an, wie einen E-Mail-Client, ein Ordner- und Dateisystem inklusive Document Viewer für PDF-, Grafik- und Videodateien, ein Tabellenkalkulations- und ein Textverarbeitungsprogramm sowie ein Enterprise Resource Planning (ERP) System mit Recherchefunktionen (read only). Der LUCA Editor ermöglicht Lehrkräften das Erstellen eigener Szenarien oder das Kopieren und Anpassen bestehender Szenarien, die in ein ebenfalls editierbares Modellunternehmen eingebettet werden können. Ein Arbeitsszenario beinhaltet PDF-Dokumente (z. B. Briefe, Rechnungen, Angebote). Zudem können bearbeitbare Tabellen und Textdokumente definiert sowie E-Mails erstellt werden, die auch erst nach einer vordefinierten Laufzeit eintreffen können. Ferner kann für ein Modellunternehmen ein umfangreicher Datenkranz im ERP-System bereitgestellt werden. Arbeitsszenarien können Interventionen und Ereignisse enthalten, die sich adaptiv an die Lernenden anpassen. Im LUCA Manager stellen Lehrkräfte Projekte aus Arbeitsszenarien und Fragebögen zusammen, laden Lernende ein, verfolgen – bei synchronen Projekten – die Bearbeitungsprozesse, können per Chat intervenieren und bewerten nach Bearbeitungsende die Lösungsqualität anhand von Scoring Rubrics, die ebenfalls individuell angelegt werden können.

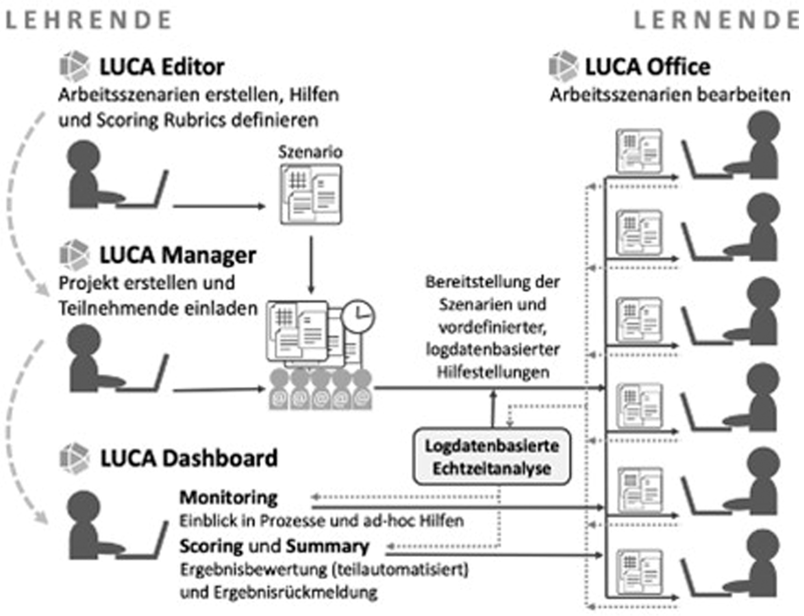


Abbildung 1: Übersicht der LUCA Software-Komponenten (Rausch, Deutscher, Seifried u. a. 2021, S. 379)

Die Lernumgebung ermöglicht in mehrfacher Weise eine adaptive Gestaltung der Lehr-Lern-Prozesse: (1) Auf Basis von Personenangaben (z. B. Name, Geschlecht) lässt sich die Interaktion in Form von E-Mails personalisieren. (2) Auf Basis von Antworten der Lernenden auf kurze Abfragen innerhalb von Ereignissen können personalisierte Interventionen/Prompts in Form von E-Mails ausgelöst werden. Ereignisse sind Overlays, die üblichen Interaktionen am Arbeitsplatz ähneln und in denen Fragen eingeblendet werden. So sind z. B. Anpassungen mit Blick auf das aktuelle Erleben, spezifische Interessen oder Vorwissen der Lernenden möglich. In Abhängigkeit von deren Reaktionen werden vordefinierte E-Mails gesendet, die auch spezifische Hilfen enthalten können. Schließlich lassen sich auf Basis von Verhaltensdaten innerhalb der Lernumgebung logdatenbasierte Interventionen – ebenfalls in Form vordefinierter E-Mails – auslösen. Interventionen erfolgen u. a., wenn nach einer bestimmten Laufzeit für die Aufgabenbearbeitung notwendige Aktionen der Lernenden nicht oder fehlerhaft erfolgt sind.

3 Didaktisches Design

Das der Bürosimulation LUCA zugrunde gelegte didaktische Design fasst – aufbauend auf konstruktivistischen Ideen – Lernen im konnektivistischen Sinne als einen Prozess auf, der v. a. in realen oder virtuellen (Wissens-)Netzwerken stattfindet, die aus Menschen (z. B. Kolleginnen und Kollegen, Kundinnen und Kunden, Lehrenden), aber auch nicht-menschlichen Entitäten (z. B. Künstliche Intelligenz) bestehen. Wissen wird dabei weiterhin konstruktivistisch als individuelle Sinnkonstruktion aufgefasst. Jedoch wird dieses Wissen konnektiv über Netzwerke als (exponentiell wachsende) geteilte Ressource bereitgestellt und existiert damit auch außerhalb des Individuums (vgl. Siemens 2004). Entsprechende Überlegungen finden sich auch in verbreiteten beruflichen Lerntheorien wieder (z. B. Situated Learning, vgl. Lave & Wenger 1991). Mit Blick auf die Gestaltung von Lehr-Lern-Situationen erscheinen folgende Designkriterien als relevant:

1. Bezüglich der Implementation von problemhaltigen Aufgaben ist eine *Orientierung an vollständigen beruflichen Handlungen* empfehlenswert (vgl. Hacker 1986). Hierzu empfiehlt es sich, auf Basis von realen Arbeitsprozessen berufliche Arbeitsaufgaben zu identifizieren und mittels einer domänenspezifischen Aufgabenanalyse in authentische Arbeitsszenarien zu überführen (zur Vorgehensweise siehe Aprea, Ebner & Müller 2010). Im Kontext der LUCA Bürosimulation sprechen wir diesbezüglich von Arbeitsszenarien, die in eine realistische „Story“ sowie in einen konkreten Unternehmenskontext (ein Modellunternehmen) eingebunden sind. Diese Arbeitsszenarien sind typischerweise problemhaltig (vgl. Jonassen 2000). Das Ausmaß der kognitiven Anforderung sollte sich hierbei zum einen am Grad der realen beruflichen Aufgabe orientieren und zum anderen am Leistungsstand der Lernenden.
2. Hinsichtlich der inhaltlichen Sequenzierung der Instruktionseinheiten erscheint die Orientierung an realen Geschäftsprozessen (vgl. Deutscher 2019) als zielführend. Hierzu werden mehrere Arbeitsszenarien in LUCA entsprechend ihrer typischen Sequenzierung im realen Geschäftsprozess dargeboten.
3. Folgt man dem konnektivistischen Netzwerk-Gedanken (s. o.), dann sollten Aufgabenstellungen und Informationen sozial situiert werden. Dies bedeutet, dass die Lernenden als zentrale Akteurinnen und Akteure im Zentrum der Aufgabenstellung stehen („social placement“) und durch Aktion („social action“) und Reaktion („social reaction“) in Interaktion („social interaction“) mit ihrer sozialen Umgebung treten (vgl. Braunstein, Deutscher, Seifried u. a. 2021). Sofern kollaboratives Lernen gefördert werden soll, können durch die Kombination mit externen Tools (z. B. Zoom oder Teams) auch Gruppenarbeiten ermöglicht werden, sodass die Lernerfahrung selbst kollaborativ stattfindet („social collaboration“) (ebd.).
4. Es sollte eine der realen Aufgabenstellung entsprechende, realistische Informationsmenge zur Verfügung stehen, um bei Lernenden Suchstrategien und den Umgang mit Information zu fördern. In LUCA kann unter Rückgriff auf das

4C/ID Modell (vgl. van Merriënboer & Kirschner 2018) zwischen Informationen zur Lernaufgabe selbst (z. B. Auftragsdetails, Aktennotizen im ERP-System), unterstützenden Informationen (z. B. Fachwissen, domänenspezifische Modelle oder Heuristiken) sowie prozeduralen Informationen (in Form von szenariospezifischen Prompts) unterschieden werden (vgl. Rausch, Deutscher, Seifried u. a. 2021). Zudem ist der Einbezug extern geteilter Wissensressourcen möglich.

5. Bedeutsam ist zudem die Anpassung der Lernumgebung an individuelle Bedürfnisse der Lernenden. Personalisierung umschreibt hierbei einen Aspekt der individuellen Förderung. Didaktisch soll durch die Bereitstellung von effektivem Feedback einer kognitiven und emotionalen Überforderung entgegengewirkt werden („Scaffolding and Fading“, s. Cognitive Apprenticeship; vgl. Collins, Brown & Newman 1989). Dies geschieht in LUCA über Prompts. Prompts sind Hilfestellungen bzw. Hinweise in Form von Fragen, Vorschlägen und Feedback, die während des Lernprozesses dargeboten werden und die Anwendung relevanter Verarbeitungsstrategien fördern (vgl. Wirth 2009). Da der Einsatz von Prompts zusätzliche mentale Ressourcen erfordert, sollten Prompts keine neuen Informationen beinhalten, sondern vielmehr den Abruf und die Ausführung von Handlungsweisen unterstützen (vgl. Bannert 2009). Zur Vermeidung von „Overprompting“ sollten die Prompts möglichst knapp bzw. wenig komplex formuliert sowie adaptiv ausgestaltet sein (i. S. von „Scaffolding und Fading“). Daneben sind sie möglichst zeitgerecht zu präsentieren, damit sie im Aufgabenverlauf nicht disruptiv wirken und es eindeutig ist, auf welchen Aufgabenaspekt Bezug genommen wird (vgl. Renkl & Scheiter 2017). Didaktisch sinnvoll eingesetzt, unterstützen Prompts Lernende bei der Selbstregulation und -steuerung (vgl. Mead, Buxner, Bruce u. a. 2019).

4 Ein Anwendungsbeispiel: Das Arbeitsszenario „Lieferantenauswahl”

Im Folgenden wird am Beispiel des Lerninhalts „Lieferantenauswahl“ gezeigt, wie sich die in Abschnitt 3 skizzierten Designprinzipien in der Bürosimulation LUCA umsetzen lassen.

4.1 Lerninhaltsanalyse

Die Angebotsauswahl mittels Nutzwertanalyse ist fester Bestandteil von kaufmännischen Rahmenlehrplänen und gilt als kaufmännische Querschnittsaufgabe. Bei der Aufbereitung der Lerninhalte für die LUCA Office Simulation sind wir wie folgt vorgegangen: Im Rahmen einer domänenspezifischen Aufgabenanalyse wurden zunächst typische Arbeitsschritte und relevante Wissensaspekte der übergeordneten Teilschritte „Angebote auswerten“, „Entscheidung treffen“ und „Entscheidung kommunizieren“ bestimmt sowie ein Ablaufszenario für die Aufgabenbearbeitung festgelegt. Hierfür wurden im Rahmen einer kognitiven Aufgabenanalyse relevante Wissensaspekte identifiziert und in Anlehnung an Anderson und Krathwohl (2001) den Wissensarten Faktenwissen (FaW), konzeptuelles Wissen (KonW), prozedurales Wissen (ProzW) und metakognitives Wissen (MetaW) zugeordnet. Aufbauend auf diesem ersten Analyseschritt wurde das Szenario auf Basis von Arbeitssituations- und Lehrbuchanalysen im Detail konzipiert. Den Lernenden wird durch eine fiktive vorgesetzte Person die Aufgabenstellung per E-Mail gesendet. Auf Basis mehrerer Angebote und weiterer Informationen ist ein Lieferant auszuwählen. Hierzu sind verschiedene Kriterien (Bezugspreis, Qualitätsbewertung, Lieferzeit, ethische und ökologische Aspekte) von Relevanz. Die Lernenden führen eine Nutzwertanalyse durch, treffen eine Vorentscheidung und begründen diese.

4.2 Prompt-Design

Auf Basis der skizzierten Design-Überlegungen wird die Bearbeitung des Arbeitsszenarios durch ein Prompt-Design unterstützt. Insgesamt wurden verschiedene kognitive, nichtkognitive und metakognitive Lernprompts in das Arbeitsszenario eingebettet (für eine Übersicht, s. Tab. 1). Dabei wurden die kognitiven Prompts auf die Eingaben der Lernenden in die Tabellenvorlage zugeschnitten. Hierfür wurden zunächst alle plausiblen Eingabewerte im Rahmen einer Analyse möglicher Fehler bestimmt. Anschließend wurde jede für die Lösung relevante Zelle mit Auslösebedingungen für die Prompts versehen. Wählen die Lernenden beispielsweise bei der Berechnung des Bezugspreises einen falschen Wechselkurs, erhalten sie zeitnah einen personalisierten Prompt mit dem Hinweis auf die potenzielle Fehlerquelle (Tab. 1, Nr. 1). Die nichtkognitiven Prompts zielen auf die Steigerung der Lernmotivation ab. In der Aufgabenstellung wird z. B. erwähnt, dass zu den vorhandenen Auswahlkriterien weitere Aspekte berücksichtigt werden können. Ergänzen die Lernenden nun eigenständig weitere Kriterien (z. B. Umweltverträglichkeit, ethische Aspekte), erhalten sie einen verstärkenden Prompt (Tab. 1, Nr. 2). Weiterhin wurden metakognitive Prompts (Tab. 1, Nr. 3) implementiert, um die Lernenden logdatenbasiert auf ggf. nicht gesichtete relevante Informationen hinzuweisen.

Tabelle 1: Übersicht über das Aufgaben- und Prompt-Design

Nr.	Lösungsschritt im Arbeitsszenario	Prompt-Art	Auslösebedingung	Prompt-Darbietung	Prompt-Inhalt
1	Die Lernenden berechnen den Bezugspreis und tragen ihn in der Tabellenvorlage vorgesehenen Zelle ein.	kognitiv	falscher Wert in Zelle L14 der Tabellenkalkulation *oder* fehlender Wert in Zelle L14 der Tabellenkalkulation (nach X Minuten)	E-Mail-Intervention	Hallo (Anrede), haben Sie bei der Währungsumrechnung den aktuellen Wechselkurs beachtet? Eine Tabelle zu den Wechselkursen finden Sie im Nachschlagewerk. Mit freundlichen Grüßen

(Fortsetzung Tabelle 1)

Nr.	Lösungsschritt im Arbeitsszenario	Prompt-Art	Auslösebedingung	Prompt-Darbietung	Prompt-Inhalt
2	Die Lernenden ergänzen selbstständig weitere Auswahlkriterien in die Tabellenvorlage für die Nutzwertanalyse.	nicht-kognitiv	Textinput in Zellen B17 bis B19 (nach X Minuten)	E-Mail-Intervention	Hallo (Anrede), Sie haben gut erkannt, dass es sinnvoll sein könnte, auch weitere Auswahlkriterien für die Nutzwertanalyse heranzuziehen. Machen Sie weiter so!
3	Die Lernenden identifizieren notwendige Informationen zur Ermittlung der Auftragswerte.	meta-kognitiv	Nicht-Öffnen einer relevanten Datei im ERP-System (nach X Minuten)	E-Mail-Intervention	Hallo (Anrede), haben Sie sich schon die Aktennotiz des Lieferanten Jinshu Gongsi anschauen und in Ihrer Auswahl berücksichtigen können?
4	Die Lernenden verschaffen sich einen Überblick über die Aufgabenanforderungen und die zur Lösung notwendigen Dokumente.	meta-kognitiv	Abhängig von der Antwortauswahl der Ereignisabfrage (Antwortmöglichkeit 3)	E-Mail-Intervention	Halle (Anrede), bevor ich mit der Lieferantenauswahl beginne, nehme ich mir immer einen Moment Zeit, um mir Notizen zu machen. Insbesondere die Erstellung einer Nutzwertanalyse erfordert verschiedene Arbeitsschritte. Im Nachschlagewerk zur Lieferantenauswahl finden Sie hierzu Informationen.

Weiterhin können sogenannte Ereignisse in LUCA zur Individualisierung der Aufgabenbearbeitung genutzt werden. Beispielsweise können auf Basis der Angaben der Lernenden (z. B. Erleben, Einschätzungen, Testfragen oder persönliche Präferenzen), die in Form von kurzen Abfragen eingeblendet werden, personalisierte Prompts ausgelöst werden. Im vorliegenden Beispiel erkundigt sich eine Kollegin nach dem Zwischenstand der Aufgabenbearbeitung (Abb. 2).

Im vorliegenden Beispiel wählen die Lernenden eine Antwortmöglichkeit, auf deren Basis verschiedene Prompts generiert werden. Bei Antwortmöglichkeit 1 (die Lernenden wissen, was zu tun ist) wird ein nichtkognitiver Prompt in Form eines Lobs angezeigt. Geben Lernende an, dass sie sich zunächst einen Überblick verschaffen müssen, wird ein metakognitiver Prompt zur Unterstützung der Aufgabenplanung generiert. Für diesen Prompt wird eine unspezifische und kurze Formulierung gewählt, um die Lernenden in der Wahl ihrer Problemlösestrategien nicht einzuschränken. Die Wahl der dritten Antwortmöglichkeit (Überforderung) löst einen detaillierten metakognitiven Prompt mit einer konkreten Anregung zur weiteren Vorgehensweise aus (Tab. 1, Nr. 4).

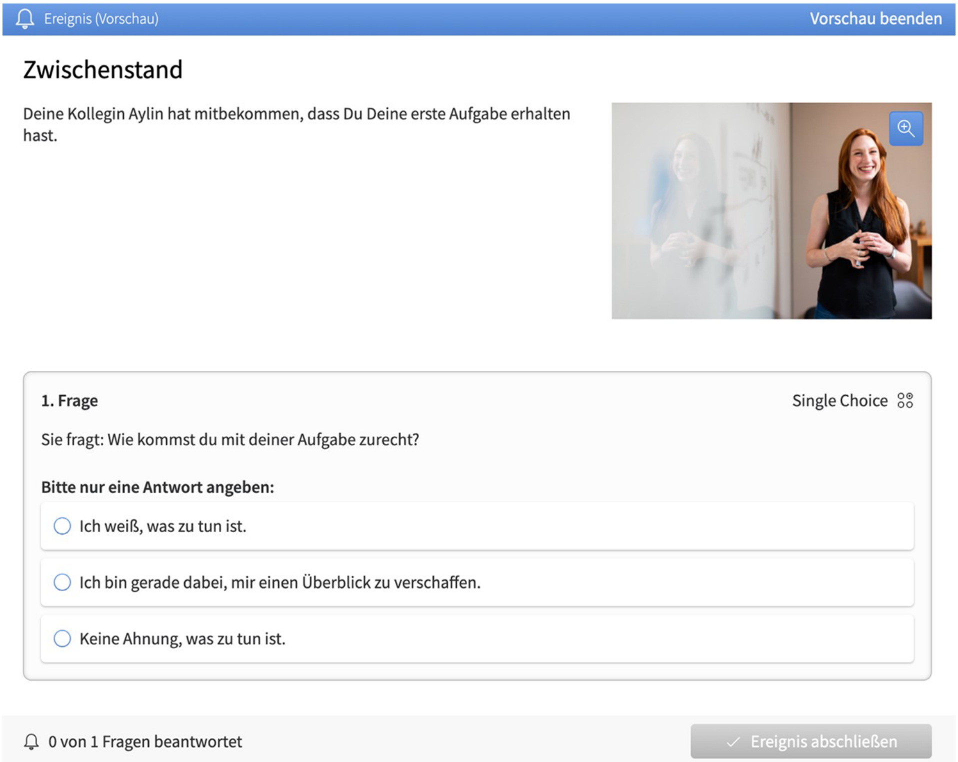


Abbildung 2: Ereignis als Abfrage zur Generierung personalisierter Prompts in LUCA

5 Modelle digitaler Unterrichtskompetenzen von Lehrkräften

Zur Einordnung der Kompetenzen, die Lehrkräfte für die didaktische Arbeit mit digitalen Tools wie der hier thematisierten LUCA Office Simulation benötigen, wird häufig auf das TPACK-Modell (vgl. Koehler & Mishra 2009) sowie das European Framework for Digital Competence of Educators (DigCompEdu, vgl. Redecker 2017) zurückgegriffen (für eine knappe Übersicht über weitere Modelle siehe Schmid & Petko 2020). TPACK bezeichnet das technisch-pädagogische Inhaltswissen (Technological Pedagogical Content Knowledge) und baut auf Shulman (1986) auf, der fachdidaktisches Wissen (Pedagogical Content Knowledge: PCK) als Schnittmenge von Inhaltswissen (Content Knowledge: CK) und pädagogischem Wissen (Pedagogical Knowledge: PK) beschreibt. Darüber hinaus wird das Modell um eine technologische Komponente (TK: Wissen über den Umgang mit digitalen Technologien) ergänzt, die Schnittstellen zu sämtlichen Wissensbereichen aufweist. TPK bezeichnet das technologisch-pädagogische Wissen über die Möglichkeiten und Grenzen der Einbeziehung von digitalen Elementen im Unterricht, wohingegen TCK das Wissen über die Möglichkeiten von Technologien zur Erarbeitung von Unterrichtsinhalten beschreibt. TPACK schließlich wird als Schnittmenge von TPK, TCK und PCK definiert (Abb. 3).

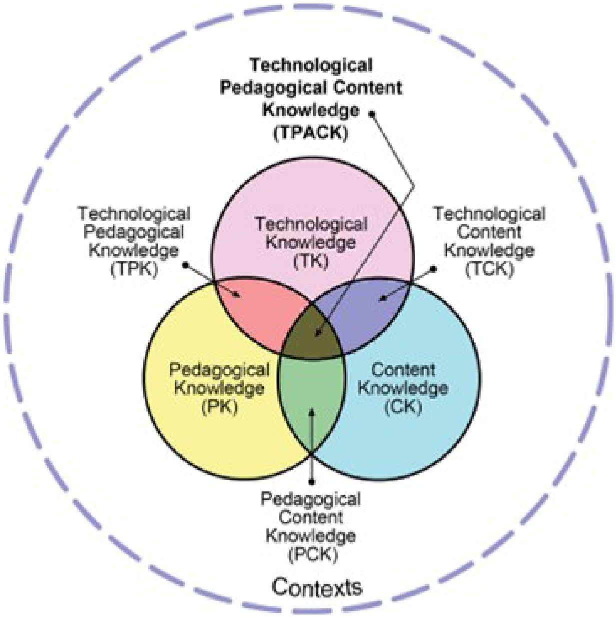


Abbildung 3: Technological Pedagogical Content Knowledge (vgl. Koehler & Mishra 2009)

Das DigCompEdu Framework wurde vom Joint Research Centre (2017) auf EU-Ebene entwickelt. Es umfasst insgesamt sechs Kompetenzbereiche (berufliches Engagement, Digitale Ressourcen, Lehren und Lernen, Evaluation, Lernendenorientierung, Förderung der digitalen Kompetenz der Lernenden) mit insgesamt 22 Kompetenzen, wobei die Bereiche zwei bis fünf (Digitale Ressourcen, Lehren und Lernen, Evaluation und Lernendenorientierung) unter der Klammer der pädagogischen und didaktischen Kompetenzen von Lehrenden den Kern bilden und die beiden restlichen Bereiche eins (Berufliches Engagement) und sechs (Entwicklung der digitalen Kompetenz der Lernenden) diesen Kernbereich flankieren. Im Vergleich zu TPACK adressiert DigCompEdu stärker auch die Lernaktivitäten und Kompetenzen der Lernenden. Tabelle 2 gibt einen Überblick über das DigCompEdu-Framework.

Das Framework wurde als Referenzrahmen für die Auseinandersetzung mit den digitalen Kompetenzen von Lehrenden auf allen Bildungsebenen entwickelt. Dabei stehen nicht technologische Aspekte im Vordergrund, sondern Ansätze, wie Lehrkräfte bessere Strategien im Umgang mit digitalen Unterrichtssettings erwerben können (Mikroebene). Auf der Mesoebene kann das Framework für die Schulentwicklung genutzt werden und auf der Makroebene eröffnet es Ansatzpunkte für die Qualitätssicherung in der Lehrerbildung (vgl. Redecker 2017).

Tabelle 2: Europäischer Rahmen für die digitale Kompetenz von Lehrenden (DigCompEdu, vgl. Redecker 2017)

1. Berufliches Engagement	2. Digitale Ressourcen	3. Lehren und Lernen	4. Evaluation	5. Lernendenorientierung	6. Förderung der digitalen Kompetenz der Lernenden
1.1 Berufliche Kommunikation 1.2 Berufliche Zusammenarbeit 1.3 Reflektierte Praxis 1.4 Digitale Weiterbildung	2.1 Auswählen digitaler Ressourcen 2.2 Erstellen und Anpassen digitaler Ressourcen 2.3 Organisieren, Schützen und Teilen digitaler Ressourcen	3.1 Lehren 3.2 Lernbegleitung 3.3 Kollaboratives Lernen 3.4 Selbstgesteuertes Lernen	4.1 Lernstand erheben 4.2 Lern-Evidenzen analysieren 4.3 Feedback und Planung	5.1 Digitale Teilhabe 5.2 Differenzierung und Individualisierung 5.3 Aktive Einbindung der Lernenden.	6.1 Informations- und Medienkompetenz 6.2 Digitale Kommunikation und Zusammenarbeit 6.3 Erstellung digitaler Inhalte 6.4 Verantwortungsvoller Umgang mit digitalen Medien 6.5 Digitales Problemlösen

6 Fazit: Kompetenzanforderungen an Lehrkräfte

Die Nutzung der LUCA Office Simulation erfordert von Lehrkräften in vielfältiger Weise professionelle Kompetenzen, die sich in den beiden skizzierten Modellen wiederfinden. Mit Blick auf den breiter angelegten DigCompEdu-Ansatz sind zunächst die reflektierte Praxis und die generelle Bereitschaft digitale Tools im Unterricht einzusetzen (Kompetenzfacette 1.3) zu nennen. Zentrale Bereiche adressieren dann Kompetenzen rund um die Auswahl bestehender bzw. die Gestaltung neuer Tools zur Durchführung von digital gestütztem Unterricht passend für die jeweilige Zielgruppe (Kompetenzbereich 2) sowie den unterrichtlichen Einsatz von digitalen Tools (Bereich 3). Die Evaluation der Effekte des Einsatzes von digitalen Tools ist Gegenstand von Kompetenzbereich 4. Aus didaktischer Sicht von Bedeutung sind daneben die Aspekte der Förderung der digitalen Teilhabe der Lernenden, der Differenzierung und Individualisierung, der aktiven Einbindung der Lernenden (Kompetenzbereich 5) sowie die Förderung verschiedenster digitaler Kompetenzen der Lernenden (Bereich 6: digitale Kommunikation und Zusammenarbeit, Erstellung digitaler Inhalte, digitales Problemlösen etc.). Mit Blick auf LUCA sind diesbezüglich insbesondere die Kompetenzbereiche 3 (Lehren und Lernen), 4 (Evaluation) sowie 5 (Lernendenorientierung) von Relevanz. Der Einsatz von LUCA erfordert von Lehrkräften, dass sie die Simulation angemessen in ihren Unterricht einbetten und entsprechende Lernszenarien gestalten oder auswählen. Zudem geht es um die Fähigkeiten von Lehrkräften, im Rahmen der Lernbegleitung die Lernfortschritte der Lernenden zu erfassen und individuell sowie auf Gruppenebene innerhalb und außerhalb des Unterrichts rückzumelden. Mit den oben skizzierten Prompts besteht zudem die Möglichkeit, neue Formen der Hilfestellung in den Unterricht zu implementieren bzw. das Lernen individuell zu begleiten. Nicht zuletzt sind Kompetenzen zur Anleitung und Begleitung kollaborativer und selbstgesteuerter Lernprozesse notwendig.

Zieht man das TPACK-Modell zur Beschreibung der einschlägigen professionellen Kompetenzen von Lehrkräften heran, so wird deutlich, dass der Einsatz von digitalen Tools wie der Bürosimulation LUCA von Lehrkräften Wissensbestände in sämtlichen Bereichen des Modells adressiert: Beispielsweise wird CK in Form von domänenspezifischem Inhaltswissen benötigt, um die domänenspezifische Aufgabenanalyse durchzuführen und ein Arbeitsszenario zu gestalten oder um ein passendes Szenario auszuwählen. Beim Umgang mit der Lernplattform ist TK von Bedeutung (z. B. für den Upload der Materialien). PK als fachübergreifendes Professionswissen über Lernprozesse und wirksame Unterstützungsmöglichkeiten wird als Hintergrundwissen z. B. bei der Wahl der unterrichtlichen Sozialform oder der Abschätzung der Wirksamkeit des Einsatzes formativ-diagnostischer Elemente bei der Planung der Unterrichtseinheit relevant. PCK benötigen die Lehrkräfte nicht nur bei der Einschätzung der Aufgabenschwierigkeit, der zu modellierenden Hilfestellungen oder der Antizipation typischer fachlicher Fehler der Lernenden, sondern vielmehr bei sämtlichen Entscheidungen zur fachdidaktischen Gestaltung der Arbeitsszenarien (u. a. in Bezug auf Authentizität sowie die Möglichkeiten der didaktischen Reduktion). TCK wird für die Nutzung der innerhalb der LUCA-Umgebung implementierten realistischen Arbeitswerkzeuge benötigt. TPK fließt bei der Wahl digitaler Lerntools und deren Einbettung in den Unterricht ein. TPACK schließlich ist für die Umsetzung der Lernsituationen in der Lernplattform auf Basis fachdidaktischer, pädagogischer und technischer Überlegungen (u. a. Gestaltung der Lernprompts) von Bedeutung.

Bei der Diskussion um digitale Kompetenzen von Lehrenden ist abschließend zu betonen, dass diesen eine entscheidende Bedeutung für Unterrichtsqualität und in der Folge für die Leistungen der Lernenden zugesprochen wird. Aktuelle Forschung zur Unterrichtsqualität benennt für den Präsenzunterricht drei zentrale Faktoren, nämlich (1) kognitive Aktivierung, (2) konstruktive Unterstützung und Strukturierung sowie (3) Classroom-Management (vgl. Praetorius, Klieme, Herbert u. a. 2018). Ergänzend können für den digital gestützten Unterricht Qualitätskriterien herangezogen werden, die der Forschungstradition des E-Learnings bzw. der Distance Education entstammen (vgl. Helm, Huber & Loisinger 2021). In einem umfassenden Framework führt beispielsweise Picciano (2017) diesbezüglich Qualitätskriterien wie „Content“, „Social/Emotional“, „Self-Paced“, „Dialectic/Questioning“, „Evaluation“, „Collaboration“, „Reflection“ sowie „Learning Community“ an. Wichtig ist an dieser Stelle der Hinweis, dass die genannten Basisdimensionen der Unterrichtsqualität für die Gestaltung von digital gestütztem Unterricht ebenfalls von zentraler Bedeutung sind und mit Blick auf virtuelle Lernumgebungen zu konkretisieren sind. Diesbezüglich geht es jenseits der Bereitstellung eines kognitiv aktivierenden Lernangebots insbesondere darum, Transparenz und Strukturen zu schaffen, Lernende dauerhaft an unterrichtlichen Interaktionen zu beteiligen sowie die Selbstregulation und Vernetzung der Lernenden in digitalen Settings zu fördern.

Literaturverzeichnis

Anderson, L. W. & Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Boston: Allyn & Bacon.

Aprea, C., Ebner, H. G. & Müller, W. (2010). „Ja mach nur einen Plan …" – Entwicklung und Erprobung eines heuristischen Ansatzes zur Planung kompetenzbasierter wirtschaftsberuflicher Lehr-Lern-Arrangements. Wirtschaft und Erziehung, 61(4), 91–99.

Bannert, M. (2009). Promoting self-regulated learning through prompts. Zeitschrift für Pädagogische Psychologie, 23(2), 139–145. doi: 10.1024/1010–0652.23.2.139.

Braunstein, A., Deutscher, V., Seifried, J., Winther, E. & Rausch, A. (2022). A taxonomy of social embedding-A systematic review of virtual learning simulations in vocational and professional learning. Studies in Educational Evaluation, 72, 101098.

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A. & Subramaniam, A. (2018). Skill shift automation and the future of the workforce. McKinsey Global Institute. Discussion Papers. https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce.

Collins, A., Brown, J. S. & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), Knowing, learning, and instruction: Essays in honor of Robert Glaser (pp. 453–494). Lawrence Erlbaum Associates, Inc.

Deutscher, V. (2019). Berufliche Handlungskompetenz und ihre Diagnostik: zwischen Bildungsanspruch und Verwertbarkeitserfordernissen. In J. Seifried, K. Beck., B.-J. Ertelt & A. Frey (Hrsg.), Beruf, Beruflichkeit, Employability (S. 95–116). Bielefeld: wbv.

Frey, C. B. & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114(C), 254–280. doi: 10.1016/j.techfore.2016.08.019

Gerholz, K.-H., Schlottmann, P., Faßhauer, U., Gillen, J. & Bals, T. (2022). Erfahrungen und Perspektiven digitalen Unterrichtens und Entwickelns an beruflichen Schulen. Berlin: Bundesverband der Lehrkräfte für Berufsbildung e. V.

Hacker, W. (1986). Arbeitspsychologie. Psychische Regulation von Arbeitstätigkeiten. Bern, Stuttgart & Toronto: Huber.

Helm, C., Huber, S. & Loisinger, T. (2021). Was wissen wir über schulische Lehr-Lern-Prozesse im Distanzunterricht während der Corona-Pandemie? – Evidenz aus Deutschland, Österreich und der Schweiz. Zeitschrift für Erziehungswissenschaft, 24, 237–311. doi: 10.1007/s11618-021-01000-z.

Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I. & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. Computers & Education, 153, 103897. doi: 1016/j.compedu.2020.103897.

Jonassen, D. H. (2000). Toward a design theory of problem solving. Educational Technology Research and Development, 48(4), 63–85. doi: 10.1007/BF02300500.

Koehler, M. J. & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? Contemporary Issues in Technology and Teacher Education, 9(1), 60–70. doi: 10.1177/002205741319300303.

Lave, J. & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. New York: Cambridge University Press.

Mayer, C., Gentner, S. & Seifried, J. (im Druck). Digitaler Unterricht an kaufmännischen Schulen in der Corona-Pandemie – Eine Bestandsaufnahme. Zeitschrift für Berufs- und Wirtschaftspädagogik.

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S. & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. Journal of Geoscience Education, 67(2), 131–142. doi: 10.1080/10899995.2019.1565285.

Picciano, A. G. (2017). Theories and frameworks for online education: Seeking an integrated model. Online Learning, 21(3). doi: 10.24059/OLJ.V21I3.1225.

Praetorius, A.-K., Klieme, E., Herbert, B. & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. ZDM, 50(3), 407–426. doi: 10.1007/s11858-018-0918-4.

Puentedura, R. (2014). SAMR, learning, and assessment. Zugriff am 04.02.2022. http://www.hippasus.com/rrpweblog/archives/2014/11/28/SAMRLearningAssessment.pdf

Rausch, A., Deutscher, V., Seifried, J., Brandt, S. & Winther, E. (2021). Die web-basierte Bürosimulation LUCA–Funktionen, Einsatzmöglichkeiten und Forschungsausblick. Zeitschrift für Berufs- und Wirtschaftspädagogik, 117(3), 372–394.

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu (JRC107466). Seville, Spain: Joint Research Centre. https://publications.jrc.ec.europa.eu/repository/handle/JRC107466

Renkl, A. & Scheiter, K. (2017). Studying visual displays: How to instructionally support learning. Educational Psychology Review, 29(3), 599–621. doi: 10.1007/s10648-015-9340-4.

Schmid, M. & Petko, D. (2020). Technological Pedagogical Content Knowledge als Leitmodell medienpädagogischer Kompetenz. Jahrbuch Medienpädagogik, 17, 121–140. doi: 10.21240/mpaed/jb17/2020.04.28.X.

Seeber, S., Weber, S., Geiser, P., Zarnow, S., Hackenberg, T. & Hiller, F. (2019). Effekte der Digitalisierung auf kaufmännische Tätigkeiten und Sichtweisen ausgewählter Akteure. Berufsbildung, 73(176), 2–7.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4–14. doi: 10.3102/0013189X015002004.

Siemens, G. (2004). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2, 3–10.

van Merriënboer, J. J. G. & Kirschner, P. (2018). Ten steps to complex Learning: A systematic approach to four-component instructional design. New York: Routledge/Taylor & Francis.

Wheeler, S. (2012). e-Learning and digital learning. In N. M. Seel (Ed.), Encyclopedia of the sciences of learning (pp. 1109–1111). New York: Springer.

Wirth, J. (2009). Promoting self-regulated learning through prompts. Zeitschrift für Pädagogische Psychologie, 23(2), 91–94. doi: 10.1024/1010–0652.23.2.91.

Abbildungsverzeichnis

Tabellenverzeichnis

4.3 Paper 3: How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects

Reproduced with permission from Elsevier.

Thomann, H., Zimmermann, J., & Deutscher, V. (2024). How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects. *Computers & Education*, 220, 105127. https://doi.org/10.1016/j.compedu.2024.105127

Computers & Education 220 (2024) 105127

Contents lists available at ScienceDirect

Computers & Education

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

How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects

Herbert Thomann *, Jan Zimmermann, Viola Deutscher

Chair of Business Education – Digital Vocational Learning, Georg-August University Goettingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany

A B S T R A C T

While Immersive Virtual Reality (IVR) technology has been predominantly employed in technical and medical academic education, it also holds significant potential for Vocational Education and Training (VET). IVR's unique properties, such as high immersion could be especially beneficial in VET, where action-oriented skills, domain-specific knowledge, and their application in new work contexts are crucial. This study investigates the effectiveness of IVR in vocational education, focusing on (1) objective knowledge acquisition, (2) subjectively perceived knowledge acquisition, and (3) motivational effects in the domain of warehouse logistics. Through a randomized controlled trial with 72 vocational students, we compared IVR-based learning to traditional paper-based methods. Results show that IVR did not improve immediate declarative knowledge acquisition; in fact, the paper-pencil group outperformed the IVR group on an objective post-test. However, IVR significantly enhanced students' perceived knowledge gains. The study also confirms higher motivation and immersion in IVR settings compared to paper-based learning environments. The identified discrepancy between perceived and actual learning may help explain the unclear state of research regarding knowledge acquisition in IVR studies, based on the measures used. Moreover, the findings underscore the necessity for a nuanced approach to IVR implementation in VET education. While IVR can be recommended for enhancing short-term learner engagement, traditional methods or a blend of IVR and non-immersive techniques may be more effective for fostering declarative knowledge in the short term.

1. Introduction

Immersive Virtual Reality (IVR) is a high-immersion technology that distinguishes itself from desktop-based VR using head-mounted displays (HMDs) and special controllers. Unlike desktop VR, which uses traditional computer screens and inputs, IVR creates a more encompassing experience by simulating reality with enhanced sensory stimuli, fostering a stronger sense of presence in the virtual environment (Makransky & Petersen, 2021). This heightened immersion is a key factor in differentiating IVR from other VR technologies and potentially influencing learning outcomes (Cummings & Bailenson, 2016).

While IVR technology has been predominantly employed in technical and medical academic education (Radianti et al., 2020), it also holds significant potential for vocational education and training (VET). VET is a structured educational framework that combines theoretical instruction at vocational schools with practical training in companies (Fürstenau et al., 2014). IVR's unique properties—such as high immersion—could be especially beneficial in VET, where imparting action-oriented skills, domain-specific knowledge, and their application in new work contexts are crucial (Buchner & Mulders, 2020; Conrad et al., 2022; Zinn, 2019). However, a significant research gap persists regarding systematically exploring IVR's utility in inculcating domain-specific competencies within VET (Conrad et al., 2022; Hellriegel & Čubela, 2018; Liu et al., 2023). It seems crucial to understand the various learning outcomes of adolescents when using IVR in terms of knowledge and ability as well as regarding emotional learning outcomes in order

* Corresponding author. Georg-August University Goettingen, Chair of Business Education – Digital Vocational Learning, Platz der Göttinger Sieben 5, 37073 Göttingen, Germany.

E-mail address: herbert.thomann@uni-goettingen.de (H. Thomann).

https://doi.org/10.1016/j.compedu.2024.105127
Received 18 March 2024; Received in revised form 14 July 2024; Accepted 24 July 2024
Available online 26 July 2024
0360-1315/

to provide recommendations for VET researchers, practitioners, and stakeholders alike. To fill this gap, we investigate the short-term effects of IVR use in the domain of warehouse logistics.

The empirical data for this study comes from Germany, where vocational education and training play a significant role in the educational system, with about half of young adolescents undergoing VET. Building on previous work by Ravichandran and Mahapatra (2023) on the challenges and possibilities of VR in VET and Liu et al.'s (2023) review of virtual and augmented reality environments, we conduct the first systematic literature review specifically focused on experimental studies using IVR in VET to systemize the existing state of research. We then use a strictly randomized experimental-control group design to investigate the effects of participating in IVR versus an identical content paper-based learning environment. To the best of our knowledge, this is the first study to combine a strictly randomized experimental-control group design with a substantial data set in the field of VET. Moreover, this study is among the first in IVR research (not only in VET) to explicitly investigate and report on the relationship between objective knowledge gains and subjectively perceived knowledge gains in IVR using correlation analysis. As far as we know, only Huang, Zhao, et al. (2023) have previously examined this relationship in an IVR context. This is of general importance as subjectively perceived knowledge acquisition is sometimes used to measure content-related learning in IVR studies (e.g., Makransky & Klingenberg, 2022; Makransky & Lilleholt, 2018). We are still unaware of how strong the relation between the two constructs is for declarative knowledge in IVR research.

2. Previous research in IVR and studies with a specific focus on VET

Research on IVR's effectiveness in knowledge acquisition compared to traditional methods has produced mixed results. Meta-analyses reveal a range of effect sizes across different educational levels and domains. Villena-Taranilla et al. (2022) found IVR significantly more effective ($g = 1.06$) than non-immersive methods in K6 education, particularly for short interventions. However, studies in K12 and higher education by Coban et al. (2022), Wu et al. (2020), and Luo et al. (2021) reported smaller effect sizes ($g = 0.20$–0.38), while other primary studies found no or negative effects. The high variance in effect sizes can be attributed to diverse methodologies, subject areas, and instructional content. Moreover, some studies distinguish between declarative and procedural knowledge within their analyses. While certain research suggests IVR's superiority over traditional methods (e.g., paper-pencil material, lectures) for both declarative and procedural knowledge (Conrad et al., 2024; Hamilton et al., 2021; Jensen & Konradsen, 2018), others find no significant benefits or even negative effects (Matovu et al., 2023; Won et al., 2023; Meyer & Pfeiffer, 2020; Makransky, Terkildsen, & Mayer, 2019).

Despite these mixed results in knowledge acquisition, IVR has consistently positive effects on affective and motivational factors. Numerous studies report enhancements in enjoyment, positive emotions, engagement, interest, motivation, and presence (Kolarik et al., 2024; Matovu et al., 2023; Makransky & Klingenberg, 2022; Radianti et al., 2020).

Conflicting findings on the learning effectiveness of IVR stem from various factors. Conrad et al. (2024) and Won et al. (2023) indicate challenges in assessing IVR's impact due to the classification ambiguity between declarative and procedural knowledge which arises from learning objectives and the influence of learner engagement levels. Additionally, factors like technical aspects (IVR hardware and software) and individual learner differences, such as prior knowledge and cognitive abilities, can significantly influence the learning outcomes in IVR (Makransky & Petersen, 2021; Won et al., 2023). Additionally, learners may face challenges due to limited familiarity with IVR, leading to increased cognitive load, and some studies report negative effects such as motion sickness, which can affect the learning experience despite high ratings in motivation and engagement (Matovu et al., 2023; Miguel-Alonso et al., 2023). Overall, many factors are involved in shaping and enhancing learning outcomes in IVR (Makransky & Lilleholt, 2018).

Regarding the state of IVR research in VET, for this introduction, we conducted a systematic literature search and identified only five experimental studies in this field (Kolarik et al., 2024; Kablitz et al., 2023; Makransky & Klingenberg, 2022; Chang, 2021; Lee, 2020).[1] These studies provide initial insights into the effectiveness of IVR in VET, though with mixed results.

In Kablitz et al.'s (2023) study, the IVR group designed a virtual salesroom using an IVR application, while the control group completed similar tasks using a paper-based floor plan. They observed a significant advantage ($\eta^2 = 0.261$) in domain-specific knowledge in the IVR group, despite some experiencing motion sickness. Lee (2020) similarly reported enhanced understanding of furniture production and processing tasks in the IVR group compared to traditional presentations and instructions. Makransky and Klingenberg (2022) compared IVR-based maritime safety training to traditional classroom instruction (personal trainer). While not reporting specific objective measurements of learning achievement, the study found the IVR group reported significantly higher perceived learning, enjoyment, and intrinsic motivation. Conversely, Chang (2021) found that IVR negatively impacted declarative knowledge learning outcomes in network cable service training, requiring a higher cognitive load from learners and affecting perception levels compared to conventional teaching methods. Kolarik et al. (2024) compared IVR learning with paper-based methods in logistics processes, finding no significant difference in learning success between groups, but higher intrinsic motivation, positive mood, and flow experience in the IVR group.

These IVR studies in VET share common limitations that impact their generalizability and robustness. Most suffer from small sample sizes (range = 29–60), with Kablitz et al. (2023) being the exception ($n = 79$). Many lacked strict randomization (Kolarik et al., 2024; Kablitz et al., 2023; Lee, 2020), potentially introducing selection bias. Specific challenges were also reported, such as participant

[1] The inclusion criteria (Table S1), literature search per database and search terms (Table S2), PRISMA Diagram (Fig. S1) and search results and categorization (excel file) can be found in the supplementary material. Two raters assessed whether studies based on the inclusion criteria should be included or excluded. Their agreement was calculated using a Cohen's Kappa coefficient of 0.865, indicating good interrater reliability between the two raters.

interaction issues in IVR conditions (Kolarik et al., 2024) and experimental fairness concerns (Lee, 2020).

Our research addresses these limitations by incorporating a substantial sample size (n = 72) and employing a pre-test/post-test design with individual-level randomization. We require participants to work individually under supervision to ensure consistent experimental conditions and minimize interference. Furthermore, we utilize comprehensive questionnaires grounded in Fokides and Antonopoulos's (2024) model and Tcha-Tokey et al.'s (2016) instruments, offering a more detailed measurement of relevant learning-related constructs in IVR. This approach aims to provide a more robust assessment of IVR's potential in VET, contributing to the nascent but growing body of research in this field.

3. Theoretical foundation and research questions

Based on the studies identified through the previous systematic literature review, we want to investigate the following three research questions:

(1) Does objectively and subjectively measured declarative knowledge acquisition differ between IVR and paper-based learning approaches?
(2) How strong is the relation between objective knowledge acquisition and subjectively perceived knowledge acquisition in both test settings (paper-based versus IVR)?
(3) To what extent do differences exist between the IVR and paper-based groups regarding mood, intrinsic motivation, and immersion during task completion?

Regarding the first research question, declarative knowledge refers to learners' understanding of facts and concepts by reproduction and reorganization in a task (Anderson & Krathwohl, 2001). The effectiveness of IVR on immediate declarative knowledge acquisition compared to paper-based learning remains unclear. While Conrad et al.'s (2024) systematic review based on four studies in higher education suggests IVR often outperforms analog media, such as paper-pencil material (e.g., Chittaro & Buttussi, 2015; Kablitz et al., 2023; Lee, 2020; Villena-Taranilla et al., 2019), other studies show no significant advantage (Kolarik et al., 2024; Makransky, Borre-Gude, & Mayer, 2019) or even negative effects (Chang, 2021; Meyer & Pfeiffer, 2020). Given the mixed state of research regarding research question one, we draw back to cognitive load theory (Sweller, 2020), and expect lower declarative knowledge gains in the IVR group due to higher cognitive load in IVR settings, particularly for unfamiliar users, which makes it harder for them to concentrate on the contents in IVR settings. Cognitive load theory suggests that working memory has limited capacity (Sweller, 2020). Educational material should minimize strain on working memory by reducing exogenous cognitive load (how information is presented) and increasing germane cognitive load (which supports learning) (Leppink et al., 2014). Research indicated that IVR can lead to higher cognitive load and therefore impede learning by overwhelming the learner (Makransky, Terkildsen, & Mayer, 2019).

Regarding the second research question, perceived knowledge acquisition refers to learners' subjective assessments of how a technology enhances their learning (Lee et al., 2010). Studies show that learners often report higher perceived learning from IVR compared to non-immersive methods such as classroom lectures and computer-based learning (Makransky & Klingenberg, 2022; Makransky & Lilleholt, 2018). However, this enhanced perception doesn't always align with objective measures of knowledge acquisition in immediate tests (Makransky, Terkildsen, & Mayer, 2019; Parong & Mayer, 2018). The current state of research on IVR reveals a significant gap in understanding the relationship between objective and subjective knowledge measures. Notably, Huang, Zhao et al. (2023) reported no correlation between objective knowledge and perceived learning in IVR environments based on n = 40 participants, marking the only study to explicitly examine this relationship. Moreover, it remains an open research question if the correlation between subjective and objective knowledge gains is lower for IVR than for paper-based learning, which could in part explain differing research results based on the measures (subjective versus objective) employed.

Regarding the third research question, learners' affective and motivational states significantly shape learner's perceptions and experiences in IVR environments (Fokides & Antonopoulos, 2024; Tcha-Tokey et al., 2016). Positive mood features include inspiration, attentiveness, excitement, and relaxation, while Negative mood features encompass anxiety, annoyance, and nervousness. Intrinsic motivation, as defined by Ryan and Deci (2000), refers to engagement in an activity for its inherent satisfaction rather than external rewards. Immersion, an objective feature determined by hardware capabilities, measures the extent to which users perceive sensory stimuli from the virtual world (Cummings & Bailenson, 2016). The rationale for using IVR is rooted in interest theory (Renninger & Hidi, 2016), which posits that students are more engaged when intrinsically interested in the material or when the learning environment elicits situational interest. IVR environments have the potential to stimulate both types of interest, potentially enhancing the learner's affective and motivational states. Empirical evidence supports this theoretical foundation. Previous studies investigating the effects of IVR on learning have consistently reported its superiority over traditional methods in fostering positive mood, motivation, and immersion (e.g., Kolarik et al., 2024; Makransky & Klingenberg, 2022; Makransky & Lilleholt, 2018; Parong & Mayer, 2018). We therefore expect that learning in IVR will result in significantly higher ratings of Intrinsic Motivation, Positive mood features, and Immersion compared to paper-based learning.

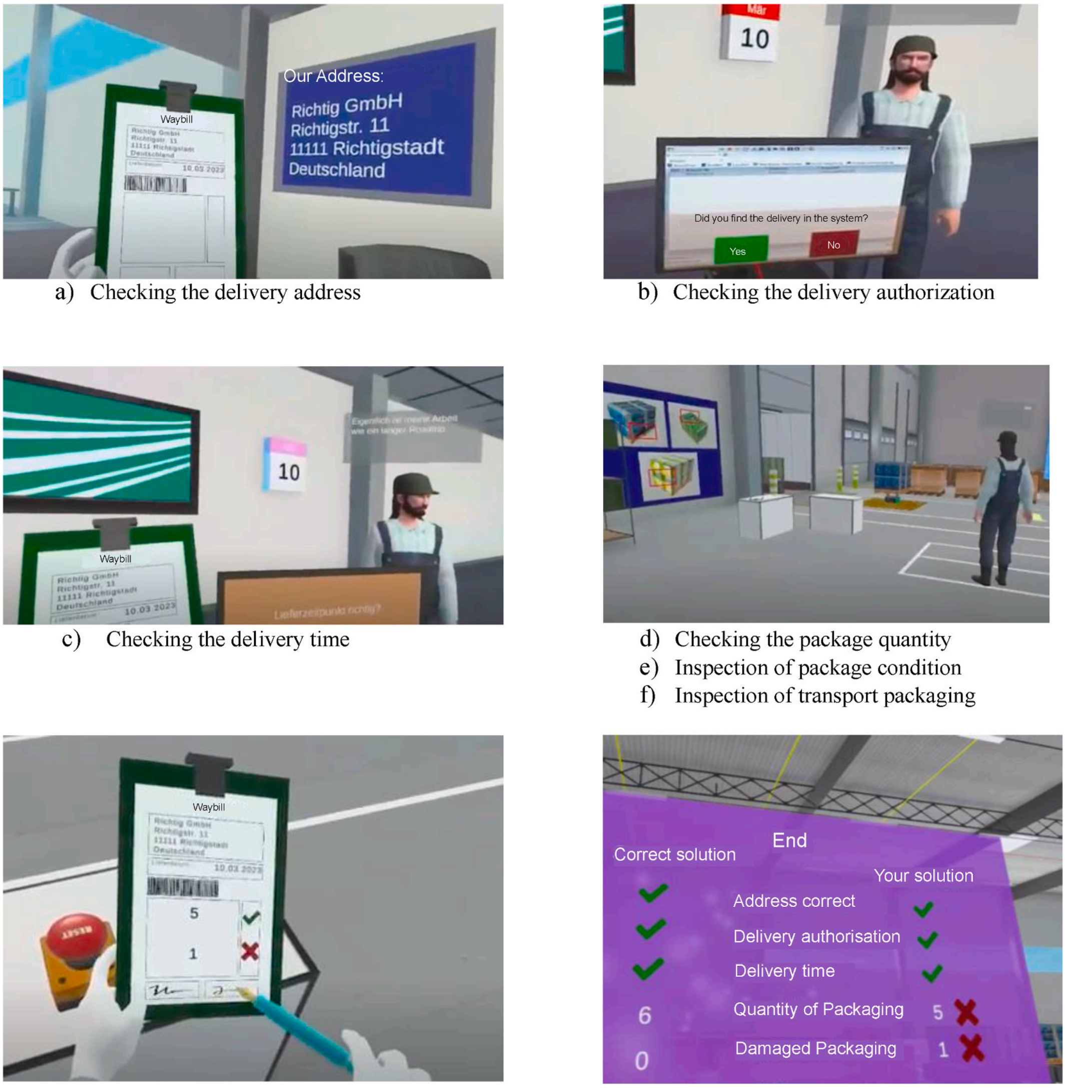
By addressing these three research questions, this study seeks to provide empirically grounded insights and recommendations for harnessing potentials of IVR in VET, ultimately enhancing the learning experience's quality and effectiveness.

4. Material

4.1. Classification of the IVR environment

As off-site teaching in the area of warehouse logistics mostly lacks action-oriented learning, IVR provides interactive simulations of real work situations, helping vocational education students experience and understand them better (Schäfer et al., 2023). This study employed the learning environment 'InGo' developed by the Fraunhofer Institute for Material Flow and Logistics to teach a goods-receiving process in the logistics industry (Schlüter & Kretschmer, 2020).

To systematically evaluate 'InGo', Won et al.'s (2023) framework for categorizing IVR design features was utilized, encompassing sensory, actional, narrative, and social aspects. The classification was independently conducted by a researcher and a virtual reality expert, achieving high interrater reliability (Cohen's kappa = 0.812). Discrepancies were resolved through discussion. 'InGo' was classified as a medium to high-quality IVR environment, featuring advanced sensory elements (e.g., medium-resolution graphics, realistic sound effects), intuitive controls with real-time feedback, coherent narrative contextualizing learning objectives, and simulated social interactions mirroring real-world professional communications in logistics (see Braunstein et al., 2022). A detailed breakdown of InGo's features is provided in Table S3 in the supplemental material.



a) Checking the delivery address
b) Checking the delivery authorization
c) Checking the delivery time
d) Checking the package quantity
e) Inspection of package condition
f) Inspection of transport packaging
g) Signing of the consignment note
h) Dashboard with detailed feedback

Fig. 1. IVR environment 'InGo', the procedure of the goods-receiving process (own representation by Kolarik et al., 2024; Schlüter & Kretschmer, 2020).

4.2. Learning objectives and content

The primary learning objective of the 'InGo' IVR environment is to teach learners the simplified goods acceptance process (Kolarik et al., 2024), which is an integral part of the formal curriculum for warehouse logistics trainees (KMK, 2004). In the IVR-based learning scenario, learners identify with a virtual avatar and receive and process goods delivered by the truck driver Ingo, simulating the actual goods-receiving process (Schlüter & Kretschmer, 2020).

4.3. Task scenario

Before commencing the task, a brief tutorial (on average 3 min and 26 s) acquaints learners with the controller and interface. During the task scenario (Fig. 1), learners must carry out several steps: checking the delivery address, delivery authorization, delivery time, parcel quantities, physical integrity of goods, and transport packaging. At each step, the scenario presents either a standard procedure or a deviation, such as an incorrect address or wrong quantity of packages, prompting users to identify deviations and respond accordingly, which may involve contacting a supervisor via phone.

4.4. Instructional design principles

Like that of Miguel-Alonso et al. (2023), our IVR environment integrated seven principles (multimedia presentation, signaling, coherence, spatial contiguity, temporal contiguity, redundancy, personalization) from Mayer's Cognitive Theory of Multimedia Learning (2021) to familiarize learners with IVR, reduce exogeneous cognitive load, and enhance learning (e.g., Albus et al., 2021). The first principle, multimedia presentation, was applied through visual cues alongside text instructions (Fig. 2) where learners verified order delivery dates. Our instructional design employs the signaling principle, using visual cues such as transparent bubbles, bold text, and color coding to direct attention. Embracing the coherence principle, unnecessary elements are excluded, and focus is maintained while playful dialogue for realism is introduced, as seen in Ingo's humorous exchanges in Fig. 2. Spatial and temporal contiguity principles were observed by positioning visual cues nearby and simultaneously enhancing coherence. The redundancy principle is addressed by only combining text and animation rather than narration. Lastly, the personalization principle is reflected in our informal narrative style, particularly evident in dialogues with the truck driver, contributing to a more engaging and relatable learning experience. Gamification elements were incorporated to enrich the gaming experience, including a decision-driven narrative guided by Ingo. The IVR environment also provides instructions for the learner at each step (Tutoring System). Interactive and animated elements, along with tailored feedback at the end (Fig. 1h), further enhance the learning experience (Kolarik et al., 2024).

5. Method

5.1. Data collection

For data collection, we collaborated with a vocational education school in South Germany. Participation was voluntary, and all participants provided written informed consent. The target group consisted of students from four warehouse logistics classes in their first year of a three-year apprenticeship. Four 90-min data collection sessions were organized (May 2023 and February 2024), in which the individual learners were randomly assigned to the intervention (IVR) or control (paper-based) group. The intervention group performed the task in the IVR environment. Before commencement, we provided a brief oral introduction of the experiment procedure,

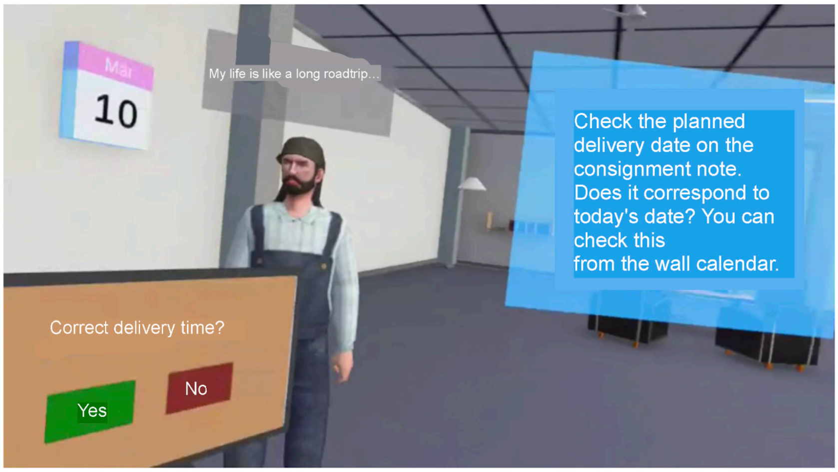


Fig. 2. Verifying the delivery time in the InGo IVR environment, based on seven principles of the Multimedia Learning Cognitive Theory of Mayer (2021) (translated from German).

learning topic, and simulation. Subsequently, each participant utilized Oculus Meta Quest (one and two) devices with associated controllers to interact with InGo. In contrast, the control group received traditional learning materials in paper format (Fig. S2). The traditional setting relied on images and materials from the IVR environment to maintain the authenticity of the narrative and keep the materials (pictures, cues) consistent. Both groups were tasked with solving the identical learning tasks. The VR group completed the simulation individually in a separate room, with each participant supervised by an experimental assistant (7–15 min). Meanwhile, the control group remained in the classroom, monitored by both the teacher and an experiment assistant, to ensure no collaboration among learners and to maintain consistent intervention duration for both groups (max. 15 min). The study protocol encompassed an initial pre-test, a general questionnaire, a post-test, and a final questionnaire.

5.2. Data sources

All tests and questionnaire items can be assessed as supplementary material. The pre-test (Fig. S3) assessed the students' prior domain-specific knowledge with three items in both groups based on the school curriculum for warehouse logistics (Omelicheva & Avdeyeva, 2008). The pre-test was modified from Kolarik et al. (2024) and improved together with a VET teacher specialized in logistics. The first two items were open-answer questions, and the last item was an assignment item. Partially correct answers were possible for each item, for a maximum score of 7.5 points.

Subsequently, a general questionnaire (Fig. S4) was used to record age, gender, previous experiences with VR technology, and the incoming goods processes of both groups ($\alpha = 0.82$), based on Shou and Olney (2021).

Following the intervention, both groups completed a general questionnaire section to provide information on how IVR-based learning differs from traditional paper-based learning (Table S4). Based on Fokides and Antonopoulos (2024), students were asked about positive ($\alpha = 0.78$) and negative ($\alpha = 0.87$) mood features using the "PANAS Scale" (Mackinnon et al., 1999). Furthermore, according to Wilde et al. (2009), motivation was assessed using the modified "Short Intrinsic Motivation Scale." Four areas of intrinsic motivation were defined: Interest and Enjoyment ($\alpha = 0.85$), Perceived competence ($\alpha = 0.83$), Perceived choice ($\alpha = 0.75$), and Pressure/Tension ($\alpha = 0.60$). The scale on Immersion ($\alpha = 0.89$) was based on Georgiou and Kyza (2017), and the single measure item perceived knowledge gain was constructed by adopting the wording from Kolarik et al. (2024) and adapting it from a questionnaire item by Lee et al. (2010).

The final questionnaire section, based on the "Unified UX Questionnaire" (Tcha-Tokey et al., 2016), was filled out by the IVR group only, to obtain more specific results for seven expressions (Table S5). According to the authors, this questionnaire has a satisfactory reliability ($\alpha = 0.718–0.908$) for the scales used in this context. Besides positive and negative emotions, experience consequence (motion sickness), engagement, flow experience, immersion, presence, and overall judgment of the IVR experience were measured (Fokides & Antonopoulos, 2024).

To avoid the post-test scores being influenced by students' memorization of the items from the pre-test, a parallel test (Blumberg, 1981) was used to measure learning outcomes based on the knowledge acquired during the intervention (Fig. S3), as previous studies have suggested (e.g., Kablitz et al., 2023). Moreover, the task difficulty was increased and adapted to the learning content of both interventions. The post-test consisted of five items: four multiple-choice items (each one point) and one assignment item (partially correct answers), for a maximum score of 7.5 points. While not psychometrically reliable due to time constraints, a limitation shared by similar studies (e.g., Kolarik et al., 2024; Lee, 2020), our tests accurately reflect all contents represented in the InGo simulation. Therefore, this approach aligns with formative rather than reflective measurement principles (Coltman et al., 2008), serving as a formative aggregated assessment of declarative knowledge acquisition from the content covered in the simulation.

5.3. Data analysis

5.3.1. G-power analysis

Our study's sample size ($n = 72$) is substantial for VR research, with high power for key measures such as post-test scores (0.84) and perceived learning gains (0.90). Most questionnaire scales demonstrate very high power (0.99), indicating robust findings. However, the Negative mood features (0.40) and Pressure/Tension (0.10) scales have lower power, necessitating cautious interpretation. Despite some limitations, our study's size and overall power allow us to contribute valuable insights to the emerging field of IVR in vocational education.

5.3.2. RQ 1: differences in objective knowledge gain and perceived knowledge gain

We tested assumptions of homogeneity of variance and normality. Levene's test confirmed homogeneity of variances for pre-test ($p = 0.59$), post-test scores ($p = 0.79$), and Perceived Learning ($p = 0.450$). According to George (2011), skewness and kurtosis values between −2 and +2 are acceptable to demonstrate normal univariate distribution. Hair et al. (2010) and Byrne (2013) also argue that data is considered normal if skewness is between −2 and +2 and kurtosis is between −7 and +7. The skewness and kurtosis values indicated that the pre-test scores for the control group (skewness = 0.4, kurtosis = 3.9) and the experimental group (skewness = 0.6, kurtosis = 4.1) were within acceptable ranges for normality. The control group's post-test scores also show near-symmetry and are nearly mesokurtic (skewness = 0.14, kurtosis = 2.6). The Jarque-Bera test indicates no significant deviation from normality ($p = 0.87$). The experimental group's post-test scores also show slight positive skewness and are nearly mesokurtic (skewness = 0.23, kurtosis = 2.6). The Jarque-Bera test confirms no significant deviation from normality ($p = 0.78$). Thus, both groups' post-test scores meet the normality assumption. To address the first research question, independent sample t-tests for pre- and post-tests and for perceived knowledge gains were conducted. Missing data (eight cases out of 72 participants in the post-test) were assessed for being Missing

Completely at Random (MCAR) using Little's test (p = 0.237). Complete Case Analysis (CCA) handled missing data to ensure the robustness of the findings.

5.3.3. RQ 2: relationship between objective knowledge and subjectively perceived knowledge

To explore the relationship between objective knowledge acquisition and subjectively perceived knowledge acquisition, we calculated the Pearson correlation coefficient (r) separately for both groups. We then applied Fisher's z-transformation to this coefficient, converting it to a z-score for more robust statistical inference and to enable comparisons across samples. For group comparisons, we used the q-statistic, representing the difference between Fisher's z-transformed correlations.

5.3.4. RQ 3: differences regarding mood, motivation, and immersion

The results of the Levene tests indicated that the variances were equal for most scales, with p-values greater than 0.05: Positive mood features (p = 0.430), Negative mood features (p = 0.170), Interest/Enjoyment (p = 0.354), Perceived competence (p = 0.199). However, significant differences in variances were observed for Perceived Choice (p = 0.013), Pressure (p = 0.025), and Immersion (p = 0.033), indicating that the homogeneity of variances assumption was violated for these scales. Thus, Welch's *t*-test was performed for these scales. In the control group, skewness values ranged from −0.61 to 1.28, and kurtosis values ranged from 1.77 to 4.10. In the experimental group, skewness values ranged from −1.56 to 2.03, and kurtosis values ranged from 2.03 to 5.99. The higher range of kurtosis was only due to the "negative mood feature" scale, which had slightly elevated values in both groups. However, the skewness and kurtosis values for all scales meet the normality criteria, supporting standard parametric tests. For the third research question, we conducted independent sample t-tests and Welch's t-tests comparing the pre- and post-questionnaire scores between the experimental (IVR) and control (paper-pencil) groups. Moreover, a Holm-Bonferroni correction was applied to all tests and questionnaire scales to control for multiple comparisons.

6. Results

6.1. Demographic data of the sample

The study sample comprised 72 participants, of whom 83.3% were male and 16.7% were female, with a mean age of M = 20.15 (SD = 2.26; Range = 15–28). The experimental group (IVR) consisted of 37 male and three female participants, with a mean age of M = 19.86 (SD = 2.53). The control group (paper-pencil) included 29 male and three female participants, with a mean age of M = 20.05 (SD = 1.83). No significant differences were found between the two groups regarding gender ($X^2_{(1)} < 0.001$, $p = 1$) and age ($t_{(72)} = -1.712$, $p = 0.246$, $d = -0.537$). Of the total sample size, 37 participants (51.4%) reported previous exposure to VR. Meanwhile, 60% of the participants in the experimental group had no previous experience with IVR, whereas 35% were in the control group. Leisure activities (e.g., gaming) were the most common context for the VR experience (64%), followed by the vocational school (10%), secondary school (10%), apprenticeship fairs (5%), activities at the workplace (5%) and not further specified contexts (5%). No significant differences were found between the intervention and control groups in terms of IVR experience ($t_{(72)} = 1.403$, $p = 0.165$, $d = 0.319$), familiarity with VR ($t_{(72)} = -0.554$, $p = 0.581$, $d = -0.250$), familiarity with the goods receiving process ($t_{(72)} = 1.607$, $p = 0.113$, $d = 0.251$), and experience with the goods-receiving process ($t_{(72)} = 0.113$, $p = 0.911$, $d = 0.280$).

6.2. Quality of the IVR experience for the treatment group

For the treatment group, we gathered comprehensive insights into various aspects of the immersive virtual reality learning experience. The findings, summarized in Table S6 in the supplementary material, reveal that participants reported predominantly positive emotions, engagement, immersion encounters, sense of presence, flow experience, and overall perception of the IVR experience. Furthermore, they experienced moderate levels of motion sickness during IVR learning. These findings align with recent studies, including Kablitz et al. (2023) and the meta-analysis by Wu et al. (2020), which reported similar moderate levels of motion sickness.

6.3. Differences in objective knowledge and perceived knowledge gain

The first research question was addressed by conducting two independent sample t-tests to compare domain-specific knowledge between the experimental (IVR) and the control (paper-pencil) group. One *t*-test was performed on the pre-test results (before the intervention), and another on the post-test results (after the intervention). Table 1 presents the results of these two t-tests, including t-

Table 1
t-test for knowledge test scores.

	t	p	CI Lower	CI Upper	IVR Mean	SD	Control Mean	SD	η^2	d
Pre-test	0.146	0.885	−0.437	0.505	2.99	0.96	2.95	1.03	<0.001	0.034
Post-test	−2.654	0.010	0.186	1.141	2.97	1.45	3.98	1.61	0.102	−0.674

values, p-values, confidence intervals, means, standard deviations, and effect sizes Cohen's d (Cohen, 1992) for both the pre-test and post-test comparisons between the IVR and control groups.

The pre-test results showed that both groups had, on average, low prior knowledge. The test scores for both tests ranged from 0 to 7.5 points. There was no significant difference between the IVR and control groups. At the pre-test, the IVR group achieved approximately 3 points on average ($M = 2.99$; $SD = 0.96$); the average for the control group was almost identical ($M = 2.95$; $SD = 1.03$). The t-value was 0.146 with a p-value of 0.885, indicating no statistical significance (95% CI: 0.437 to 0.505). The effect size was negligible with $\eta^2 < 0.001$ and Cohen's $d = 0.034$.

For the post-test, a significant difference was found between the groups, with the control group scoring higher. The mean post-test score for the IVR group was 2.97 ($SD = 1.45$) and for the control group, it was 3.98 ($SD = 1.61$). The t-value was −2.654 with a p-value of 0.010, indicating statistical significance (95% CI: 0.186 to 1.141). The effect size was moderate with $\eta^2 = 0.102$ and Cohen's $d = -0.674$, indicating the control group outperformed the IVR group.

Regarding the learners' perceived knowledge gain measured in the post-questionnaire, there was again a significant difference between the two groups ($t_{(72)} = 2.898$, $p = 0.005$, $\eta^2 = 0.107$, $d = 0.695$), but this time in the reverse direction, indicating that participants in the IVR group reported a significantly higher perceived knowledge gain compared to those in the paper-pencil group (Table 2). After applying the Holm-Bonferroni correction, the post-test comparison ($p = 0.04$) and the questionnaire item "Perceived learning gain" ($p = 0.025$) remained statistically significant.

6.4. Relationship between objective knowledge and subjectively perceived knowledge

Regarding the second research question, our analysis of the relationship between objective and subjective knowledge acquisition revealed differing patterns. The IVR group showed a weak, non-significant correlation ($r = 0.107$, $p = 0.555$), while the paper-pencil group demonstrated a moderate, significant correlation ($r = 0.401$, $p = 0.025$). However, Fisher's z-transformation and Q-test ($z = 1.21$, $p = 0.114$) indicated that the difference between these correlations was not statistically significant. This suggests that while the control group showed a stronger association between objective and subjective knowledge acquisition, we cannot conclude that this relationship significantly differs between the experimental and control conditions.

6.5. Differences regarding mood, motivation, and immersion

The third research question investigated the students' mood, motivation, and perception of immersion during the intervention. Table 3 presents the results of the independent samples t-tests and Welch's t-tests comparing the pre-and post-questionnaire scores between the experimental (IVR) and control (paper-pencil) groups.

Results indicated that participants in the IVR group reported significantly higher levels of Positive mood features than the control group ($t_{(4)} = 6.61$, $p < 0.001$, $\eta^2 = 0.384$, $d = 1.579$). Specifically, IVR participants demonstrated greater feelings of inspiration, attentiveness, excitement, and relaxation compared to the control group. Conversely, no significant differences were observed between the IVR and control groups regarding Negative mood features ($t_{(3)} = 1.47$, $p = 0.146$, $\eta^2 = 0.030$, $d = 0.352$) such as anxiety, annoyance, and nervousness.

Regarding the intrinsic motivation scale I (Interest/Enjoyment), participants in the IVR group exhibited significantly higher levels compared to the control group ($t_{(3)} = 5.82$, $p < 0.001$, $\eta^2 = 0.326$, $d = 1.391$). Specifically, they reported enhanced experiences of fun, interest, and enjoyment.

There were significant differences in the intrinsic motivation scale II (Perceived competence) between the IVR and control groups ($t_{(3)} = 4.32$, $p < 0.001$, $\eta^2 = 0.211$, $d = 1.034$). Specifically, participants in the IVR group reported higher levels of satisfaction with their performance during the intervention, perceived competence, and performance self-evaluation compared to the control group.

Participants in the IVR group exhibited significantly higher levels of intrinsic motivation in scale III (Perceived choice), compared to the control group ($t_{(3)} = 6.12$, $p < 0.001$, $\eta^2 = 0.350$, $d = 1.468$). Specifically, they reported greater self-choice and self-management in terms of their ability to choose to do the activity.

No significant differences were found in the intrinsic motivation scale IV (Pressure/Tension) between the IVR and control groups ($t_{(3)} = 1.12$, $p = 0.246$, $\eta^2 = 0.002$, $d = 0.089$). Participants in both groups reported similar levels of pressure, tension, and self-doubt.

Furthermore, participants in the IVR group reported higher Immersion levels than the control group ($t_{(3)} = 5.16$, $p < 0.001$, $\eta^2 = 0.277$, $d = 1.238$). Compared to the control group, IVR participants reported fewer irrelevant thoughts, were more focused, and more often lost track of time.

After applying the Holm-Bonferroni correction, all previously identified significant results remained statistically significant (range = 0.001 - 0.04), indicating the robustness of our findings.

Table 2
t-test for perceived learning gains.

Items	t	p	CI Lower	CI Upper	IVR Mean	SD	Control Mean	SD	η^2	d
Perceived learning gain	2.90	0.005	0.26	1.40	3.30	1.30	2.47	1.08	0.107	0.695

Table 3
T-tests comparing the pre-and post-questionnaire scores.

Items	t	p	CI Lower	CI Upper	IVR (n = 40)		Control (n = 32)			d
					Mean	SD	Mean	SD	η^2	
Positive mood features (4)	6.61	<0.001	1.03	1.91	3.97	0.96	2.50	0.91	0.384	1.579
Negative mood features (3)	1.47	0.146	−0.11	0.74	1.70	1.12	1.39	0.52	0.030	0.352
Interest/Enjoyment (3)	5.82	<0.001	0.82	1.68	4.17	0.88	2.92	0.93	0.326	1.391
Perceived competence (3)	4.32	<0.001	0.50	1.35	3.97	1.01	3.04	0.74	0.211	1.034
Perceived choice (3)*	6.12	<0.001	0.78	1.52	4.03	0.92	2.88	0.68	0.350	1.468
Pressure/Tension (3)*	1.12	0.264	−0.20	0.73	2.33	1.15	2.06	0.82	0.002	0.089
Immersion (3)*	5.16	<0.001	1.08	2.46	4.50	1.67	2.73	1.24	0.277	1.238

Likert scale 1 = "Strongly disagree" to 5 = "Strongly agree." Exception: Immersion scale, Likert scale 1 = "Strongly disagree" to 7 = "Strongly agree."
*Welch's *t*-test performed.

7. Discussion

7.1. Differences in objective knowledge and perceived knowledge gain

Our first research question compared the effectiveness of IVR and paper-pencil methods on domain-specific knowledge acquisition and perceived learning gains. Pre-test results confirmed comparable low levels of prior knowledge across both IVR and control groups, ensuring that post-test performance differences could be attributed to the interventions rather than pre-existing knowledge disparities. Post-test results revealed a significant difference in declarative knowledge acquisition. Interestingly, the control group achieved higher scores, indicating that traditional paper-pencil methods enhanced domain-specific knowledge more than the IVR method, as indicated by the negative Cohen's d value.

Our findings contrast with several previous studies: Kablitz et al. (2023), Lee (2020), Villena-Taranilla et al. (2019), and Chittaro and Buttussi (2015) reported significant advantages of IVR over paper-pencil methods. Additionally, our results differ from studies showing equal effects between IVR and paper-based learning on immediate declarative knowledge tests (Kolarik et al., 2024; Makransky, Borre-Gude, & Mayer, 2019). However, it is noteworthy that our study employed the same IVR environment as the pilot study of Kolarik et al. (2024) but tested the hypothesis against a larger sample (n = 72 vs. n = 41). Similarly to our study, Kolarik et al. (2024) noted that the paper-based learning group achieved higher scores in the knowledge test than the IVR group (7.38 vs. 8.35), but those differences were not significant for n = 41 participants. Our larger sample provides a clearer picture here and the results now align with studies indicating that IVR may be less effective for declarative knowledge acquisition: Chang (2021) and Meyer and Pfeiffer (2020) compared IVR directly with paper-based learning and found IVR to be less effective. Parong and Mayer (2018) and Makransky, Terkildsen, and Mayer (2019) compared IVR with computer-based learning and also found IVR less effective. While the comparison groups differ (paper-based vs. computer-based), the consistency in IVR's underperformance across these studies is noteworthy. This finding is particularly significant as it suggests that the potential benefits of VR's immersive nature may not translate directly into improved declarative knowledge outcomes, at least in the short term and in the context of our specific learning scenario.

Several factors may explain the superiority of the paper-pencil method in our study. First, the lack of familiarity with VR technology may have initially hindered learning by increasing exogeneous cognitive load in the form of distractions (Miguel-Alonso et al., 2024; Chang, 2021). Of students in the VR group, 60% had no previous experience with VR, and the time needed to complete the tasks varied widely based on the technical affinity of the students (7–15 min). Although efforts were made to familiarize participants with the IVR environment and hardware, the learning curve associated with the technology could have temporarily impeded knowledge acquisition. Secondly, both groups demonstrated low prior knowledge. Research suggests that incorporating a traditional learning method after the pre-test and before the IVR experience can enhance learners' autonomous learning capabilities (Paxinou et al., 2022; Meyer et al., 2019). Lastly, while the IVR simulation incorporated Mayer's multimedia principles (2021) to aid learners in familiarizing themselves with the IVR environment and gamification elements to enhance learning, the paper-based materials may have provided a more focused learning experience for declarative knowledge acquisition (Makransky, Terkildsen, & Mayer, 2019; Meyer et al., 2019).

Interestingly, despite the lower objective test scores, participants in the IVR group reported a significantly higher perceived knowledge gain than control group participants. This suggests that the immersive experience of VR enhanced learners' engagement and confidence in their learning, even if it did not translate into higher test scores. This is in line with the results of previous research (Makransky, Terkildsen, & Mayer, 2019; Makransky & Lilleholt, 2018).

The discrepancy in the results between objectively versus subjectively perceived learning could be partly attributed to the novelty effect. The learning experience's engaging and interactive VR setting was new for most participants and likely fostered heightened immersion, focus, and positive emotional responses (Miguel-Alonso et al., 2024). These factors may have influenced participants' subjective assessments, leading them to perceive greater knowledge gains. Moreover, the engaging nature of IVR increases the likelihood that learners will mistake increased engagement with actual learning progress (Sung et al., 2021). VR also offers greater control over the learning experience, providing students with a sense of autonomy that can likewise enhance their perceived learning experience (Makransky & Lilleholt, 2018). Assessment methods may also contribute to this discrepancy. Immediate retention tests may not fully capture the benefits of IVR learning. In contrast, more realistic tests that align closely with the immersive experience could show better performance, highlighting the gap between perceived and objective learning outcomes (Makransky, Terkildsen, & Mayer, 2019).

7.2. Relationship between objective knowledge and subjectively perceived knowledge

Our analysis of the relationship between objective and subjective knowledge acquisition further illuminates the discrepancy between perceived and actual learning outcomes. The IVR group exhibited a low and non-significant correlation between perceived and actual learning, suggesting a potential misalignment between students' confidence and their performance. This finding aligns with Huang, Zhao, et al. (2023), who similarly reported no correlation between objective knowledge and perceived learning in IVR environments. Although we observed a stronger correlation in the paper-based group with a medium-sized effect, the difference was not statistically significant, possibly due to an insufficient number of participants in each group. Comparing this to a medium-sized significant correlation for the paper-pencil group, this indicates that the misalignment between objective and subjective measures of knowledge gains appears significantly more pronounced in IVR environments.

The disconnect between perceived and actual learning in IVR challenges our understanding of how immersive technologies impact the learning process and self-assessment. Future research should investigate long-term learning outcomes and the effects of repeated IVR exposure to determine whether this disparity is a persistent feature of IVR learning or a result of novelty effects (Huang, Zhao, et al., 2023). Additionally, it needs to be considered whether this disconnect is possibly a result of the measures of knowledge employed, which are more similar to the paper-pencil learning environments than to IVR environments.

7.3. Differences regarding mood, motivation and immersion

The results on the third research question reveal significant differences in mood, motivation, and immersion between the IVR and paper-pencil group. Compared to the control group, participants in the IVR group reported higher levels of Positive mood features, such as inspiration, attentiveness, excitement, and relaxation. The large effect sizes underscore the impact of the IVR intervention on enhancing positive emotional states. Conversely, no significant differences were observed between the IVR and paper-pencil groups regarding Negative mood features, such as anxiety, annoyance, and nervousness. While the control group reported slightly lower levels of these negative emotions, the differences were not statistically significant. This suggests that the IVR intervention did not heighten negative emotional states despite 60% of students lacking prior VR exposure.

Furthermore, the IVR group exhibited significantly higher levels of intrinsic motivation across multiple scales compared to the paper-pencil group. Specifically, they reported enhanced experiences of fun, interest, enjoyment, satisfaction with their performance, perceived competence, and perceived freedom of choice. These results align with previous studies (e.g., Kolarik et al., 2024; Makransky & Klingenberg, 2022) that suggest IVR can positively influence mood and motivation, potentially enhancing learning outcomes.

Notably, participants in the VR group also reported higher immersion levels than the control group. They demonstrated fewer irrelevant thoughts, increased focus, and a greater tendency to lose track of time, suggesting heightened engagement and absorption in the IVR experience. The moderate to large effect sizes observed for immersion further support the potential of VR to create an immersive learning environment that captivates learners' attention and promotes sustained focus, as Makransky and Petersen (2021) hypothesized.

All these positive findings have to be interpreted cautiously, since they could be attributed partly or fully to the novelty effect, as also mentioned in the work of Allcoat and von Mühlenen (2018), as well as Miguel-Alonso et al. (2023). As Csikszentmihalyi and Larsen (1984) posited, learners are more motivated to learn in a different and new context than in traditional lessons. In the context of our study, the novelty and interactivity of IVR may have contributed to participants' heightened Positive mood features, Motivation, and Immersion. However, it is important to note that the novelty effect is often temporary, and its impact on emotional states, motivation, and immersion may diminish as learners become more familiar with VR technology (Huang, Huss, et al., 2023).

8. Limitations, future research and practical implications

Several limitations of our study are related to the sample and intervention setting. Here, the study focused solely on the warehouse logistics sector, leading to a pronounced gender imbalance with 83.3% male and 16.7% female participants. Additionally, the young age of the target group ($M = 20.15$) suggests the potential for conducting future research with older subjects. Another limitation regarding the intervention was the brief duration of tutorial and technical support provided to students engaging with IVR, stemming from time constraints within the vocational education school. This may have attributed to low familiarity with IVR and lowered objective knowledge gain in the experimental group.

Moreover, the study has two major limitations regarding the general study design:

(1) A clear limitation of our study must be seen in the focus on assessing declarative knowledge immediately after the IVR intervention. Two studies implementing one-week delayed post-tests (Chittaro & Buttussi, 2015; Meyer & Pfeiffer, 2020) found a positive effect for IVR groups compared to paper-based groups for long-term learning. Several theoretical assumptions could explain this potential long-term benefit. Since IVR provides a highly immersive learning environment, this may enhance memory retention through emotional experience. Moreover, the realistic nature of IVR environments could make learning experiences more memorable compared to traditional materials. Additionally, the active engagement required in IVR environments is known to improve learning and retention compared to passive methods like reading (Chittaro & Buttussi, 2015;

Meyer & Pfeiffer, 2020). However, these studies again have limitations, including small sample sizes and specific content areas. Thus, future research should address long-term memorization effects through IVR.

(2) Our study provided participants with a one-time-experience of IVR. It's unclear whether our positive findings regarding mood, motivation, and engagement for the IVR group persist over time or diminish due to a wear-off of positive novelty affects. As the novelty of IVR fades over time, learners may experience reduced motivation (Tsay et al., 2020). On the other hand, the lower declarative knowledge gains for the IVR group could be attributed again to negative novelty effects resulting from heightened exogeneous cognitive load due to unfamiliar interactions (Huang, Huss, et al., 2023). Additionally, according to Merchant et al. (2014), novice learners often prioritize exploration over learning tasks, diminishing learning gains in novel situations. Given VR's limited presence in schools and the brief familiarization period, an extended study with repeated interventions could yield different findings for the long-term use of IVR.

Summing up, the relationships between immediate emotional and cognitive experiences and long-term learning effects are not well understood. Addressing this requires an acknowledgement of the dynamic nature of memorization effects as well as the novelty effect in IVR experiences in longitudinal research designs with repeated IVR use and long-term assessments of cognitive outcomes.

9. Conclusion

Overall, our study offers a differentiated picture of IVR effectiveness in vocational education. A key finding is that, in the short term, the immersive nature of VR did not translate into improved declarative knowledge acquisition compared to traditional paper-pencil methods. Interestingly, despite lower objective test scores, the IVR group reported significantly higher perceived knowledge gains. This discrepancy challenges assumptions about IVR's general superiority for learning success in vocational domains due to higher immersion and highlights significant differences between perceived and actual learning in immersive environments. Based on this finding, we must caution researchers against the use of perceived learning as a proxy for objective learning, particularly in IVR settings. However, despite showing no immediate advantages in objective knowledge acquisition, IVR demonstrated significant benefits for motivation, immersion, and positive emotional responses as important outcomes of learning situations. This underscores the importance of considering factors beyond immediate knowledge gains when evaluating educational technologies (Kolarik et al., 2024; Conrad et al., 2022).

Moreover, the mixed findings of this study highlight the need for a nuanced approach to IVR implementation in education. Teachers could aim to harness IVR's motivational benefits and perceived learning gains to improve actual learning outcomes by combining paper-based and IVR learning environments if the goal is to foster declarative knowledge. Studies by Kablitz et al. (2023) and Lee (2020) found that after all students received basic training via paper-pencil material, those who then learned using IVR outperformed those who continued with paper-based materials. Apart from that, based on the state of IVR research, we hypothesize that IVR might be more suitable for learning procedural and automated knowledge than for the learning of declarative knowledge. To provide more concrete pedagogical advice, future IVR research should explore the key success factors in regular pedagogical use (natural experimental settings) for both short- and long-term outcomes, as well as their dynamic interplay. This will help determine not only if, but when and how to effectively use IVR as a complementary tool in vocational education and training.

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used Chat-GPT 3.5 in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used Chat-GPT 3.5 in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

CRediT authorship contribution statement

Herbert Thomann: Writing – review & editing, Writing – original draft. **Jan Zimmermann:** Writing – original draft. **Viola Deutscher:** Writing – review & editing.

Declaration of competing interest

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.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2024.105127.

References

Albus, P., Vogt, A., & Seufert, T. (2021). Signaling in virtual reality influences learning outcome and cognitive load. *Computers & Education*. https://doi.org/10.1016/j.compedu.2021.104154

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology, 26*.

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.

Blumberg, P. (1981). A practical methodology for developing content parallel multiple-choice tests. *The Journal of Experimental Education, 50*(2), 56–63.

Braunstein, A., Deutscher, V., Seifried, J., Winther, E., & Rausch, A. (2022). A taxonomy of social embedding-A systematic review of virtual learning simulations in vocational and professional learning. *Studies In Educational Evaluation, 72*, Article 101098.

Buchner, J., & Mulders, M. (2020). Lernen in immersiven virtuellen Welten aus der Perspektive der Mediendidaktik. *Medienimpulse, 58*(2).

Byrne, B. M. (2013). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. routledge.

Chang, Y. (2021). Effects of virtual reality application on skill learning for optical-fibre fusion splicing. *British Journal of Educational Technology, 52*(6), 2209–2226. https://doi.org/10.1111/bjet.13118

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review, 36*. Article 100452.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, 61*(12), 1250–1262.

Conrad, M., Dölker, J., Kablitz, D., & Schumann, S. (2022). VR in der kaufmännischen Berufsbildung: Potenziale–Befunde–Perspektiven. In S. Schumann, S. Seeber, & S. AbeleHrsg (Eds.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (pp. 231–255). Bielefeld. https://doi.org/10.3278/9783763971381.

Conrad, M., Kablitz, D., & Schumann, S. (2024). Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. *Computers & Education: X Reality, 4*, Article 100053.

Csikszentmihalyi, M., & Larsen, R. (1984). *Being adolescent: Conflict and growth in the teenage years*. New York, NY, USA: Basic Books.

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the efect of immersive technology on user presence. *Media Psychology, 19*(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Fokides, E., & Antonopoulos, P. (2024). Development and testing of a model for explaining learning and learning-related factors in immersive virtual reality. *Computers & Education: X Reality, 4*, Article 100048.

Fürstenau, B., Pilz, M., & Gonon, P. (2014). The dual system of vocational education and training in Germany–what can be learnt about education for (other) professions. *International handbook of research in professional and practice-based learning* (pp. 427–460).

George, D. (2011). *SPSS for windows step by step: A simple study guide and reference, 17.0 update, 10/e*. Pearson Education India.

Georgiou, Y., & Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. *International Journal of Human-Computer Studies, 98*, 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014

Hair, J., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2010). *Multivariate data analysis* (6th). Upper Saddle River NJ.

Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*(1), 1–32.

Hellriegel, J., & Čubela, D. (2018). Das Potenzial von Virtual Reality für den schulischen Unterricht - Eine konstruktivistische Sicht. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 58–80. https://doi.org/10.21240/mpaed/00/2018.12.11.X

Huang, X., Huss, J., North, L., Williams, K., & Boyd-Devine, A. (2023). Cognitive and motivational benefits of a theory-based immersive virtual reality design in science learning. *Computers and Education Open, 4*, Article 100124. https://doi.org/10.1016/j.caeo.2023.100124

Huang, X., Zhao, Q., Liu, Y., Harris, D., & Shawler, M. (2023). Learning in an immersive VR environment: Role of learner Characteristics and relations between learning and psychological outcomes. *Journal of Educational Technology Systems*. , Article 00472395231216943. https://doi.org/10.1177/00472395231216943

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies, 23*, 1515–1529.

Kablitz, D., Conrad, M., & Schumann, S. (2023). Immersive VR-based instruction in vocational schools: Effects on domain-specific knowledge and wellbeing of retail trainees. *Empirical Research in Vocational Education and Training, 15*(1), 9.

Kolarik, S., Schlüter, C., & Ziolkowski, K. (2024). Impact of VR on learning experience compared to a paper based approach. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 12*.

Kulturministerkonferenz. (2004). Rahmenlehrplan für den Ausbildungsberuf Fachkraft für Lagerlogistik. Retrieved from https://www.kmk.org/fileadmin/pdf/Bildung/BeruflicheBildung/rlp/FKLagerlogistik.pdf. (Accessed 8 March 2024).

Lee, I.-J. (2020). Applying virtual reality for learning woodworking in the vocational training of batch wood furniture production. *Interactive Learning Environments, 31*(3), 1448–1466. https://doi.org/10.1080/10494820.2020.1841799

Lee, E. A.-L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education, 55*(4), 1424–1442. https://doi.org/10.1016/j.compedu.2010.06.006

Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & Van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42. https://doi.org/10.1016/j.learninstruc.2013.12.001

Liu, Y., Zhan, Q., & Zhao, W. (2023). A systematic review of VR/AR applications in vocational education: Models, affects, and performances. *Interactive Learning Environments*, 1–18. https://doi.org/10.1080/10494820.2023.2263043

Luo, H., Li, G., Feng, Q., Yang, Y., & Zuo, M. (2021). Virtual reality in K-12 and higher education: A systematic review of the literature from 2000 to 2019. *Journal of Computer Assisted Learning, 37*(3), 887–901.

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the positive and negative affect schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences, 27*(3), 405–416. https://doi.org/10.1016/S0191-8869(98)00251-7

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691–707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Klingenberg, S. (2022). Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-WEIRD sample. *Journal of Computer Assisted Learning, 38*(4), 1127–1140. https://doi.org/10.1111/jcal.12670

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research & Development, 66*(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C.-C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education, 59*(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Mayer, R. E. (2021). The multimedia principle. In R. E. Mayer, & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 296–303). Cambridge University Press.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*. https://doi.org/10.1016/j.compedu.2019.103603

Meyer, L., & Pfeiffer, T. (2020). Comparing virtual reality and screen-based training simulations in terms of learning and recalling declarative knowledge. In *Delfi 2020 – Die 18* (pp. 55–66). Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.. Bonn: Gesellschaft für Informatik e.V.. PISSN, 1617-5468. ISBN: 978-3-88579-702-9.

Miguel-Alonso, I., Checa, D., Guillen-Sanz, H., & Bustillo, A. (2024). Evaluation of the novelty effect in immersive Virtual Reality learning experiences. *Virtual Reality, 28*(1), 27. https://doi.org/10.1007/s10055-023-00926-5

Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., & Bustillo, A. (2023). Countering the novelty effect: A tutorial for immersive virtual reality learning environments. *Applied Sciences, 13*(1), 593.

Omelicheva, M. Y., & Avdeyeva, O. (2008). Teaching with lecture or debate? Testing the effectiveness of traditional versus active learning methods of instruction. *PS: Political Science & Politics, 41*(3), 603–607. https://doi.org/10.1017/S1049096508080815

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*, 785–797. https://doi.org/10.1037/edu0000241

Paxinou, E., Georgiou, M., Kakkos, V., Kalles, D., & Galani, L. (2022). Achieving educational goals in microscopy education by adopting virtual reality labs on top of face-to-face tutorials. *Research in Science & Technological Education, 40*(3), 320–339.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 1–29.

Ravichandran, R., & Mahapatra, J. (2023). Virtual reality in vocational education and training: Challenges and possibilities. *Journal of Digital Learning and Education, 3*(1), 25–31. https://doi.org/10.52562/jdle.v3i1.602

Renninger, K. A., & Hidi, S. E. (2016). *The power of interest for motivation and engagement*. New York: Routledge.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Schäfer, C., Rohse, D., Gittinger, M., & Wiesche, D. (2023). Virtual Reality in der Schule: Bedenken und Potenziale aus Sicht der Akteur: innen in interdisziplinären Ratingkonferenzen. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung, 51*, 1–24.

Schlüter, C., & Kretschmer, V. (2020). Next level training in logistics: Evaluation of a virtual reality-based serious game for warehouse logistics. *Proceedings of the 19th International Conference on modeling & applied simulation (MAS 2020)*. https://doi.org/10.46354/i3m.2020.mas.018

Shou, Y., & Olney, J. (2021). Attitudes toward risk and uncertainty: The role of subjective knowledge and affect. *Journal of Behavioral Decision Making, 34*(3), 393–404. https://onlinelibrary.wiley.com/doi/10.1002/bdm.2217.

Sung, B., Mergelsberg, E., Teah, M., D'Silva, B., & Phau, I. (2021). The effectiveness of a marketing virtual reality learning simulation: A quantitative survey with psychophysiological measures. *British Journal of Educational Technology, 52*(1), 196–213.

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research & Development, 68*(1), 1–16.

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality, 16*(1), 33–48. https://doi.org/10.20870/IJVR.2016.16.1.2880

Tsay, C. H. H., Kofinas, A. K., Trivedi, S. K., & Yang, Y. (2020). Overcoming the novelty effect in online gamified learning systems: An empirical evaluation of student engagement and performance. *Journal of Computer Assisted Learning, 36*(2), 128–146.

Villena-Taranilla, R., Cózar-Gutiérrez, R., González-Calero, J. A., & López Cirugeda, I. (2019). Strolling through a city of the Roman Empire: An analysis of the potential of virtual reality to teach history in Primary Education. *Interactive Learning Environments, 30*(4), 608–618. https://doi.org/10.1080/10494820.2019.1674886

Villena-Taranilla, R., Tirado-Olivares, S., Cozar-Gutierrez, R., & Gonz′alez-Calero, J. A. (2022). Effects of virtual reality on learning outcomes in K-6 education: A metaanalysis. *Educational Research Review, 35*. https://doi.org/10.1016/j.edurev.2022.100434. Article 100434.

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzskala intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften ZfDN ; Biologie, Chemie, Physik, 15*, 31–45.

Won, M., Ungu, D. A. K., Matovu, H., Treagust, D. F., Tsai, C. C., Park, J., … Tasker, R. (2023). Diverse approaches to learning with immersive Virtual Reality identified from a systematic review. *Computers & Education, 195*, Article 104701.

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using headmounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology, 51*(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Zinn, B. (2019). Lehren und Lernen zwischen Virtualität und Realität. *Journal of Technical Education (JOTED), 7*(1).

5 Discussion and Outlook

This concluding chapter synthesizes the key findings and implications of the dissertation's three papers. Section 5.1 presents the main findings, Section 5.2 examines their practical implications, and Section 5.3 addresses both general and specific limitations and future research directions.

5.1 Summary of Findings

Papers 1, 2, and 3 investigated the effectiveness and impact of personalized learning in vocational education, with particular emphasis on digital learning prompts, simulation-based learning environments, and immersive virtual reality.

The first paper examined three key research questions: (1) Which types of prompts are distinguished throughout the literature? (2) What is the overall effect of prompts on learning achievement? (3) How is the effectiveness moderated by prompt features and study demographics? The investigation employed a two-stage analysis of 68 experimental studies, beginning with a qualitative literature analysis to identify prompt types and features, which were organized into a conceptual model (see Section 2.5.2), followed by a quantitative meta-analysis of effect sizes.

The overall analysis revealed that prompts have a positive, moderate effect on learning achievement ($d = .394$), indicating their potential as a valuable personalization tool. However, the study uniquely identified a significant publication bias, yielding a more conservative effect size estimate ($d = .22$) that indicates a smaller yet still positive impact.

Several key moderating variables emerged from the analysis. While no significant differences in effectiveness were observed among cognitive types of prompts, research on non-cognitive prompts remained limited ($n = 4$). Action-based prompts ($d = .447$) demonstrated greater effectiveness than time-based prompts ($d = .240$), suggesting that prompts triggered by learner actions yield better outcomes. Time-based prompts showed mixed results, with many studies reporting minimal effects and some indicating negative impacts. This variability may stem from cognitive load considerations, where poorly timed prompts can overwhelm learners, particularly those with limited prior knowledge (Sweller, 1988; Thillmann et al., 2009; Wang & Lajoie, 2023).

Personalization emerged as a crucial factor, with prompts tailored to groups of learners ($d = .513$) showing higher efficacy compared to prompts for all learners. This suggests that personalized prompts serve as an effective safety net, enabling learners to navigate complex digital learning environments while receiving targeted support when needed (Mead et al., 2019). Additionally, combining generic and directed prompts (d = .571) yielded the highest effect sizes, indicating that a balanced approach to prompt specificity is optimal.

Demographic and contextual factors significantly influenced prompt effectiveness. Notable regional differences emerged, with studies from East Asia showing the highest effect sizes (d = .862) compared to Europe (d = .278). The analysis revealed that prompts were most effective for high school students (d = .412) compared to higher education students (d = .320) and working adults (d = .071). With participants averaging 25 years of age, the research emphasized outcomes for younger learners. Different learning domains demonstrated varying effects, with social sciences (d = .479) and technological sciences (d = .432) showing the strongest outcomes. Prompts demonstrated similar effectiveness across immersive environments (d = .341), traditional settings (d = .364), and ITS (d = .384), with no significant differences (p = .08).

Building on the findings presented in Paper 1 (DBR Phase 0), the second paper focused only on applying the LUCA office simulation in vocational education, exploring how a personalized prompt design can create authentic, adaptive learning scenarios for future vocational education. These prompts were log data-based and adaptively integrated into the learning environment, triggering based on learners' specific actions or inactions as outlined in Section 2.6.2 and 3.2.2 and tested by three research scientists (DBR Phase 1).

During DBR Phase 2, four research assistants tested the working scenario, evaluating both the content and timing of prompts. Teacher feedback in free-text responses indicated uncertainty about the note-taking prompt implementation (Appendix Table A1, No. 4). Consequently, the event pop-up and associated prompts were reformulated with greater clarity and detail to provide enhanced learner guidance. Additionally, teachers suggested that prompt utilization required initial familiarization and recommended including an exemplar prompt before task commencement. In response, an onboarding video was incorporated demonstrating *Luca* office functionalities, particularly highlighting event and prompt mechanisms, aligning with the fifth rule of thumb established in Paper 1.

Initial empirical evidence from Phase 3 showed no significant differences between the experimental (prompt) and control groups (p = .51, d = .05). Most prompts appeared ineffective in facilitating successful solution adjustments, particularly those designed to assist with more complex tasks such as price calculations. Both questionnaire responses and free-text feedback indicated that the number of prompts (learners received an average of 12 out of 23 possible prompts) was excessive and perceived as partly disruptive to learning. This aligns with research (Thillmann et al., 2009; Wang & Lajoie, 2023) suggesting that an increased number of prompts can elevate workload and cause cognitive overload (Sweller, 1988), resulting in irritation or feelings of being overwhelmed. As detailed in Section 3.2.2, this feedback led to substantial changes in the prompt design, specifically a reduction in frequency and increasing detail.

Subsequent empirical evidence from the second data collection (DBR Phase 4) of Paper 2 (Thomann et al., 2024) demonstrated significant improvements in task performance between experimental and control groups (d = .34). These findings suggest that the refined prompts effectively enhanced learners' problem-solving competence, as reflected in overall scores. The reduced number of prompts (an average of 6 out of 18 available prompts) demonstrated higher compliance, with about 90% of prompts being accessed and recognized by learners. Notably, the reformulated and consolidated prompts for supplier price calculation demonstrated substantial positive effects on task performance (d = .47 - .50).

The third paper investigated three research questions regarding IVR in vocational education. First, it looked at the differences between IVR and paper-based learning in terms of objective and subjective knowledge acquisition. Second, it explored the relationship between objective and subjective knowledge acquisition in both settings. Third, it assessed the differences in mood, motivation, and immersion between IVR and paper-based learning.

Addressing the first question, the study found that IVR did not improve immediate declarative knowledge acquisition compared to traditional paper-based methods. Contrary to expectations, students using paper-based learning demonstrated better performance in objective knowledge tests. This finding challenges assumptions about the superiority of immersive technologies for short-term knowledge retention. Moreover, a notable discrepancy emerged as IVR users reported significantly higher perceived knowledge gains despite their lower actual performance.

For the second research question, the findings revealed a misalignment between perceived and actual learning in IVR environments. While paper-based learning showed a moderate correlation between perceived and actual knowledge gains, IVR learning demonstrated a weak, non-significant correlation between these measures, though the difference between correlations was not statistically significant. This discrepancy between objective and subjective learning outcomes highlights the complex nature of learning experiences in immersive virtual environments.

Regarding the third research question, IVR proved to be highly effective in enhancing the learning experience's affective and motivational dimensions. Students using IVR showed increased positive emotions, higher intrinsic motivation across multiple dimensions, and greater immersion in the learning process compared to those using traditional methods. These findings highlight a complex relationship between immersive technology, learning outcomes, and student engagement in vocational education, particularly demonstrating the distinction between perceived and actual learning gains in IVR environments.

Synthesizing findings across all three papers reveals several key insights about personalized learning in VET. First, the effectiveness of digital interventions appears to be enhanced through targeted personalization, as evidenced by the superior performance of action-based, group-tailored prompts in Paper 1 and the positive effects of personalized prompts in the LUCA office simulation in Paper 2. While Papers 1 and 2 demonstrated the value of digital support tools, Paper 3's investigation of IVR revealed a more nuanced picture. Despite increasing engagement and motivation, IVR showed lower immediate declarative knowledge gains compared to traditional methods. This pattern across papers suggests that successful implementation of personalized learning tools depends not on technological sophistication alone, but rather on careful alignment between learning objectives, learner needs, and implementation design.

5.2 Practical Implications

The findings from this dissertation yield practical implications for educators and students, software developers, and curriculum designers at the micro-, meso-, and macro-levels.

At the micro-level, the findings provide concrete guidance for daily teaching and learning practices. Paper 1 established five key rules of thumb (p. 23) for designing effective learning prompts. While prompts can significantly enhance learning achievement, their effectiveness depends on proper design and implementation, requiring careful consideration of multiple factors, including personalization, timing, cultural context, and learner characteristics (Bannert & Mengelkamp, 2013; van Alten et al., 2020). The modest effect size uncovered through publication bias analysis and various moderating variables should caution practitioners that simply using prompts is no silver bullet for enhancing learning achievement. Instead, prompts should be viewed as one component of a comprehensive, personalized learning approach rather than a standalone solution for enhancing learning achievement.

The design-based research process in Paper 2 required four iterative cycles to refine the prompt design, with Paper 1 serving as a blueprint for personalized prompts (see Section 3.2.2). Metacognitive prompts directing learners to access supplementary materials (such as documents, spreadsheets, and file notes) did not elicit the anticipated behavioral responses (e.g., Appendix Table A1, No. 6). Following Mayers' (2021) coherence principle (see Section 2.4.2), these prompts were largely eliminated, retaining only those essential for the most essential documents. Initial empirical evidence (Phase 4) identified cognitive prompts (Appendix Table A3, No. 6-12) as particularly effective for more complex calculations, especially when referencing specific examples in the manual (supplementary material). These findings indicate that prompts are more beneficial when supporting well-defined cognitive procedures with clear solution paths, rather than serving as standalone navigational aids. Moreover, simplicity enhances prompt compliance. Basic prompts, such as reminders for recording dates or product names, showed higher adherence rates. This suggests that, in the design of learning environments, practitioners should prioritize straightforward, easy-to-follow prompts for routine tasks while reserving detailed prompting strategies with solution examples for complex cognitive operations where additional support is crucial.

While IVR demonstrated significantly higher levels of motivation, positive mood, and immersion (Paper 3), heightened engagement did not automatically translate to improved learning outcomes. The weak correlation between perceived and actual learning in the IVR group suggests that students need to develop metacognitive strategies for accurately assessing their learning progress in immersive environments. Such strategies should include regular self-assessment checks, explicit reflection on learning objectives, and comparison of perceived learning with objective measures. Learners should also be prepared to engage with multiple learning modalities, as the research supports a combined approach of traditional (e.g., paper-based learning) and IVR-based learning (Meyer et al., 2019; Paxinou et al., 2022).

At the meso-level, the findings suggest several key considerations for software designers developing educational technologies. For the personalized prompt design in Paper 2, delivering prompts through a more sophisticated triggering mechanism would have been optimal. However, technical limitations in LUCA Office restricted this capability. As described in Section 2.6.2 (p. 31), only self-selected prompts were viable, necessitating that learners actively click on email prompts to view and read them. While this approach minimized learning interruptions, it relied heavily on learner initiative. Initial empirical data from the first wave of data collection indicated that learners occasionally overlooked prompts entirely or failed to adjust their solutions as needed. A proposed enhancement involves implementing a pop-up window system (similar to EES) that activates with clear instructions if email prompts remain unopened or if solutions stay unchanged after a specified duration.

The findings on cognitive load management and adaptive support can be effectively integrated through the FAIRI framework (Obourdin et al., 2024) for IVR software design. By incorporating Mayer's multimedia principles (2021) alongside ITS, developers can create learning environments that balance engagement with focused learning experiences (Makransky, et al., 2019). The FAIRI model's structured approach to adaptive support addresses the need to manage cognitive load while accommodating varying levels of user experience with VR technology (Obourdin et al., 2024). Software designers can implement adaptive support features that provide prompts based on individual learner needs while maintaining immersion through naturally embedded support mechanisms. Additionally, the integration of built-in assessment tools within the FAIRI framework can help bridge the observed gap between perceived and actual learning gains by enabling accurate measurement of learning outcomes while providing personalized learning support.

At the macro-level, the findings suggest pathways for curriculum modification. The tailored prompt design employed in *Luca* office effectively addresses varied student needs while enhancing problem-solving abilities, supported by empirical analysis from Paper 2 (see Rausch et al., 2021; Thomann et al., 2024). Curriculum designers must systematically integrate IVR into vocational education programs, considering Paper 3's findings that IVR might be better suited for procedural rather than declarative knowledge acquisition.

5.3 Limitations and Research Outlook

While providing valuable insights, the studies on personalized learning in vocational education are subject to several limitations. This section discusses these limitations and outlines future research directions to address them, connecting the themes across all three papers.

5.3.1 Overarching Limitations and Outlook

The first major limitation concerns sample characteristics and generalizability. The research primarily focused on the German vocational education system, particularly within warehouse logistics and commercial sectors, while the meta-analysis studies were predominantly conducted in Western educational contexts such as Europe and North America, with minimal representation from regions like East Asia. This geographical and systemic focus, while enabling detailed analysis, potentially limits the findings' applicability to other vocational domains and educational systems. Demographic considerations further constrain generalizability. The relatively young age of participants, averaging around 20 years across the three papers, raises questions about the findings' relevance to adult learners and continuing education contexts. Although this age group represents the primary demographic for initial vocational training, the effectiveness of digital interventions may vary significantly for older learners who typically have different levels of technology acceptance and learning preferences. Moreover, while the studies covered various vocational areas such as commercial management and logistics, they may not fully represent the broad spectrum of vocational education fields, each of which may have unique characteristics affecting the success of personalized learning approaches.

The second major limitation concerns the temporal scope of the research, specifically the predominant focus on short-term effects. The studies presented relied on immediate post-intervention measurements, lacking exploration of long-term retention and transfer effects. This limitation is particularly evident in Paper 3, where the effectiveness of IVR was assessed solely through immediate post-test measurements. However, some exceptions exist, such as studies by Chittaro and Buttussi (2015) and Meyer and Pfeiffer (2020), which found positive long-term effects for IVR groups compared to paper-based groups. This aligns with findings from Paper 1, where the majority of analyzed studies focused on the immediate effects of prompts, with only a few studies implementing follow-up measurements to assess sustainable learning outcomes. Studies by Bannert et al. (2015) and Engelmann et al. (2021) stand out as rare exceptions that examined long-term effects of prompting interventions. Similarly, the empirical testing of the personalized prompt design of Paper 2 (beyond this dissertation) focused on problem-solving skills during the learning task without examining their sustained impact on learning behaviors and knowledge retention. This methodologically practical yet limited timeframe raises questions about the sustainable impact of digital learning interventions on the development of vocational competence.

To overcome these limitations, future research should focus on two main areas. First, it should conduct cross-cultural and cross-domain studies to examine how PL approaches vary across different contexts and populations. Second, it should implement longitudinal studies with multiple assessment points to analyze how the effectiveness of personalized approaches evolves over time and whether initial novelty effects diminish.

5.3.2 Specific Limitations and Outlook

Paper 1's analysis focused exclusively on learning achievement outcomes, overlooking potential emotional and motivational impacts of prompts. The variability and underdetermination of test formats across studies complicated comparisons, as many studies provided incomplete details on test characteristics, length, response types, and scoring rubrics. The notable heterogeneity across included studies—encompassing diverse methodologies, contexts, and prompt implementations—further complicates result interpretation. Additionally, the analysis could not account for learners' prompt compliance, a factor likely influencing prompt effectiveness on learning outcomes. The analysis of cognitive prompt types was constrained by the limited number of studies that included all prompt types, restricting the generalizability of conclusions regarding differences between categories. While the analysis detected a publication bias related to the overall effect, which likely extends to the moderator models, a Bonferroni correction was omitted to preserve statistical power, following recommendations by Polanin and Pigott (2015) and Schmidt and Hunter (2014).

The empirical analysis of Paper 2 (beyond this dissertation) will explore the research question: "What learner behaviors indicate prompt effectiveness?" by observing changes in behavior after the prompt, such as modifications in spreadsheets and document access. Analyzing metrics like prompt openings, viewing duration, and efforts to correct responses offers an unmatched depth of interaction analysis in the prompt literature. However, this analysis relies solely on log data. Future research should incorporate additional data types, particularly eye-tracking protocols (see Gorshid et al., 2022). Integrating log and eye-tracking data would enable a more robust interpretation of prompt effectiveness and interaction strategies, without disrupting the learning process as think-aloud protocols do. Eye-tracking could reveal whether learners notice email prompt notifications, actively defer engagement, or miss prompts entirely. This data would illuminate prompt timing perception and reading decisions. Additionally, eye-tracking would show fixation patterns during prompt reading and capture nuanced problem-solving behavioral changes beyond basic log data of document access and spreadsheet modifications.

The number of provided prompts and events emerged as a critical consideration in the personalized prompt design of LUCA office. Several researchers (see Lim et al., 2024; Schumacher & Ifenthaler, 2021; Sonnenberg & Bannert, 2016; Van der Graaf et al., 2023; Wang & Lajoie, 2023) have raised important questions about the optimal number of prompts and whether varying quantities might potentially disrupt rather than support the learning process. Future research should systematically investigate the relationship between prompt frequency and learning performance through multiple methodological approaches. For instance, experimental studies comparing different prompt quantities, think-aloud protocols to analyze students' cognitive processes and perceptions, and the inclusion of cognitive load measurements to evaluate the mental demands of prompt interaction.

The limitations of Paper 3 revealed crucial considerations for advancing IVR research in vocational education. A fundamental methodological limitation stems from the media comparison approach. Despite employing a rigorous experimental design with randomization and a substantial dataset (n=72), the study encounters inherent issues with media comparison studies (Buchner & Kerres, 2023). The study's novel examination of relationships between objective and subjectively perceived knowledge gains in IVR, previously only investigated by Huang et al. (2023), adds valuable insights while working within these methodological constraints. Media comparison studies inherently assume direct technological influence on learning while overlooking instructional complexity (Clark, 1983, 1994; Hastings & Tracey, 2005). As established in the Clark (1983, 1994) and Kozma debate (1991, 1994), media primarily serve as vehicles for delivering instruction rather than causal factors in learning outcomes. Even with careful experimental design in Paper 3, controlling conditions between groups proves challenging (Buchner & Kerres, 2023).

The experimental design faced interconnected challenges affecting internal validity. Unequal familiarity with IVR technology likely creates different cognitive loads that potentially confound results (see Chang, 2021; Miguel-Alonso et al., 2024). The assessment structure introduced further complications through its narrow focus on declarative knowledge, potentially understating IVR's broader educational value, particularly for procedural skills crucial in vocational education. Time constraints within the vocational setting affected test instrument reliability, though the tests comprehensively covered simulation content following formative measurement principles (Coltman et al., 2008). This limitation echoes challenges faced in similar research (Kolarik et al., 2024; Lee, 2020). The unequal feedback structure between groups - with IVR learners receiving a concluding dashboard while paper-based learners did not - further complicated the result interpretation.

External validity faces constraints through temporal and contextual limitations. The short-term intervention complicates distinguishing between genuine learning effects and novelty-driven engagement (Allcoat & von Mühlenen, 2018). Single-intervention results, regardless of sample size or randomization, may not replicate across contexts due to learning environments' situational nature. Teachers typically employ various media and methods across multiple lessons rather than relying on single approaches (Buchner & Kerres, 2023). Thus, comparative data between IVR and paper-based learning in isolated sessions provides limited actionable insights for practitioners (Honebein & Reigeluth, 2021). Additionally, the sample's pronounced gender imbalance (over 80% male), while reflecting the logistics sector's composition, limits conclusions about gender-specific effects in IVR learning environments (Makransky & Petersen, 2021).

These limitations inform several integrated research directions. First, future research should build upon the methodological strengths demonstrated in Paper 3 while moving beyond simple media comparisons. Value-added or learner-treatment-interaction designs could better inform IVR implementation in vocational education by investigating specific instructional features or learner characteristics that enhance effectiveness (Buchner & Kerres, 2023). This evolution connects directly to examining feedback mechanisms in IVR environments. Future studies could investigate how different types and timing of feedback affect learning outcomes, comparing immediate versus delayed feedback and exploring feedback modalities unique to IVR environments. This priority is demonstrated by planned research comparing feedback variations in IVR presentation skills training. Furthermore, research indicates that incorporating traditional learning methods before IVR experiences can enhance learner capabilities (Meyer et al., 2019; Paxinou et al., 2022), suggesting the need for integrated pedagogical approaches.

Longitudinal studies should examine sustained learning effects beyond novelty periods, tracking both cognitive and affective outcomes to understand how initial technology enthusiasm translates to sustained engagement (Conrad et al., 2022; Kolarik et al., 2024). Assessment methodology requires development of instruments measuring both immediate and long-term outcomes while accommodating practical constraints. Gender-specific learning patterns and technology acceptance merit particular attention, especially in male-dominated vocational fields. Research should also expand beyond single-task scenarios to explore IVR's effectiveness across diverse vocational tasks, incorporating both declarative and procedural knowledge assessments (see Conrad et al., 2024).

These research directions collectively aim to provide more nuanced, practical insights for effectively integrating IVR into vocational education while addressing the methodological challenges identified in current research. Moving beyond comparative studies toward more sophisticated research designs will help educators understand not only whether IVR should be integrated into vocational education and training context, but also how to do so effectively (Buchner & Kerres, 2023).

5.3.3 Overarching Research Directions

The findings from this dissertation's studies, along with their limitations and implications, contribute to our understanding of personalized learning in vocational education while simultaneously highlighting the vast territory yet to be explored. Returning to the elephant metaphor introduced in Section 2, this research has illuminated specific aspects of the larger phenomenon of (digital) personalized learning - namely, personalized prompting and immersive virtual reality environments. Just as the blind men in the parable each touched different parts of the elephant, these studies have provided valuable insights into particular facets of (digital) personalized learning, while acknowledging the broader complexity of the field as illustrated in Bernacki et al.'s (2021) framework.

The personalized learning design process model presented in Section 2 revealed the intricate interplay between learner characteristics, learning environments, and desired outcomes. While this dissertation focused on specific implementations - digital prompts and IVR environments - these represent only a subset of the possible personalization approaches within the broader framework. This selective focus, while necessary for deep empirical investigation, mirrors the parable's lesson about partial perspectives contributing to a larger truth. The studies in this dissertation have provided detailed insights into how certain personalization methods can enhance vocational learning, yet they also point to the need for more comprehensive theoretical frameworks that can accommodate the full spectrum of personalization approaches.

This recognition of both specific insights and broader contexts leads us to several critical directions for future research. First, there is a need to advance personalized learning theory in ways that can bridge the gap between individual implementation studies and overarching theoretical frameworks. Such advancement requires accounting for the complex interactions between learner characteristics, learning content, and personalization methods - relationships that were only partially explored in the current studies but are fundamental to the comprehensive framework outlined by Bernacki et al. (2021).

Building on this foundation, the research directions can be organized into three interconnected domains, each representing different aspects of the broader personalized learning landscape. The first domain focuses on the integration of personalized learning technologies, reflecting the need to move beyond isolated implementations toward more comprehensive learning ecosystems. Future research should explore how different personalized learning approaches - such as prompts, simulations, and IVR - can be meaningfully combined. The FAIRI Framework (Obourdin et al., 2024) mentioned in this dissertation offers one potential pathway for such integration, particularly in combining personalized prompts with IVR simulations.

The second domain concerns the critical alignment with industry needs in vocational education. This alignment requires a sophisticated understanding of how different learner characteristics and environmental adaptations, as outlined in Bernacki's model (2021), translate to workplace performance. Future research should emphasize collaborative studies between educational institutions and industry partners, examining how personalized learning experiences in VET contribute to workplace success. The development of open educational resources such as the LUCA office and collaborative development efforts with the *InGo* IVR simulation make advanced personalized learning tools more accessible to a wider range of vocational education providers.

The third domain addresses the practical challenges of scalability and cost-effectiveness (see Ravichandran & Mahapatra, 2023). This includes exploring how artificial intelligence might enhance the dynamic adaptation capabilities described in Bernacki's framework (2021), particularly in creating responsive learning systems that can address multiple learner characteristics simultaneously. Drawing from Dillenbourg's (2013) orchestration framework, future research must investigate how IVR implementation in VET can balance intrinsic learning activities with extrinsic constraints while minimizing "orchestration load" - the cognitive burden on teachers managing multiple activities and constraints simultaneously. This involves examining cost-benefit relationships, considering alternative implementations, and ensuring the accessibility of learning materials within resource-constrained educational environments. Considering this, sufficient training for teachers is also necessary for using IVR (Ravichandran & Mahapatra, 2023; Scheiter, 2021) and computer-based simulations like LUCA office (see Paper 2, p. 116).

In conclusion, advancing our understanding of personalized learning in vocational education requires a delicate balance between detailed investigation of specific approaches and integration of these findings into broader theoretical frameworks. Future research must embrace both the complexity revealed by Bernacki's model (2021) and the practical demands of educational implementation, always keeping the diverse needs of vocational learners at the forefront. Like the blind men in the parable working together to understand the elephant, researchers, educators, and industry partners must collaborate to develop a more complete and nuanced understanding of how to implement personalized learning in vocational education effectively.

6 References

Alamri, H. A., Watson, S., & Watson, W. (2021). Learning technology models that support personalization within blended learning environments in higher education. *Tech Trends*, 65, 62–78. https://doi. org/10.1007/s11528-020-00530-3

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). *Instruction based on adaptive learning technologies.* Handbook of research on learning and instruction, 522-560.

Allcoat, D., & von Mühlenen, A. (2018). Learning in virtual reality: Effects on performance, emotion and engagement. *Research in Learning Technology*, 26.

Álvarez, R. P., Jivet, I., Pérez-Sanagustín, M., Scheffel, M., & Verbert, K. (2022). Tools designed to support self-regulated learning in Online Learning environments: A systematic review. *IEEE Transactions on Learning Technologies*, 15(4), 508–522. https://doi.org/10.1109/TLT.2022.3193271

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, 33, 381–412.

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo, & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427–449). Springer. https://doi.org/10.1007/978-1-4419-5546-3_28.

Bandura, A. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986(23-28), 2.

Bannert, M. (2009). Promoting self-regulated learning through prompts: A discussion. *Zeitschrift für Pädagogische Psychologie*, 22(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. *In International handbook of metacognition and learning technologies* (pp. 171-186). New York, NY: Springer New York.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Beck, K., Landenberger, M., & Oser, F. (2016). Technologiebasierte Kompetenzmessung in der beruflichen Bildung. *Ergebnisse aus der BMBF-Förderinitiative ASCOT. Bielefeld.*

Beese, E. B. (2019). A process perspective on research and design issues in educational personalization. *Theory and Research in Education*, 17(3), 253–279.

Bernacki, M. L. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. *In Handbook of self-regulation of learning and performance* (2nd ed., pp. 370–387). Routledge/Taylor & Francis Group.

Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A Systematic Review of Research on Personalized Learning: Personalized by Whom, to What, How, and for What Purpose(s)? *Educational Psychology Review*, *33*(4), 1675–1715. https://doi.org/10.1007/s10648-021-09615-8

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer. https://doi.org/10.1007/978-94-007-2324-5_2.

Block, J. H., & Burns, R. B. (1976). Mastery learning. *Review of Research in Education*, 4, 3–49.

Blumberg, P. (1981). A practical methodology for developing content parallel multiple-choice tests. *The Journal of Experimental Education*, 50(2), 56–63.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Braunstein, A., Deutscher, V., Seifried, J., Winther, E., & Rausch, A. (2022). A taxonomy of social embedding—A systematic review of virtual learning simulations in vocational and professional learning. *Studies in Educational Evaluation*, *72*, 101098. https://doi.org/10.1016/j.stueduc.2021.101098

Buchner, J., & Kerres, M. (2023). Media comparison studies dominate comparative research on augmented reality in education. *Computers & Education*, *195*, 104711. https://doi.org/10.1016/j.compedu.2022.104711

Buchner, J., & Mulders, M. (2020). Lernen in immersiven virtuellen Welten aus der Perspektive der Mediendidaktik. *Medienimpulse*, 58(2).

Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., & Subramaniam, A. (2018). Skill shift automation and the future of the workforce. Discussion Papers. *McKinsey Global Institute*, 1(2018), 3-84. Retrieved October, 10, 2024 from https://www.mckinsey.com/featured-insights/future-ofwork/skill-shift-automation-and-the-future-of-the-workforce.

Bulger, M. (2016). Personalized learning: The conversations we're not having. *Data and Society*, 22(1), 1-29.

Cedefop (2023). *Teaching professionals: skills opportunities and challenges. Skills intelligence data insight.* Retrieved November, 11, 2024 from https://www.cedefop.europa.eu/en/data-insights/teaching-professionals-skills-opportunities-and-challenges-2023-update#_looking_forward

Cedefop and OECD (2024). *Apprenticeships and the digital transition: modernising apprenticeships to meet digital skill needs.* Publications Office of the European Union. Cedefop reference series; 125. Retrieved November, 04, 2024 from https://data.europa.eu/doi/10.2801/074640

Chang, Y. (2021). Effects of virtual reality application on skill learning for optical-fibre fusion splicing. *British Journal of Educational Technology*, 52(6), 2209–2226. https://doi.org/10.1111/bjet.13118

Chen, O., Paas, F., & Sweller, J. (2023). A Cognitive Load Theory Approach to Defining and Measuring Task Complexity Through Element Interactivity. *Educational Psychology Review*, *35*(2), 63. https://doi.org/10.1007/s10648-023-09782-w

Chernikova, O., Heitzmann, N., Holzberger, D., Neuhaus, B., Kron, S., Ufer, S., & Fischer, F. (2024). Fostering Diagnostic Skills with Simulations: A Research Synthesis of 237 Studies in Higher Education. In Lindgren, R., Asino, T. I., Kyza, E. A., Looi, C. K., Keifert, D. T., & Suárez, E. (Eds.), *Proceedings of the 18th International Conference of the Learning Sciences - ICLS 2024* (pp. 1291-1294). International Society of the Learning Sciences.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, 90(4), 499–541. https://doi.org/10.3102/0034654320933544

Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T., & Fischer, F. (2023). Where salience goes beyond authenticity: A meta-analysis on simulation-based learning in higher education. *Zeitschrift Für Pädagogische Psychologie*, 38(1–2), 15–25. https://doi.org/10.1024/1010-0652/a000357

Cheung, S. K., Wang, F. L., Kwok, L. F., & Poulova, P. (2021). In search of the good practices of personalized. *Interactive Learning Environments*, 29(2), 179–181. https://doi.org/10.1080/10494820.2021.1894830

Chittaro, L., & Buttussi, F. (2015). Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Clariana, R. B. & Hooper, S. (2012). Adaptive evaluation systems. In N. M. Seel (Ed.), *Encyclopaedia of the Sciences of Learning* (pp.104–106). Springer.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research & Development*, 42(2), 21–29.

Clark, R. E., & Hannafin, M. J. (2012). Debate about the benefits of different levels of instructional guidance. In R. A. Reiser & J. V. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed., pp. 367–383). Allyn & Bacon.

Cognition and Technology Group at Vanderbilt (1997). *The Jasper Project: Lessons in curriculum, instruction, assessment, and professional development.* Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Collins, A. (1992). Towards a design science of education. In E. Scanlon, & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Lawrence Erlbaum Associates, Inc.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12), 1250–1262.

Conrad, M., Dölker, J., Kablitz, D., & Schumann, S. (2022). VR in der kaufmännischen Berufsbildung: Potenziale–Befunde–Perspektiven. In S. Schumann, S. Seeber, & S. Abele Hrsg (Eds.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (pp. 231–255). Bielefeld. https://doi.org/10.3278/9783763971381.

Conrad, M., Kablitz, D., & Schumann, S. (2024). Learning effectiveness of immersive virtual reality in education and training: A systematic review of findings. *Computers & Education: X Reality*, *4*, 100053. https://doi.org/10.1016/j.cexr.2024.100053

Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the efect of immersive technology on user presence. *Media Psychology*, 19(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740

Davis, E. A. (2003). Prompting Middle School Science Students for Productive Reflection: Generic and Directed Prompts. *Journal of the Learning Sciences*, *12*(1), 91–142. https://doi.org/10.1207/S15327809JLS1201_4

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268.

Deutscher, V. und Winther, E. (2018). Instructional sensitivity in vocational education. *Learning and Instruction*, 53, 21–33.

Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. https://doi.org/10.1016/j.compedu.2013.04.013

Dillenbourg, P., Schneider, D., & Synteta, P. (2002). Virtual learning environments. In *3rd Hellenic Conference" Information & Communication Technologies in Education"* (pp. 3-18). Kastaniotis Editions, Greece.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, 101859.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, 34(3), 169–189.

Engelmann, K., Bannert, M., & Melzner, N. (2021). Do self-created metacognitive prompts promote short- and long-term effects in computer-based learning environments? *Research and Practice in Technology Enhanced Learning*, *16*(1), 3. https://doi.org/10.1186/s41039-021-00148-w

Errington, E. P. (2011). Mission possible: Using near-world scenarios to prepare graduates for the professions. *International Journal of Teaching and Learning in Higher Education*, 23(1), 84-91.

Errington, E. P. (Ed.) (2010). *Preparing graduates for the professions using scenario-based learning.* Brisbane: Post Pressed.

Fokides, E., & Antonopoulos, P. (2024). Development and testing of a model for explaining learning and learning-related factors in immersive virtual reality. *Computers & Education: X Reality,* 4, Article 100048.

Frank, I., & Schreiber, D. (2006). Bildungsstandards – Herausforderungen für das duale System [Educational standards – challenges for the dual system]. *Berufsbildung in Wissenschaft und Praxis*, 35(4), 6–10.

Fraulini, N. W., Marraffino, M. D., Garibaldi, A. E., Johnson, C. I., & Whitmer, D. E. (2024). Adaptive training instructional interventions: A meta-analysis. *Military Psychology*. https://doi.org/10.1080/08995605.2024.2377884

Funke, J. (2003). *Problemlösendes denken*. Kohlhammer Verlag.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: Problem solving in the twenty-first century. *Assessment and teaching of 21st century skills: Research and applications*, 41-53.

Georgiou, Y., & Kyza, E. A. (2017). The development and validation of the ARI questionnaire: An instrument for measuring immersion in location-based augmented reality settings. International *Journal of Human-Computer Studies*, 98, 24–37. https://doi.org/10.1016/j.ijhcs.2016.09.014

Gorshid, G. D., Mayer, C., Rausch, A. & Seifried, J. (2022). Das LUCA-Dashboard im Usability-Test – Eine gaze-cued retrospective Think-Aloud-Studie. In S. Schumann, S. Seeber, S. Abele (Eds.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (Band 41, S. 189–212). Bielefeld: wbv.

Green, T. D., & Donovan, L. C. (2018). Learning anytime, anywhere through technology. *The Wiley handbook of teaching and learning* (pp. 225–256).

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, *16*(1), 94-143.

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12650

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition Learning*, 2, 107–124. https://doi.org/10.1007/s11409-007-9016-7, 2007.

Harteis, C., & Billett, S. (2023). Knowledge and Learning at the Workplace in Times of Digital Transformation. In K. Evans, W. O. Lee, J. Markowitsch, & M. Zukas (Eds.), *Third International Handbook of Lifelong Learning* (S. 163–182). Springer International Publishing. https://doi.org/10.1007/978-3-031-19592-1_4

Hastings, N. B., & Tracey, M. W. (2005). Does media affect learning: Where are we now? *TechTrends*, 49(2), 28–30. https://doi.org/10.1007/BF02773968

Hawlitschek, A., & Joeckel, S. (2017). Increasing the effectiveness of digital educational games: The effects of a learning instruction on students' learning, motivation and cognitive load. *Computers in Human Behavior*, 72, 79-86.

Heinrichs, K., & Reinke, H. (2019). *Heterogenität in der beruflichen Bildung: im Spannungsfeld von Erziehung, Förderung und Fachausbildung*. wbv Media.

Hellriegel, J., & Cubela, D. (2018). Das Potenzial von Virtual Reality für den schulischen Unterricht - Eine konstruktivistische Sicht. *MedienPäadagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 58–80. https://doi.org/10.21240/mpaed/00/2018.12.11.X

Hidi, S., & Renninger, K. (2006). The four-phase model of interest development. *Educational Psychologist*, 41(2), 111–127.

Higgins, J. P., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470712184.fmatter

Holmes, W., Anastopoulou, S., Schaumburg, H., & Mavrikis, M. (2018). *Technology-enhanced Personalised Learning: Untangling the Evidence.* Robert Bosch Stiftung GmbH, Stuttgart. Retrieved March, 27, 2023 from http://www.studie-personlisiertes-lernen.de/en.

Honebein, P. C., & Reigeluth, C. M. (2021a). Making good design judgments via the instructional theory framework. In J. K. McDonald & R. E. West (Eds.), *Design for learning: Principles, processes, and praxis*. EdTech Books.

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

Huang, X., Zhao, Q., Liu, Y., Harris, D., & Shawler, M. (2023). Learning in an immersive VR environment: Role of learner Characteristics and relations between learning and psychological outcomes. *Journal of Educational Technology Systems*, Article 00472395231216943. https://doi.org/10.1177/00472395231216943

Hwang, G.-J., Xie, H., Wah, B. W., & Gasevic, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 1(2020), 100001.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society*, 15(1)

Järvelä, S., & Bannert, M. (2021). Temporal and adaptive processes of regulated learning—what can multimodal data tell? *Learning and Instruction*, 72, Article 101268. https://doi.org/10.1016/j.learninstruc.2019.101268

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23,

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63-85. https://doi.org/10.1007/BF02300500

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, 2 (2021), 100017.

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

Kalyuga, S. (2011). Cognitive Load Theory: How Many Types of Load Does It Really Need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kerr, P. (2016). Adaptive learning. *Elt Journal*, 70(1), 88-93, https://doi.org/10.1093/elt/ccv055

Klotz, V. K. (2015). *Diagnostik beruflicher Kompetenzentwicklung. Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne.* Diagnosis of professional development. A didactic modeling for the commercial domain. Wiesbaden: Springer.

Klotz, V. K., & Winther, E. (2016). Zur Entwicklung domänenverbundener und domänenspezifischer Kompetenz im Ausbildungsverlauf: Eine Analyse für die kaufmännische Domäne. *Zeitschrift für Erziehungswissenschaft*, 19(4), 765–782. https://doi.org/10.1007/s11618-016-0712-7

Kolarik, S., Schlüter, C., & Ziolkowski, K. (2024). Impact of VR on learning experience compared to a paper based approach. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 12.

Komalawardhana, N., & Panjaburee, P. (2024). Trends and development of technology-enhanced personalized learning in science education: a systematic review of publications from 2010 to 2022. *Journal of Computers in Education*, 11(3), 721-742. https://doi.org/10.1007/s40692-023-00276-w

Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61(2), 179–211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research & Development*, 42(2), 7–19. https://doi.org/10.1007/BF02299087

Kremer, H. H., Beutner, M., & Zoyke, A. (2012). *Individuelle Förderung und berufliche Orientierung im berufsschulischen Übergangssystem.* Ergebnisse aus dem Forschungs-und Entwicklungsprojekt InLab. Paderborn: Eusl (2012), S. 212.

Lajoie, S. P. (2005). Extending the scaffolding metaphor. *Instructional Science*, 33(5–6), 541–557. https://doi.org/10.1007/s11251-005-1279-2

Langone, J. (1998). The Effects of Technology-Enhanced Anchored Instruction and Situated Learning on Preservice Teachers in a Special Education Methods Course: An Exploratory Study. *Journal of Developmental and Physical Disabilities.* 10, 35–54. https://doi.org/10.1023/A:1022809500853

Lee, E. A.-L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, 55(4), 1424–1442. https://doi.org/10.1016/j.compedu.2010.06.006

Lee, I.-J. (2020). Applying virtual reality for learning woodworking in the vocational training of batch wood furniture production. *Interactive Learning Environments*, 31 (3), 1448–1466. https://doi.org/10.1080/10494820.2020.1841799

Lepper, M. R., Drake, M. F., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), *Advances in learning & teaching. Scaffolding student learning: Instructional approaches and issues* (pp. 108–144). Brookline Books.

Li, K. C., & Wong, B. T. M. (2021). Features and trends of personalised learning: A review of journal publications from 2001 to 2018. *Interactive Learning Environments*, 29(2), 182–195. https://doi.org/ 10.1080/10494820.2020.1811735

Li, K. C., & Wong, B. T. M. (2023). Personalisation in STE(A)M education: A review of literature from 2011 to 2020. *Journal of Computing in Higher Education*, 35, 186–201. https://doi.org/10.1007/ s12528-022-09341-2

Li, T., Fan, Y., Tan, Y., Wang, Y., Singh, S., Li, X., Raković, M., Van Der Graaf, J., Lim, L., Yang, B., Molenaar, I., Bannert, M., Moore, J., Swiecki, Z., Tsai, Y.-S., Shaffer, D. W., & Gašević, D. (2023). Analytics of self-regulated learning scaffolding: Effects on learning processes. *Frontiers in Psychology*, *14*, 1206696. https://doi.org/10.3389/fpsyg.2023.1206696

Lim, L., Bannert, M., Van Der Graaf, J., Fan, Y., Rakovic, M., Singh, S., Molenaar, I., & Gašević, D. (2024). How do students learn with real-time personalized scaffolds? *British Journal of Educational Technology*, *55*(4), 1309–1327. https://doi.org/10.1111/bjet.13414

Lim, L., Bannert, M., Van Der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Lin, L., Lin, X., Zhang, X., & Ginns, P. (2024). The Personalized Learning by Interest Effect on Interest, Cognitive Load, Retention, and Transfer: A Meta-Analysis. *Educational Psychology Review*, *36*(3), 88. https://doi.org/10.1007/s10648-024-09933-7

Liu, H., Zhang, Y., & Jia, J. (2024). The Design of Guiding and Adaptive Prompts for Intelligent Tutoring Systems and Its Effect on Students' Mathematics Learning. *IEEE Transactions on Learning Technologies*, *17*, 1379–1389. https://doi.org/10.1109/TLT.2024.3382000

Ludwig, S., Rausch, A., Deutscher, V., & Seifried, J. (2024). Predicting problem-solving success in an office simulation applying N-grams and a random forest to behavioral process data. *Computers & Education*, *218*, 105093. https://doi.org/10.1016/j.compedu.2024.105093

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., et al. (2021). The future of work after COVID-19. Full Report. *McKinsey Global Institute.* Retrieved May, 28, 2024 from https://www.mckinsey.com/featured-insights/future-of-work/the-futureof-work-after-covid-19.

Mackinnon, A., Jorm, A. F., Christensen, H., Korten, A. E., Jacomb, P. A., & Rodgers, B. (1999). A short form of the positive and negative affect schedule: Evaluation of factorial validity and invariance across demographic variables in a community sample. *Personality and Individual Differences*, 27(3), 405–416. https://doi.org/ 10.1016/S0191-8869(98)00251-7

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, *3*, 100080. https://doi.org/10.1016/j.caeai.2022.100080

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low-and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, 52(5), 1935–1964. https://doi.org/10.1111/bjet.13116

Makransky G. (2021) The Immersion Principle in Multimedia Learning. In: Mayer RE, Fiorella L, eds. The Cambridge Handbook of Multimedia Learning. *Cambridge Handbooks in Psychology.* Cambridge University Press; 296-303.

Makransky, G., & Klingenberg, S. (2022). Virtual reality enhances safety training in the maritime industry: An organizational training experiment with a non-WEIRD sample. *Journal of Computer Assisted Learning*, 38(4), 1127–1140. https://doi.org/10.1111/jcal.12670

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research & Development*, 66(5), 1141–1164. https://doi.org/10.1007/s11423-018-9581-2

Makransky, G., & Petersen, G. B. (2021). The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review*, *33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Makransky, G., Andreasen, N. K., Baceviciute, S., & Mayer, R. E. (2021). Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. *Journal of Educational Psychology*, *113*(4), 719–735. https://doi.org/10.1037/edu0000473

Makransky, G., Mayer, R.E. (2022) Benefits of Taking a Virtual Field Trip in Immersive Virtual Reality: Evidence for the Immersion Principle in Multimedia Learning. *Educ Psychol Rev* **34**, 1771–1798. https://doi.org/10.1007/s10648-022-09675-4

Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 60, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007

Marougkas, A., Troussas, C., Krouska, A., & Sgouropoulou, C. (2023). Virtual Reality in Education: A Review of Learning Theories, Approaches and Methodologies for the Last Decade. *Electronics*, *12*(13), 2832. https://doi.org/10.3390/electronics12132832

Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Educational Technology Research and Development*, 68, 1903–1929. https://doi.org/10.1007/s11423-020-09793-2

Marzano, G., Abuze, A., & Nur, Y. (2021). Improving Adaptive Learning in a Smart Learning Environment. *ENVIRONMENT. TECHNOLOGIES. RESOURCES. Proceedings of the International Scientific and Practical Conference*, *2*, 93–99. https://doi.org/10.17770/etr2021vol2.6509

Matovu, H., Ungu, D. A. K., Won, M., Tsai, C.-C., Treagust, D. F., Mocerino, M., & Tasker, R. (2023). Immersive virtual reality for science learning: Design, implementation, and evaluation. *Studies in Science Education*, 59(2), 205–244. https://doi.org/10.1080/03057267.2022.2082680

Mayer, R. E. (2001). *Multimedia learning.* Cambridge University Press.

Mayer, R. E. (2021). *The multimedia principle*. In R. E. Mayer, & L. Fiorella (Eds.), The Cambridge handbook of multimedia learning (3rd ed., pp. 296–303). Cambridge University Press.

Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, *36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education*, *67*(2), 131–142. https://doi.org/10.1080/10899995.2019.1565285

Meyer, L., & Pfeiffer, T. (2020). Comparing virtual reality and screen-based training simulations in terms of learning and recalling declarative knowledge. *In Delfi 2020 – Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.* (pp. 55–66). Bonn. Gesellschaft für Informatik e.V.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, 140. https://doi.org/10.1016/j.compedu.2019.103603

Miguel-Alonso, I., Checa, D., Guillen-Sanz, H., & Bustillo, A. (2024). Evaluation of the novelty effect in immersive Virtual Reality learning experiences. *Virtual Reality*, 28(1), 27. https://doi.org/10.1007/s10055-023-00926-5

Miller, C. J., & Bernacki, M. L. (2019). Training preparatory mathematics students to be high ability self-regulators: Comparative and case-study analyses of impact on learning behavior and achievement. *High Ability Studies*, 30(1–2), 167–197. https://doi.org/10.1080/13598139.2019.1568829

Mislevy, R. J., & Riconscente, M. M. (2005). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Erlbaum.

Moos, D. C., & Bonde, C. (2016). Flipping the Classroom: Embedding Self-Regulated Learning Prompts in Videos. *Technology, Knowledge and Learning*, *21*(2), 225–242. https://doi.org/10.1007/s10758-015-9269-1

Mulders, M. (2022). Vocational training in virtual reality: A case study using the 4C/id model. *Multimodal Technologies and Interaction*, 6(7), 49. 08–224. https://doi.org/10.3991/ijet.v15i24.16615

Mulders, M., Buchner, J., & Kerres, M. (2020). A framework for the use of immersive virtual reality in learning environments. *International Journal of Emerging Technologies in Learning (iJET)*, 15, 2

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction*, *58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011

Munshi, A., Biswas, G., Baker, R., Ocumpaugh, J., Hutt, S., & Paquette, L. (2023). Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. *Journal of Computer Assisted Learning*, *39*(2), 351-368.

Niegemann, H., & Weinberger, A. (Hrsg.). (2020*). Handbuch Bildungstechnologie: Konzeption und Einsatz digitaler Lernumgebungen*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-54368-9

Obourdin, G., De Maeyer, S., & Van Den Bossche, P. (2024). Unlocking the power of immersive learning: The FAIRI instructional design proposition for adaptive immersive virtual reality. *Computers & Education: X Reality*, *5*, 100084. https://doi.org/10.1016/j.cexr.2024.100084

Oliver, R., & Herrington, J. (2000). Using situated learning as a design strategy for Web-based learning. *Instructional and cognitive impacts of web-based education* (pp. 178-191). IGI Global.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. https://doi.org/10.1136/bmj.n71

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110, 785–797. https://doi.org/10.1037/ edu0000241

Paxinou, E., Georgiou, M., Kakkos, V., Kalles, D., & Galani, L. (2022). Achieving educational goals in microscopy education by adopting virtual reality labs on top of face-to-face tutorials. *Research in Science & Technological Education*, 40(3), 320–339.

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In *International Handbook of Emotions in Education* (pp. 130–151). Routledge.

Pieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. *Computers in Human Behavior*, 86, 165–173. https://doi.org/10.1016/j.chb.2018.04.022

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3.

Plass, J. L., & Pawar, S. (2020). Toward a taxonomy of adaptivity for learning. *Journal of Research on Technology in Education*, *52*(3), 275–300. https://doi.org/10.1080/15391523.2020.1719943

Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods*, *6*(1), 63-73.

Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1–12.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 1–29.

Rausch, A., Deutscher, V., & Seifried, J. (2024). Szenario-basierte Diagnostik kaufmännischer Handlungskompetenz mit LUCA Office Simulation. *In Online-Assessment: Grundlagen und Anwendung von Online-Tests in der Unternehmenspraxis* (pp. 165-180). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-68684-3_12

Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA–Funktionen, Einsatzmöglichkeiten und Forschungsausblick. In *Zeitschrift für Berufs-und Wirtschaftspädagogik* (Bd. 3, S. 372–394).

Rausch, A., Goller, M., & Steffen, B. (2022). Uncovering Informal Workplace Learning by Using Diaries. In M. Goller, E. Kyndt, S. Paloniemi, & C. Damşa (Eds.), *Methods for Researching Professional Learning and Development* (Vol. 33, pp. 43–70). Springer International Publishing. https://doi.org/10.1007/978-3-031-08518-5_3

Rausch, A., Kögler, K., & Seifried, J. (2019). Validation of embedded experience sampling (EES) for measuring non-cognitive facets of problem-solving competence in scenario-based assessments. *Frontiers in Psychology*, 10, 1200. https://doi.org/10.3389/fpsyg.2019.01200

Rausch, A., Kögler, K., Frötschl, C., Bergrab, M., & Brandt, S. (2017). Problemlöseprozesse sichtbar machen: Analyse von Logdaten aus einer computerbasierten Bürosimulation. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *113*(4), 569–594. https://doi.org/10.25162/zbw-2017-0024

Ravichandran, R., & Mahapatra, J. (2023). Virtual reality in vocational education and training: Challenges and possibilities. *Journal of Digital Learning and Education*, 3 (1), 25–31. https://doi.org/10.52562/jdle.v3i1.602

Reigeluth, C. M., Aslan, S., Chen, Z., Dutta, P., Huh, Y., Lee, D., & Watson, S. L. (2015). Personalized integrated educational system: technology functions for the learner-centered paradigm of education. *Journal of Educational Computing Research*, 53(3), 459–496.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, 13(3), 273–304. https://doi. org/10.1207/s15327809jls1303_2

Renkl, A., & Scheiter, K. (2017). Studying Visual Displays: How to Instructionally Support Learning. *Educational Psychology Review*, *29*(3), 599–621. https://doi.org/10.1007/s10648-015-9340-4

Robertson, K., Rosasco, C., Feuz, K., Schmitter-Edgecombe, M., & Cook, D. (2015). Prompting technologies: A comparison of time-based and context-aware transition-based prompting. *Technology and Health Care*, *23*(6), 745–756. https://doi.org/10.3233/THC-151033

Sangmeister, J., Winther, E., Deutscher, V., Bley, S., Kreuzer, C., & Weber, S. (2019). *Digital workplace learning: Bridging formal and informal learning with digital technologies* https://doi.org/10.1007/978-3-319-97934-2_5

Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. *Zeitschrift für Erziehungswissenschaft*, *24*(5), 1039–1060. https://doi.org/10.1007/s11618-021-01047-y

Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation in school* (pp. 197–223). New York: Taylor Francis.

Schlüter, C., & Kretschmer, V. (2020). Next level training in logistics: Evaluation of a virtual reality-based serious game for warehouse logistics. Proceedings of the 19th *International Conference on modeling & applied simulation (MAS 2020).* https://doi.org/10.46354/i3m.2020.mas.018

Schmid, R., & Petko, D. (2019). Does the use of educational technology in personalized learning environments correlate with self-reported digital skills and beliefs of secondary-school students? *Computers & Education*, *136*, 75–86. https://doi.org/10.1016/j.compedu.2019.03.006

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage publications.

Schoor, C., Hahnel, C., Mahlow, N., Klagges, J., Kroehne, U., Goldhammer, F., & Artelt, C. (2020). Multiple document comprehension of university students: Test development and relations to person and process characteristics. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student learning in German higher education: Innovative measurement approaches and research results* (pp. 221–240). Springer VS. https://doi.org/10.1007/978-3-658-27886-1_11.

Schoor, C., Rouet, J. F., Artelt, C., Mahlow, N., Hahnel, C., Kroehne, U., & Goldhammer, F. (2021). Readers' perceived task demands and their relation to multiple document comprehension strategies and outcome. *Learning and Individual Differences*, 88, 102018.

Schumacher, C., & Ifenthaler, D. (2021). Investigating prompts for supporting students' self-regulation – A remaining challenge for learning analytics approaches? *The Internet and Higher Education*, *49*, 100791. https://doi.org/10.1016/j.iheduc.2020.100791

Schumann, S., Seeber, S., & Abele, S. (Eds.). (2022). *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (Bd. 41). wbv Publikation. https://doi.org/10.3278/9783763971381

Sedlmeier, P. (2001). Intelligent tutoring systems. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 7674–7678). Pergamon. Red.

Seifried, J., Rausch, A., Kögler, K., Brandt, S., Eigenmann, R., Schley, T., Siegfried, C., Egloffstein, M., Küster, J., Wuttke, E., Martens, T., & Wolf, K. D. (2016). *Kapitel 7 Problemlösekompetenz angehender Industriekaufleute – Konzeption des Messinstruments und ausgewählte empirische Befunde (DomPL-IK).*

Serge, S. R., Priest, H. A., Durlach, P. J., & Johnson, C. I. (2013). The effects of static and adaptive performance feedback in game-based training. *Computers in Human Behavior*, *29*(3), 1150–1158. https://doi.org/10.1016/j.chb.2012.10.007

Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments*, *7*(1), 33. https://doi.org/10.1186/s40561-020-00140-9

Shemshack, A. & Spector, J. M. (2021). A comprehensive analysis of personalized learning components. *Journal of Computers in Education*, *8*(4), 485–503. https://doi.org/10.1007/s40692-021-00188-7

Shih, K. P., Chen, H. C., Chang, C. Y., & Kao, T. C. (2010). The development and implementation of scaffolding-based self-regulated Learning System for e/m-Learning. *Journal of Educational Technology & Society*, 13(1), 80–93.

Shou, Y., & Olney, J. (2021). Attitudes toward risk and uncertainty: The role of subjective knowledge and affect. *Journal of Behavioral Decision Making*, 34(3), 393–404. https://onlinelibrary.wiley.com/doi/10.1002/bdm.2217.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: teleoperators and virtual environments*, 8(5), 560–565.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Sonnenberg, C., & Bannert, M. (2016). Evaluating the Impact of Instructional Support Using Data Mining and Process Mining: A Micro-Level Analysis of the Effectiveness of Metacognitive Prompts. *Journal of Educational Data Mining*, 8(2), 51-83

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2011). Cognitive load theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Academic Press

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality*, 16(1), 33–48. https://doi.org/10.20870/IJVR.2016.16.1.2880

Thillmann, H., Künsting, J., Wirth, J., & Leutner, D. (2009). Is it Merely a Question of “What” to Prompt or Also “When” to Prompt?: The Role of Point of Presentation Time of Prompts in Self-Regulated Learning. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 105–115. https://doi.org/10.1024/1010-0652.23.2.105

Thomann, H., Deutscher, V., Rausch, A., & Seifried, J. (22.08.2024). Personalized Prompts to Support Problem-Solving in the Digital Office Simulation LUCA [Conference presentation]. *EARLI SIG 14 Learning and Professional Development*, Jyväskylä, Finland.

Umutlu, D., & Gursoy, M. E. (2022). Leveraging artificial intelligence techniques for effective scaffolding of personalized learning in workplaces. *Artificial Intelligence Education in the Context of Work*, 59-76.

U.S. Department of Education. (2016). Future ready learning: reimagining the role of technology in education. *Office of Educational Technology*, Washington, D.C. Retrieved December, 20, 2024 from http://tech.ed.gov/files/2015/12/NETP16.pdf van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Effects of self-regulated learning prompts in a flipped history classroom. *Computers in Human Behavior*, *108*, 106318. https://doi.org/10.1016/j.chb.2020.106318

Van Der Graaf, J., Raković, M., Fan, Y., Lim, L., Singh, S., Bannert, M., Gašević, D., & Molenaar, I. (2023). How to design and evaluate personalized scaffolds for self-regulated learning. *Metacognition and Learning*, *18*(3), 783–810. https://doi.org/10.1007/s11409-023-09361-y

van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153-160. https://doi.org/10.1016/j.compedu.2012.11.025

van Merriënboer, J. J. G. & Kirschner, P. (2018). *Ten steps to complex Learning: A systematic approach to four-component instructional design*. New York: Routledge/Taylor & Francis.

van Merriënboer, J. J., Clark, R. E., & De Croock, M. B. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational technology research and development*, 50(2), 39-61.

Van Schoors, R., Elen, J., Raes, A., & Depaepe, F. (2021). An overview of 25 years of research on digital personalised learning in primary and secondary education: A systematic review of conceptual and methodological trends. *British Journal of Educational Technology*, *52*(5), 1798–1822. https://doi.org/10.1111/bjet.13148

Van Schoors, R., Elen, J., Raes, A., Vanbecelaere, S., & Depaepe, F. (2023). The Charm or Chasm of Digital Personalized Learning in Education: Teachers' Reported Use, Perceptions and Expectations. *TechTrends, 67*(2), 315–330. https://doi.org/10.1007/s11528-022-00802-0

Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, *27*(1), 118–130. https://doi.org/10.1016/j.chb.2010.07.038

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika*, *60*(3), 419–435. https://doi.org/10.1007/BF02294384

Viechtbauer, W. (2007). Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in medicine*, *26*(1), 37-52.

Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the metafor Package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

Walkington, C., & Bernacki, M. L. (2020). Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions. *Journal of Research on Technology in Education*, *52*(3), 235–252. https://doi.org/10.1080/15391523.2020.1747757

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23. https://doi.org/10.1007/ BF02504682

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review*, 35(3), 69.

Wang, Y.-H. (2020). Design-based research on integrating learning technology tools into higher education classes to achieve active learning. *Computers & Education*, *156*, 103935. https://doi.org/10.1016/j.compedu.2020.103935

What Works Clearinghouse. (2022). What Works Clearinghouse standards handbook (Version 5.0). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, Retrieved August, 10, 2022 from https://ies.ed.gov/ncee/wwc/handbooks

Wilde, M., Bätz, K., Kovaleva, A., & Urhahne, D. (2009). Überprüfung einer Kurzskala intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften ZfDN* ; Biologie, Chemie, Physik, 15, 31–45.

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Winther E, Seeber S, Festner D, et al. (2016) Large scale-assessments in der kaufmännischen Berufsbildung (CoBALIT). In: Oser F, Landenberger M, Beck K (eds) *Technologiebasierte Kompetenzmessung in der beruflichen Bildung – Resultate aus dem Forschungsprogramm ASCOT*. Bielefeld: Bertelsmann, pp.55–73.

Winther, E., & Achtenhagen, F. (2009). Simulationsaufgaben als innovatives Testverfahren für Industriekaufleute im Rahmen eines VET-LSA. [Simulation tasks as an innovative test procedure for industrial clerks in the context of a VET-LSA]. *Wirtschaft & Erziehung*, 10, 317–325.

Winther, E., & Achtenhagen, F. (2010). Berufsfachliche Kompetenz. Messinstrumente und empirische Befunde zur Mehrdimensionalität beruflicher Handlungskompetenz. [Vocational competence. Measurement instruments and empirical findings on the multidimensionality of vocational competence]. *Berufsbildung in Wissenschaft und Praxis*, 1, 18–21.

Wirth, J. (2009). Promoting Self-Regulated Learning Through Prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 91–94. https://doi.org/10.1024/1010-0652.23.2.91

Won, M., Ungu, D. A. K., Matovu, H., Treagust, D. F., Tsai, C. C., Park, J., … Tasker, R. (2023). Diverse approaches to learning with immersive Virtual Reality identified from a systematic review. *Computers & Education*, 195, Article 104701.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in Problem Solving*. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

World Economic Forum. (2023). Future of jobs report 2023. Retrieved November, 26, 2024 https://www.weforum.org/reports/the-future-of-jobs-report-2023.

Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, 140, 103599. https://doi.org/10.1016/j.compedu.2019.103599

Young, M. F., & Kulikowich, J. M. (1992). Anchored instruction and anchored assessment: An ecological approach to measuring situated learning. *Annual Meeting of the American Educational Research Association*, San Francisco, CA, 1-21.

Zahabi, M., & Abdul Razak, A. M. (2020). Adaptive virtual reality-based training: A systematic literature review and framework. *Virtual Reality*, *24*(4), 725–752. https://doi.org/10.1007/s10055-020-00434-w

Zhang, L., Basham, J. D., & Yang, S. (2020a). Understanding the implementation of personalized learning: A research synthesis. *Educational Research Review*, *31*, 100339. https://doi.org/10.1016/j.edurev.2020.100339

Zhang, L., Yang, S., & Carter, R. A. (2020b). Personalized learning and ESSA: What we know and where we go. *Journal of Research on Technology in Education*, 52(3), 253–274. https://doi.org/10.1080/ 15391523.2020.1728448

Zhang, L., Carter Jr, R. A., Basham, J. D., & Yang, S. (2022). Integrating instructional designs of personalized learning through the lens of universal design for learning. *Journal of Computer Assisted Learning*, *38*(6), 1639-1656.

Zhao, S., Hai, G., & Ma, H. (2024). Adaptive Learning Systems: Exploring Personalized Paths in Vocational Education. *Curriculum Learning and Exploration*, *2*(2).

Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education*, *2*(4), 399–420. https://doi.org/10.1007/s40692-015-0036-z

Zheng, L., Long, M., Zhong, L., & Gyasi, J. F. (2022). The effectiveness of technology-facilitated personalized learning on learning achievements and learning perceptions: A meta-analysis. *Education and Information Technologies*, 27(8), 11807–11830. https://doi.org/10.1007/s10639-022-11092-7

Zhong, L. (2022). A systematic review of personalized learning in higher education: learning content structure, learning materials sequence, and learning readiness support. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2022.2061006

Zimmerman, B. J., & Moylan, A. R. (2009). *Self-regulation: Where metacognition and motivation intersect.* In Handbook of metacognition in education (pp. 299-315). Routledge.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Handbook of self-regulation of learning and performance* (pp. 1–12). Routledge.

Zinn, B. (2019). Lehren und Lernen zwischen Virtualität und Realität. *Journal of Technical Education* (JOTED), 7(1).

7 Appendix

Table A 1: First prompt concept for the supplier selection task scenario (phase 1)

	Cognitive Level	**Prompt Text**	**Personalization**	**Appearance**
1	Ereignisabfrage (nach 20min)	**Zwischenstand** Deine Kollegin Aylin hat mitbekommen, dass Du Deine erste Aufgabe erhalten hast. Sie fragt: „Wie kommst du mit deiner Aufgabe zurecht?“ - Ich weiß was zu tun ist. (Nr. 2) - Ich bin gerade dabei, mir einen Überblick zu verschaffen. (Nr. 3) - Keine Ahnung was zu tun ist. (Nr. 4)	All learners	Time-based
2	Ereignisabfrage (nach 50min)	**Endspurt** Sie haben nur noch 5 Minuten! Denken Sie daran noch eine Antwortmail zu formulieren! Hat Ihnen die Bearbeitung des Szenarios Spaß gemacht? - Ja, die Bearbeitung des Szenarios hat mir Spaß gemacht. - Nein, die Bearbeitung des Szenarios hat mir nicht gefallen.	All learners	Time-based
1	non-cognitive	Sie haben gut erkannt, dass es sinnvoll sein könnte, auch weitere Auswahlkriterien für die Nutzwertanalyse heranzuziehen. Machen Sie weiter so!	Group of learners	Action-based; Time-based
2		Alles klar! Das klingt doch sehr gut.	Individual Learners	Action-based; Time-based
3	meta-cognitive	Obwohl ich schon ein bisschen Berufserfahrung habe, hilft es mir, die Aufgabe in verschiedene Arbeitsschritte aufzuteilen. Weiter so!	Individual Learners	Action-based; Time-based
4		Bevor ich mit der Lieferantenauswahl beginne, nehme ich mir immer einen Moment Zeit, um mir Notizen zu machen. Am besten verschaffst Du dir zunächst einen Überblick über alle Angebote, Aktennotizen und Mails. Im nächsten Schritt gilt es, die Arbeitsdatei zur Nutzwertanalyse zu vervollständigen. Hierbei fand ich die alte Nutzwertanalyse zur Orientierung sehr hilfreich. Zum Schluss formulierst Du deine Entscheidung mit Begründung in der Antwortmail.	Individual Learners	Action-based; Time-based
5		**Arbeitsschritte Lieferantenauswahl (Ereignis)** Ein Kollege kommt zu Ihnen an den Arbeitsplatz und möchte sich über ihre derzeitigen Aufgaben erkundigen. Er fragt: „Wie gehen Sie bei der	All Learners	Time-based

		Lieferantenauswahl vor und welche Arbeitsschritte gilt es zu berücksichtigen?“		
6		Hast Du dir schon die die Aktennotiz vom Lieferanten DRIVEN angeschaut und in deiner Auswahl berücksichtigt?	Individual Learners	Action-based; Time-based
7		Hast Du dir schon die Aktennotiz vom Lieferanten Jinshu Gongsi angeschaut und in deiner Auswahl berücksichtigt?	Individual Learners	Action-based; Time-based
8		Hast Du dir schon das Angebot des Lieferanten DRIVEN angesehen und in deiner Lieferantenauswahl berücksichtigt?	Individual Learners	Action-based; Time-based
9		Hast Du dir schon das Angebot des Lieferanten POWER angesehen und in deiner Lieferantenauswahl berücksichtigt?	Individual Learners	Action-based; Time-based
10		Hast du die Mail vom neuen Testbericht zu den Akkus der Firma POWER SE von Frau Hellwig aus dem Qualitätsmanagement schon gesehen?	Individual Learners	Action-based; Time-based
11	cognitive	Hast du die Bezeichnung der Nutzwertanalyse schon ausgefüllt? Für weitere Vorgänge ist es wichtig direkt identifizieren zu können, welche Person wann und für welches Produkt die Nutzwertanalyse erstellt hat.	Individual Learners	Action-based; Time-based
12		Berechnen Sie stets den Stückpreis, der gezahlt werden muss.	Individual Learners	Action-based; Time-based
13		Berechnen Sie stets den Stückpreis, der gezahlt werden muss und berücksichtigen Sie besondere Zahlungskonditionen.	Individual Learners	Action-based; Time-based
14		Berechnen Sie stets den Stückpreis, der gezahlt werden muss und berücksichtigen Sie besondere Zahlungskonditionen.	Individual Learners	Action-based; Time-based
15		Hast Du bei der Währungsumrechnung den aktuellen Wechselkurs beachtet? Eine Tabelle zu den Wechselkursen findest du im Nachschlagewerk.	Individual Learners	Action-based; Time-based
16		Es ist wichtig, dass du dir das Angebot von Jinshu Gongsi sorgfältig ansiehst und den korrekten Angebotspreis erfasst hast.	Individual Learners	Action-based; Time-based
17		Wenn Sie sich unsicher sind, die Qualität in Worten zu bewerten, dann können Sie sich an der Nutzwertanalyse vom Jahr 2020 orientieren.	Individual Learners	Action-based; Time-based
18		Schauen Sie sich das Angebot von DRIVEN an und prüfen Sie, ob die korrekte Lieferzeit in die Vorlage eingetragen wurde.	Individual Learners	Action-based; Time-based
19		Schauen Sie sich das Angebot von Jinshu Gongsi an und prüfen Sie, ob die korrekte Lieferzeit in die Vorlage eingetragen wurde.	Individual Learners	Action-based; Time-based
20		Mir ist jetzt schon öfters aufgefallen, dass das Unternehmen Steiger & Söhne fehlerhafte Angebote schickt.	Individual Learners	Action-based; Time-based

21		Wir schätzen die Testergebnisse von unseren Kollegen aus dem Qualitätsmanagement sehr.	Individual Learners	Action-based; Time-based
22		Die Gewichtung hängt von unterschiedlichen Faktoren ab.	Individual Learners	Action-based; Time-based
23		Die Kollegen betonen immer wieder, dass die Gewichtung insgesamt 1 ergeben muss. Kommst du auf diese Summe?	Individual Learners	Action-based; Time-based
24		Bei der Vergabe von Punkten orientiere ich mich immer an ähnlichen Bewertungen aus dem Vorjahr.	Individual Learners	Action-based; Time-based
25		Gewichtete Punkte berechne ich ganz einfach durch Multiplikation (Gewichtung x Punktzahl).	Individual Learners	Action-based; Time-based
26		Herr Böschek meinte zu mir mal, dass ihm ein Vergleichswert sehr wichtig sei. Deshalb berechne ich am Ende immer die Summe aller gewichteten Punkte, um alle Kriterien eines Lieferanten zu berücksichtigen.	Individual Learners	Action-based; Time-based
27		neben den bereits vorhandenen Auswahlkriterien (Preis, Qualität, Lieferzeit) kannst Du auch noch weitere Kriterien bei Deiner Begründung in die Antwortmail einbeziehen. Unsere Unternehmenswerte können Dir dabei einen Hinweis geben. Diese findest Du in der Unternehmensbroschüre.	Individual Learners	Action-based; Time-based

Table A 2: Redefined prompt concept for the supplier selection task scenario (phase 2)

	Cognitive Level	**Prompt Text**	**Time/Action**	**Number of Learners**	**Appearance**
1	Ereignisabfrage	**Zwischenstand** Deine Kollegin Aylin hat mitbekommen, dass Du Deine erste Aufgabe erhalten hast. Sie fragt: „Wie kommst du mit deiner Aufgabe zurecht?“ - Ich weiß was zu tun ist. (Nr. 2) - Ich bin gerade dabei, mir einen Überblick zu verschaffen. (Nr. 3) - Keine Ahnung was zu tun ist. (Nr. 4)	20min	All learners	Time-based
2	Ereignisabfrage	**Mail-Programm**	30min	All learners	Time-based

		es ist zwischendurch immer mal wichtig, dass du überprüfst, ob du Mails bekommen hast. Häufig erhältst du hilfreiche Informationen zur Lösung der Aufgabe mit diesen Mails.			
3	Ereignisabfrage	**Endspurt** Sie haben nur noch 5 Minuten! Denken Sie daran noch eine Antwortmail zu formulieren! Hat Ihnen die Bearbeitung des Szenarios Spaß gemacht? - Ja, die Bearbeitung des Szenarios hat mir Spaß gemacht. - Nein, die Bearbeitung des Szenarios hat mir nicht gefallen.	50min	All learners	Time-based
1	non-cognitive	Sie haben gut erkannt, dass es sinnvoll sein könnte, auch weitere Auswahlkriterien für die Nutzwertanalyse heranzuziehen. Machen Sie weiter so!		Individual Learners	Action- & Time-based
2		Alles klar! Das klingt doch sehr gut.	Antwortauswahl 1 von Ereignis 1	Individual Learners	Action-based
3	meta-cognitive	Obwohl ich schon ein bisschen Berufserfahrung habe, hilft es mir, die Aufgabe in verschiedene Arbeitsschritte aufzuteilen. Weiter so!	Antwortauswahl 2 von Ereignis 1	Individual Learners	Action-based
4		Bevor ich mit der Lieferantenauswahl beginne, nehme ich mir immer einen Moment Zeit, um mir Notizen zu machen. Insbesondere die Erstellung einer Nutzwertanalyse erfordert verschiedene Arbeitsschritte. Am besten verschaffst Du dir zunächst einen Überblick über alle Angebote, Aktennotizen und Mails, indem Du dir eine kurze Checkliste im Notiz-Programm machst. Diese könnte z.B. so aussehen: Angebote: - Lieferant DRIVEN - ... Aktennotizen: - Lieferant DRIVEN	Antwortauswahl 3 von Ereignis 1	Individual Learners	Action-based

		- ... E-Mails: - Qualitätsmanagement - ... Weitere Dokumente: - ... Im nächsten Schritt gilt es, die Arbeitsdatei zur Nutzwertanalyse zu vervollständigen. Hierbei fand ich die alte Nutzwertanalyse zur Orientierung sehr hilfreich. Zum Schluss formulierst Du deine Entscheidung mit Begründung in der Antwortmail.			
5		Hast Du dir schon die die Aktennotizen vom Lieferanten DRIVEN und Jinshu Gongsi im ERP-System angeschaut und in deiner Auswahl berücksichtigt?		Individual Learners	Action- & Time-based
6		Hast Du dir schon alle Angebote der Lieferanten angesehen und in deiner Lieferantenauswahl berücksichtigt?		Individual Learners	Action- & Time-based
7		Hast du die Mail vom neuen Testbericht zu den Akkus der Firma POWER SE von Frau Hellwig aus dem Qualitätsmanagement schon gesehen? Der Testbericht dürfte für deine Lieferantenauswahl hilfreich sein.		Individual Learners	Action- & Time-based
9	cognitive	Hast du die Bezeichnung der Nutzwertanalyse schon ausgefüllt? Für weitere Vorgänge ist es wichtig direkt identifizieren zu können, welche Person wann und für welches Produkt die Nutzwertanalyse erstellt hat.		Individual Learners	Action- & Time-based
10		Berechnen Sie stets den Stückpreis, der gezahlt werden muss und berücksichtigen Sie ggf. Rabatt und Skonto als besondere Zahlungskonditionen.		Individual Learners	Action- & Time-based
11		Berechnen Sie stets den Stückpreis, der gezahlt werden muss und berücksichtigen Sie den Rabatt als besondere Zahlungskondition.		Individual Learners	Action- & Time-based
12		Berechnen Sie stets den Stückpreis, der gezahlt werden muss und berücksichtigen Sie Skonto als besondere Zahlungskondition.		Individual Learners	Action- & Time-based
13		Hast Du bei der Währungsumrechnung den aktuellen Wechselkurs beachtet?		Individual Learners	Action- & Time-based

		~~Eine Tabelle zu den Wechselkursen findest du im Nachschlagewerk.~~			
14		Es ist wichtig, dass du dir das Angebot von Jinshu Gongsi sorgfältig ansiehst und den korrekten Angebotspreis erfasst hast.		Individual Learners	Action- & Time-based
15		Schauen Sie sich das Angebot von DRIVEN an und prüfen Sie, ob die korrekte Lieferzeit in die Vorlage eingetragen wurde. Neben den Preis wird auch der Lieferzeit eine große Bedeutung beigemessen.		Individual Learners	Action- & Time-based
16		Schauen Sie sich das Angebot von Jinshu Gongsi an und prüfen Sie, ob die korrekte Lieferzeit in die Vorlage eingetragen wurde. Neben den Preis wird auch der Lieferzeit eine große Bedeutung beigemessen.		Individual Learners	Action- & Time-based
17		Mir ist jetzt schon öfters aufgefallen, dass das Unternehmen Steiger & Söhne fehlerhafte Angebote schickt. Einmal habe ich sogar ein Angebot mit dem falschen Produkt erhalten. Dummerweise habe ich es nicht bemerkt und alles umsonst gerechnet!		Individual Learners	Action- & Time-based
18		Wir schätzen die Testergebnisse von unseren Kollegen aus dem Qualitätsmanagement sehr. Haben Sie auf Frau Hellwigs Mail mit den Akku-Testlauf reagiert und entschieden, wie Sie mit dem Angebot von POWER umgehen?		Individual Learners	Action- & Time-based
19		Die Gewichtung hängt von unterschiedlichen Faktoren ab. Ich habe mitbekommen, dass die Lieferzeit in gleicher Höhe wie der Preis und die Qualität gewichtet werden soll.		Individual Learners	Action- & Time-based
20		Die Kollegen betonen immer wieder, dass die Gewichtung insgesamt 1 ergeben muss. Kommst du auf diese Summe?		Individual Learners	Action- & Time-based
21		Bei der Vergabe von Punkten orientiere ich mich immer an ähnlichen Bewertungen aus dem Vorjahr. Beispielsweise machen 5 Cent pro Stück nicht zwingend einen Unterschied.		Individual Learners	Action- & Time-based

22		Gewichtete Punkte berechne ich ganz einfach durch Multiplikation (Gewichtung x Punktzahl).		Individual Learners	Action- & Time-based
23		neben den bereits vorhandenen Auswahlkriterien (Preis, Qualität, Lieferzeit) kannst Du auch noch weitere Kriterien bei Deiner Begründung in die Antwortmail einbeziehen. Unsere Unternehmenswerte können Dir dabei einen Hinweis geben. Diese findest Du in der Unternehmensbroschüre.		Individual Learners	Action- & Time-based

Table A 3: Final prompt concept for the supplier selection task scenario (after phase 3)

	Cognitive Level	**Prompt Text**	**Time/Action**	**Item**
1	Ereignisabfrage (cognitive & meta-cognitive)	**Effizientes Arbeiten** Ist der Testbericht zu den Akkus von Power SE angekommen? Bei einem schlechten Testbericht kommt der Lieferant für einen Vergleich nicht infrage. Der Lieferant Steiger & Söhne hat öfters schon fehlerhafte Angebote verschickt. Gegebenenfalls brauchst du mache Angebote im Vergleich nicht berücksichtigen und kannst Dir damit viel Arbeit sparen! Zutreffendes ist anzukreuzen: - Der Hinweis zur Arbeitseffizienz hat mir weitergeholfen - Der Hinweis zur Arbeitseffizienz hat mir nicht weitergeholfen	15min	Alle POWER Items dok007_hellwigmail
2	Ereignisabfrage (meta-cognitive)	**Mails checken** Während deiner Aufgabenbearbeitung ist es ratsam hin und wieder zu überprüfen, ob Du neue Mails bekommen hast. Häufig bekommst du hier wertvolle Tipps, die Dir weiterhelfen können! Hast Du dir die Nutzwertanalyse GreenLi vom Vorjahr aus unserem internen Ordnersystem schon zur Orientierung angesehen? - Ja, habe ich. - Nein, habe ich aber noch vor.	29min	dok009_greenli

		- Nein, brauche ich nicht.		
3	Ereignisabfrage (meta-cognitive)	**Mails checken** Vergiss nicht Deine Mails zu checken, damit Dir keine Tipps entgehen! Falls du Probleme bei Deinen Berechnungen hast, findest Du eine ausführliche Anleitung im Nachschlagewerk. - Der Hinweis hat mir weitergeholfen. - Der Hinweis hat mir nicht weitergeholfen.	39min	NA
4	Ereignisabfrage (meta-cognitive)	**Endspurt** Sie haben nur noch 5 Minuten! Denken Sie daran noch eine Antwortmail zu formulieren! Hat Ihnen die Bearbeitung des Szenarios Spaß gemacht? - Ja, die Bearbeitung des Szenarios hat mir Spaß gemacht. - Nein, die Bearbeitung des Szenarios hat mir nicht gefallen.	50min	antw001_bedarf-mail
1	non-cognitive	Sie haben gut erkannt, dass es sinnvoll sein könnte, auch weitere Auswahlkriterien für die Nutzwertanalyse heranzuziehen. Machen Sie weiter so!	40min	NA
1	cognitive	Hast du die Bezeichnung der Nutzwertanalyse schon ausgefüllt? Für weitere Vorgänge ist es wichtig direkt identifizieren zu können, welche Person wann und für welches Produkt die Nutzwertanalyse erstellt hat.	20min	tbl001_datum tbl002_sachbearbeiter tbl003_produkt
2		Beachte bei der Lieferantenanalyse auch die Gewichtung der einzelnen Kriterien (Preis, Qualität, Lieferzeit). Hierbei kann Dir die Nutzwertanalyse GreenLI aus dem Vorjahr helfen. Die Summe Deiner Gewichtung muss dabei immer 1 ergeben! Die gewichteten Punkte kannst Du einfach durch Multiplikation berechnen (Gewichtung x Punktzahl). Berechne am Ende auch immer die Summe Deiner gewichteten Punkte.	23min	dok009_greenli tbl007_sum-gew driven007_gewpunkte-preis driven008_gewpunkte-qualitaet driven009_gewpunkte-lieferzeit power007_gewpunkte-preis power008_gewpunkte-qualitaet power009_gewpunkte-lieferzeit steiger007_gewpunkte-preis steiger008_gewpunkte-qualitaet steiger009_gewpunkte-lieferzeit jinshu007_gewpunkte-preis jinshu008_gewpunkte-qualitaet jinshu009_gewpunkte-lieferzeit

3		Schau dir das Angebot von DRIVEN an und prüfe, ob Du die korrekte Lieferzeit in die Vorlage eingetragen hast.	26min	driven003_lieferzeit
4		zur Berechnung des Angebotspreises von DRIVEN solltest Du sowohl Rabatt als auch Skonto berücksichtigen. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
5		berücksichtige neben Skonto auch Rabatt als besondere Zahlungskondition. Bei der Kalkulation wird zuerst der Rabatt abgezogen und anschließend das Skonto. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
6		bei der Berechnung des Angebotspreises für den Lieferanten DRIVEN solltest du darauf achten den Rabatt vor dem Skonto abzuziehen. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
7		es ist wichtig, dass Du sowohl Rabatt als auch Skonto für das Angebot von DRIVEN berücksichtigst. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
8		neben Rabatt solltest Du auch Skonto als besondere Zahlungskondition für den Lieferanten DRIVEN berücksichtigen. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
9		bei der Berechnung des Angebotspreises für den Lieferanten DRIVEN solltest Du Rabatt und Skonto einzeln verrechnen. Achte zudem darauf Rabatt vor Skonto abzuziehen. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	29min	driven001_preis
10		berücksichtige bei dem Lieferanten Jinshu Gongsi stets das Skonto als besondere Zahlungskondition. Eine ausführliche Anleitung findest Du im Nachschlagewerk.	37min	jinshu001_preis
11		es ist wichtig, dass du bei der Berechnung des Angebots von Jinshu Gongsi sowohl dem Wechselkurs als	37min	jinshu001_preis

		auch das Skonto berücksichtigst. Eine ausführliche Anleitung findest Du im Nachschlagewerk.		
12		hast du bei der Währungsumrechnung für den Lieferanten Jinshu Gongsi den aktuellen Wechselkurs beachtet? Eine Tabelle zu den Wechselkursen findest du im Nachschlagewerk.	37min	jinshu001_preis
13		neben den bereits vorhandenen Auswahlkriterien (Preis, Qualität, Lieferzeit) kannst Du auch noch weitere Kriterien bei Deiner Begründung in die Antwortmail einbeziehen. Unsere Unternehmenswerte können Dir dabei einen Hinweis geben. Diese findest Du in der Unternehmensbroschüre.	45min	dok010_unternehmensbroschuere antw005_grund-oekolog antw006_grund-soz

Declaration in lieu of oath

Declaration in lieu of oath according to section 8 subsection 2 No. 1(b) of the Regulations and Procedures Governing the Doctoral Dissertation to Earn a Doctoral Degree in Business at the University of Mannheim

Eidesstattliche Versicherung

Eidesstattliche Versicherung gemäß § 8 Absatz 2 Satz 1 Buchstabe b) der Promotionsordnung der Universität Mannheim zur Erlangung des Doktorgrades der Betriebswirtschaftslehre (Dr. rer. pol.)

(1) The submitted doctoral dissertation on the subject 'Personalized Learning in Vocational Education' is my own work and to the rules of proper scientific conduct.
Bei der eingereichten Dissertation mit dem Titel 'Personalisiertes Lernen in der beruflichen Bildung' handelt es sich um mein eigenständig erstelltes Werk, das den Regeln guter wissenschaftlicher Praxis entspricht.

(2) I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any direct and indirect quotations derived from the works of others.
Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtliche und nicht wörtliche Zitate aus anderen Werken als solche kenntlich gemacht.

(3) I did not yet present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

(4) I hereby confirm the accuracy of the affirmation above.
Die Richtigkeit der vorstehenden Erklärung bestätige ich.

(5) I am aware of the significance of this affirmation and the legal ramifications in case of untrue or incomplete statements. I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.
Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

I agree that for the purpose of assessing plagiarism the dissertation may be electronically forwarded, stored, and processed.

Ich bin damit einverstanden, dass die Arbeit zum Zwecke des Plagiatsabgleichs in elektronischer Form versendet, gespeichert und verarbeitet wird.

Mannheim, 31.01.2025	Herbert Thomann
Place, Date	Signature
Ort, Datum	*Unterschrift*

Author's addendum: During the preparation of this work, the author used ChatGPT and Grammarly in order to optimize the language style. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ergänzung des Autors: Bei der Vorbereitung dieser Arbeit hat der Autor ChatGPT und Grammarly verwendet, um den Sprachstil zu optimieren. Nach der Nutzung dieser Tools/Dienste hat der Autor den Inhalt nach Bedarf überprüft und bearbeitet und übernimmt die volle Verantwortung für den Inhalt der Veröffentlichung.

Doctoral Study Program

GESS Graduate School of Economic and Social Sciences

Systematic Reviews and Meta-Analyses

Spring 2022

Grade: 1.0

Data Science in Action

Fall 2022

Grade: PASS

MET Intensive Longitudinal Methods: ILM in Contexts of Work and Learning

Fall 2023

Grade: 1.0

Curriculum Vitae Herbert Thomann

Academic Experience

since 03/2024 **Research Associate at the Georg-August University Göttingen**
Area Economic and Business Education (Prof. Dr. Deutscher)
Chair of Business Education and Digital Vocational Learning

06/2021 – 02/2024 **Research Associate at the University of Mannheim**
Chair for Business and Economic Education - Competence Development and Training Quality (Prof. Dr. Deutscher)

Education

11/2021 - 04/2025 **Doctoral Study Program, University of Mannheim**

09/2019 – 10/2021 **Economic and Business Education (M.Sc.), University Mannheim**

09/2015 – 07/2019 **Economic and Business Education (B.Sc.), University Mannheim**

Selected Teaching Activities

10/2024 – 03/2025 (HWS25) **Vertiefende Fachdidaktik & Unterrichtsforschung (Master)**
04/2024 – 09/2024 (FSS24) **Forschungsmethoden (Bachelor)**
Georg-August University Göttingen

09/2023 – 12/2023 **Empirische Instruktionsforschung (EIF) - Virtual & Augmented Reality in der kaufmännischen Bildung (Master)**
University of Mannheim

since 01/2022 **Supervision of bachelor's theses and master's theses**
University of Mannheim & Göttingen

Publications

2025 Thomann, H., & Deutscher, V. (2025). Scaffolding through prompts in digital learning: A systematic review and meta-analysis of effectiveness on learning achievement. *Educational Research Review*, 47, 100686. https://doi.org/10.1016/j.edurev.2025.100686

2024 Thomann, H., Zimmermann, J., & Deutscher, V. (2024). How effective is immersive VR for vocational education? Analyzing knowledge gains and motivational effects. *Computers & Education*, 220, 105127.

2022 Deutscher, V., Seifried, J., Rausch, A., Thomann, H. und Braunstein, A. (2022). Die LUCA Office Simulation in der Lehrerinnen- und Lehrerbildung – didaktische Design-Empfehlungen und erforderliche Lehrkompetenzen. In *Digital Literacy in der beruflichen Lehrer:innenbildung : Didaktik, Empirie und Innovation* (S. 107–121). Bielefeld: wbv.