Learning and Instruction 101 (2026) 102255

Contents lists available at ScienceDirect

Learning and Instruction

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

Effects of instructional design, instructional preferences, and cognitive load on problem solving and knowledge acquisition in a computer-based office simulation

Sabrina Ludwig [a,*], Andreas Rausch [a], Michelle Taub [b]

[a] Chair of Economic and Business Education – Workplace Learning, University of Mannheim, Germany
[b] Department of Learning Sciences and Educational Research, and Core Faculty of the Learning Sciences Cluster, University of Central Florida, USA

ARTICLE INFO

Keywords:
Simulation-based learning
Instructional design
Problem solving
Video-based instruction
Experimental study
Log file analysis
Instructional preferences
Cognitive load theory

ABSTRACT

Background: Instructional designs based on problem solving and self-regulation have been extensively studied in the context of computer-based learning environments. However, the question of how problem-solving phases should be embedded in instructional designs remains. Several studies, especially in STEM, suggest that 'direct instruction followed by problem-solving (DI-PS)' benefits procedural knowledge acquisition while the instructional design 'problem-solving followed by direct instruction' (PS-DI) benefits conceptual knowledge.
Aims: This study aimed to investigate the effects of DI-PS and PS-DI, learners' instructional preferences, and cognitive load on problem-solving and knowledge acquisition in a computer-based office simulation.
Sample(s): Eighty-one German undergraduate business education students participated in a pre-post-test experimental study, randomly assigned to either the DI-PS or PS-DI condition. The students worked on a supplier selection task within a computer-based office simulation.
Methods: Knowledge tests and questionnaires were employed to measure knowledge acquisition, cognitive load, and learning preferences. Log data analyses as well as variance and regression analyses were conducted.
Results: Results showed that PS-DI led to higher conceptual knowledge gains, whereas DI-PS tended to yield slightly higher problem-solving performance. Both groups improved in conceptual and procedural knowledge, yet no statistically significant differences between the sequencing conditions emerged. The PS-DI group reported higher intrinsic cognitive load during the problem-solving phase. Moreover, the DI-PS group reported higher retrospective satisfaction with the instructional design than the PS-DI group, aligning with existing literature on instructional preferences.
Conclusions: These findings provide insights for improving learning designs, personalised recommendations, and adaptive support in computer-based learning environments to enhance learners' performance. Further research is needed to examine long-term learning and transfer.

1. Introduction

Computer-based simulations are popular tools to foster self-regulated learning and domain-specific problem-solving (Chernikova et al., 2020, 2023). Simulations offer various benefits for learners, such as self-regulated (Azevedo & Gašević, 2019), personalised (van Schoors et al., 2021), and inquiry-based learning (Thibaut et al., 2018). In simulation-based learning environments, learners can acquire domain-specific practice and problem-solving competences beyond factual domain-specific knowledge without the risk of making mistakes in a real-life environment (e.g., medical simulations; Chernikova et al., 2020, 2023). Simulations can represent authentic scenarios (Ludwig & Rausch, 2023) and provide feedback on various learner interactions with the material (Saba et al., 2023). Problem solving (PS) in simulations is characterised as a cognitive process in which students actively engage and take on challenging roles during their learning. Instead of receiving information passively, they explore, analyse, and synthesise information to address the presented authentic problem. However, students might be overwhelmed during problem solving and suffer from cognitive load (CL) due to limited working memory (Sweller et al.,

* Corresponding author. University of Mannheim Chair of Economic and Business Education – Workplace Learning, University of Mannheim, L 4, 1 68161, Mannheim, Germany.
E-mail address: sabrina.ludwig@uni-mannheim.de (S. Ludwig).

https://doi.org/10.1016/j.learninstruc.2025.102255
Received 26 June 2024; Received in revised form 28 May 2025; Accepted 17 October 2025
Available online 18 November 2025

1998), resulting in ineffective attempts (Likourezos & Kalyuga, 2017) and negative affect such as confusion or frustration (Ludwig et al., 2024). Moreover, PS also represents a problem-based instructional approach, which is designed to support and structure the problem-solving process through specific pedagogical strategies. Starting an instructional sequence with highly complex and overwhelming problems can be impractical (van Merriënboer et al., 2003) and can adversely affect motivation when no training with highly complex problems is provided (Sweller et al., 1998).

Traditional approaches, in contrast, using direct instruction (DI), such as step-by-step instructions with worked examples (WE) (Sweller et al., 1998), are commonly found in teacher-centered instruction. More recently, these methods have been applied in video-based instruction (Kulgemeyer & Geelan, 2024), often placing students in an externally guided role. Instead of actively engaging in the learning process, students tend to assume the role of passive listeners. According to DeCaro et al. (2023), several studies showed that this can lead to challenges in accurately assessing the depth of acquired knowledge, since students can overestimate their gained learning outcome during traditional teacher-based instructions. Nevertheless, direct instructions can also demonstrate examples, or schemas of the learning content, and convey background knowledge. Yet, the optimal integration of PS phases into instructional sequences remains uncertain.

Against this background, questions have arisen on how to combine PS and DI to acquire knowledge and reduce cognitive load. Two dominant instructional designs have emerged from years of discussions regarding their effectiveness (Gorbunova et al., 2023; Prince & Felder, 2006; Tobias & Duffy, 2009): (1) the deductive instructional design 'direct instruction followed by problem-solving' (DI-PS) and (2) the inductive instructional design 'problem-solving followed by instruction' (PS-DI). Still, empirical findings regarding these designs are inconsistent (Chen & Kalyuga, 2020). While many studies, especially in STEM, show that DI-PS supports procedural knowledge (e.g., Loibl et al., 2020), other studies showed that PS-DI enhances conceptual knowledge (e.g., Kapur, 2014) and others found no difference in learning outcome (Van Gog et al., 2011). Moreover, research in the business field has not studied the effect of both instructional designs in combination with problem-solving processes and CL. To close this gap, this study aims to understand the effects of different instructional designs on problem-solving performance and knowledge acquisition in the business context. Thus, the study contributes to the field by providing insights into effective instructional strategies that can enhance learning and problem-solving skills in business education. Additionally, the study sheds light on the problem-solving processes and cognitive load of learners with a business background, which can support the development of more effective educational interventions.

We address the following research questions.

1) How do instructional designs (DI-PS vs. PS-DI) and instructional preferences influence a) problem-solving performance and b) knowledge acquisition?
2) How do instructional designs (DI-PS vs. PS-DI) influence a) CL and b) problem-solving processes?
3) How does problem-solving behaviour predict both problem-solving performance and knowledge acquisition within the DI-PS and PS-DI groups?

2. Theoretical background

2.1. Instructional design (DI-PS and PS-DI) and instructional preferences

Regarding the sequencing of more self-regulated and more receptive phases, two instructional designs have been widely discussed: (1) the deductive instructional design 'direct instruction followed by problem-solving' (DI-PS) and (2) the inductive instructional design 'problem-solving followed by instruction' (PS-DI) (Gorbunova et al., 2023; Prince & Felder, 2006). In DI-PS, learners acquire the necessary task-related schema to complete the task by first receiving instruction or worked examples (WE) (Sweller, 1994). Moreover, learners pay attention to relevant information before solving a problem (Kirschner et al., 2006). Hence, prior lectures, video tutorials, or worked examples (DI) can reduce cognitive load. Many studies on PS-DI were conducted within the conceptual framework of 'productive failure' (PF) (e.g., Kapur, 2012, 2014) that is motivated by Bruner's concept of 'discovery learning' (1961). PF encourages students to engage in problem-solving tasks before receiving DI, leading to initial failures (Hartmann et al., 2021). Under the PS-DI condition, learners can work on an either unsupported or supported problem, followed by instruction in the subsequent phase. Learners can benefit from improved transfer of learning and long-term retention since prior knowledge is activated (e.g., Kalyuga & Singh, 2016; Kapur & Bielaczyc, 2012), and learners are made aware of their knowledge gaps during the PS phase (Loibl & Rummel, 2014b).

Regarding the possible cognitive mechanisms in favour of PS-DI settings, Sinha & Kapur (2021), based on further literature, conclude that the advantages of PS-DI might be due to prior knowledge activation, knowledge gap awareness, deep feature recognition, and cognitive activation in general. In favour of direct instruction, Eiriksdottir and Catrambone (2011) outline that procedural instruction that explains how to complete certain tasks step by step is particularly helpful for the immediate performance of the same task. Providing information in the form of more abstract principles aims at a deeper understanding, problem solving in the domain, and transfer of learning. In conclusion, they suggest a combination of the three types of direct instruction including procedural information, principles, and examples. Based on goal achievement theory (Dweck, 1986), the performance goal orientation decreases in the face of obstacles that are very likely under the PS condition. 'In contrast, it is hypothesised that learning goals, which focus individuals on increasing their ability over time, will promote the mastery-oriented response to obstacles: strategy formulation, positive affect, and sustained performance' (Elliott & Dweck, 1988, p. 5).

It is not only the cognitive processes that are different, but also the learning outcomes in terms of cognitive representations that result from DI vs. PS. While both instructional designs aim to support knowledge acquisition, they likely result in different internal knowledge structures. In DI-PS, learners are typically exposed to step-by-step explanations and worked examples that aim to build accurate and complete procedural schemas (Sweller, 1994) or production rules (Anderson, 1993) that can be readily applied during subsequent problem solving. This may facilitate performance on near-transfer tasks when similar problems are encountered. According to Duncker (1945), a problem is defined as a situation in which one lacks knowledge of how to reach a concrete goal. From this perspective, one might even question whether DI-PS settings involve genuine problem solving at all—since learners are equipped with the necessary knowledge in advance during the instructional phase. In contrast, in PS-DI, learners initially explore the problem, experience confusion, and encounter failure, which activates possibly incomplete or naïve prior knowledge. This initial problem-solving phase leads to the formation of episodic traces of problem-specific solution approaches, which – although incorrect and incomplete – form the basic representation of the problem structure. Instruction then allows learners to reframe or revise their earlier, often flawed mental models, integrating accurate principles with their experiential traces. This integrative process supports deeper conceptual understanding. Cognitive representations are supposed to be rich, well-connected, and adaptive, particularly for far-transfer tasks (Kapur, 2016; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014b). Therefore, we focus on two knowledge types that are often discussed and refer to the above differences in internal knowledge structures (Anderson & Krathwohl, 2001; McCormick, 1997): *conceptual knowledge* (e.g., the understanding of principles and relationships of connected concepts, principles or categories) and *procedural knowledge* (e.g., the knowledge of a process and knowledge of applying rules, algorithms, techniques or a sequence of actions to perform procedure operations). According to Loibl et al. (2020), several studies demonstrate that learners who receive instructions before solving a problem (DI-PS) tend to acquire more procedural knowledge than students who do not receive prior instructions (e.g. Rittle-Johnson et al., 2001). Contrarily, studies on PS-DI demonstrate that learners achieve superior conceptual understanding and transfer outcomes compared to those in DI-PS conditions (e.g., Kapur, 2016; Loibl et al., 2017).

Furthermore, individual instructional preferences can also affect learning outcomes, as some learners may prefer the PS-DI condition and may benefit more from a PS-DI condition. In contrast, other learners may favour and benefit from a DI-PS condition. Considering, for instance, an everyday problem scenario, like adjusting a thermostat (Hahnel & Stemmann, 2023), individuals may approach the problem in different ways. Some individuals may first study the manual before engaging in exploratory problem solving, while others may opt for the reverse sequence. However, it is essential to highlight that the preferred sequence may not always lead to better learning outcomes, since learners tend to prefer instructional approaches ' … that involve the least amount of cognitive effort' (Eiriksdottir & Catrambone, 2011, p. 756). Instead, students tend to underestimate their learning outcomes in more self-directed settings, while overestimating learning outcomes in more traditional, direct instruction (Deslauriers et al., 2019). However, it can be assumed that students' perceived satisfaction with the instructional design is higher if their instructional preferences are met.

Lastly, instructional preferences may be influenced by cultural background. For instance, DI-PS — representing the "principle-first approach"— is more prevalent, for instance, in Germany. Typically, in German university lectures, students receive clear and concise instructions and explanations with general or abstract concepts, theories, rules, and procedures before practicing. In contrast, the PS-DI — representing the "application-first approach"— is more commonly found, for instance, in the US, reflecting potential cultural differences in preferred instructional sequencing (Meyer, 2014). US American students are sometimes expected to apply concepts and skills before discussing them in class. Research has shown "cultural differences in the complex problem-solving process" in that German participants were keener to plan (Smith et al., 2022). This is confirmed by Strohschneider and Güss (1998, as cited in Smith et al., 2022), who stated that German students showed a higher tendency to plan when facing a problem than (South) American students.

2.2. Self-regulated learning in problem-based learning environments

Problem-based learning environments require self-regulated learning (SRL) skills (Greene et al., 2011; Zhang et al., 2022). SRL represents the processes and strategies individuals use to autonomously manage their learning by engaging in processes for monitoring and controlling their own learning, fostering responsibility when navigating through complex problems in computer-based learning environments and developing solutions to challenging tasks (Winne & Azevedo, 2022). SRL involves cognitive, affective, metacognitive and motivational processes (e.g., Azevedo et al., 2017) enabling learners to actively plan, set sub-goals, organise information, monitor their own progress, and adapt their strategies accordingly (Winne & Hadwin, 1998; Zimmerman, 2000). When applying strategies (such as note taking, reading, or rereading documents, etc.), learners build mental models, acquire new knowledge, and apply their knowledge to solve complex tasks successfully (Taub & Azevedo, 2019). Moreover, SRL is influenced by learner characteristics such as prior knowledge (Taub et al., 2014; Taub & Azevedo, 2019 and cognitive ability (Sternberg, 1997).

2.3. Cognitive load in simulation-based learning

Cognitive Load Theory (CLT) describes how problems and corresponding learning material should be designed for the specific learner group to keep (extraneous) cognitive load low and enhance self-regulated learning processes and problem solving (Klepsch & Seufert, 2020; Sweller et al., 1998). CLT is a framework in educational psychology that is used to examine learning by investigating how the human brain stores and processes information (Sweller et al., 1998). Human working memory, responsible for temporary information storage and processing, has a limited capacity and can process only a limited number of cognitive elements (Sweller et al., 1998). Element interactivity describes the extent to which the elements to be processed are interrelated and cannot be processed successively and in isolation from each other. This element interactivity causes cognitive load (Chen et al., 2019; Sweller et al., 1998). To mitigate this, humans organise information into schemas, reducing cognitive load by linking new information to existing knowledge. In CLT, cognitive load had originally been divided into three categories: intrinsic, extraneous, and germane load (Sweller et al., 1998), while more recent publications only differentiate between intrinsic and extraneous load (Kalyuga, 2011; Sweller, 2010). Intrinsic cognitive load (ICL) relates to the intrinsic complexity (i.e., the element interactivity) of the material that has to be learned, regardless of the design of the learning environment. In contrast, extraneous cognitive load, on the other hand, results from a non-optimal design of the learning environment, for instance due to the cumbersome presentation of information or the inclusion of irrelevant information (Sweller, 2010). However, this is often the case in simulation-based learning environments, since learning goals include distinguishing relevant from irrelevant information and making sense of information that is presented in a cumbersome way. Unsurprisingly, advocates of simulation-based learning environments criticise the resulting denial of exploratory learning environments that resonates in early CLT publications (Plass & Schwartz, 2014). However, "the same information may impose an intrinsic or an extraneous cognitive load depending on what needs to be learned" (Sweller, 2010, p. 125). Thus, if the learning goals include distinguishing relevant from irrelevant information and making sense of information that is presented in a cumbersome way, as is often the case within problem-based learning, the resulting element interactivity is intrinsic. Still, less experienced problem solvers with less prior knowledge can be cognitively overwhelmed by these complex learning environments (Fischer et al., 2022). To reduce the intrinsic cognitive load, parts of these learning goals could be reduced or abandoned, while extraneous cognitive load can be reduced through instructional design (Sweller, 1994, 2010; Sweller et al., 1998), for instance, through various kinds of scaffolding (Fischer et al., 2022; Rogers & Franklin, 2021). Furthermore, previous direct instruction can help to ensure a minimum of necessary prior knowledge in order to successfully master the more complex problems in a simulation-based environment.

2.4. Related research on the effects of DI-PS and PS-DI

2.4.1. Effects of DI-PS on knowledge acquisition

Several studies investigated the effects of DI-PS and PS-DI on cognitive load, problem solving and knowledge acquisition (Ashman et al., 2020; Loibl et al., 2020). Ashman et al. (2020) investigated fifth-grade primary school students working on a light energy efficiency task. DI-PS was more effective for problem solving compared to PS-DI. Additionally, DI-PS provided a performance advantage on transfer tasks, especially in problems involving high element interactivity. Referring to CLT, the PS-DI condition caused cognitive overload (Ashman et al., 2020). Several other studies used worked examples as an instructional approach that resembles DI (Leppink et al., 2014; Van Gog et al., 2011). For instance, Van Gog et al. (2011) investigated the effect of three instructional designs *worked examples only (WE)*, *example-problem pairs (WE-PS)*, *problem-example pairs (PS-WE)* and compared them to *problem solving only (PS)* on novices' learning. A total of 103 secondary education students completed the electrical circuit troubleshooting tasks. The instructional designs WE-PS and WE had a higher learning effectiveness than the instructional designs PS-WE and PS. Additionally, the WE-PS and WE groups showed significantly lower cognitive load. Leppink et al. (2014), investigated the four instructional designs *example-example (E-E), example-problem (E-P), problem-example (P-E), and problem-problem (P-P)* in the field of mathematical and statistical problems. They observed higher learning outcomes in E-E and E-P groups compared to P-P and P-E pairs. Moreover, the investigation into cognitive load surprisingly showed that the four different conditions did not differ in intrinsic and extraneous cognitive load (maybe due to a short learning phase or the benefits of worked example). The researchers also confirmed their hypotheses by showing that the E-E and E-P conditions increase germane cognitive load in the studying post-test phases compared to the other conditions. Moreover, the researchers reported a negative correlation between intrinsic cognitive load and exam performance.

2.4.2. Effects of PS-DI on knowledge acquisition

Other studies showed positive effects of PS-DI on knowledge (e.g., DeCaro et al., 2023; Kapur, 2014; Loibl & Rummel, 2014a). For instance, Kapur (2014) investigated the effects of both PF-DI and DI-PF on the learning of ninth-grade mathematics students in a paper-pencil setting. Within the PF phase, students solved a problem on standard deviation individually without any help. During the DI phase, the teacher presented worked examples explaining the concept of standard deviation to the students. Kapur (2014) showed that both instructional designs (DI-PF and PF-DI) lead to higher student procedural knowledge. However, students that were assigned to the PS-DI group showed higher scores in conceptual knowledge as well as in a transfer test compared to the DI-PS group despite a higher perceived cognitive load. Loibl and Rummel (2014a) examined three conditions: problem-solving prior to instruction (PS-I), guided problem-solving prior to instruction (PS_{guided}-I) and direct instruction (I-PS) in the field of mathematics. Tenth-grade students in secondary schools without prior knowledge on the mathematical concept “mean absolute deviation” participated in two quasi-experimental studies (Loibl & Rummel, 2014a). Loibl and Rummel (2014a) reported a positive effect in favour of PS-I on *conceptual knowledge*. Students in the I-PS condition outperformed PS-I and PS_{guided}-I conditions in *procedural knowledge* (only marginally significant difference in the first study, but not in the second study).

A recent study such as by DeCaro and colleagues' (2023) shows the effects of both instructional designs on procedural and conceptual knowledge of 78 undergraduate physics students in online courses. The DI-PS group was asked to use the mathematical equation learned in the instruction while the PS-DI group was asked to invent a mathematical equation. DeCaro and colleagues (2023) implemented a 2x2 design (condition: instruction-first and exploration-first) and found that PS-DI fosters students' procedural and conceptual knowledge compared to the DI-PS approach. Only learning activities were higher in the DI-PS group. To sum up, DeCaro et al. (2023) stress that the instruction order matters in online settings.

2.4.3. Effects of PS-DI vs. DI-PS on knowledge acquisition

Other studies did not find any differences between instructional designs in knowledge acquisition. Fässler et al. (2022), for instance, investigated the effect of declarative and conceptual knowledge in a medical simulation with students of a third-year medical course assigned to either the ‘instructions prior to problem-solving in computer-based virtual environment’ group (I-CVE) or ‘problem-solving in computer-based virtual environment prior to instruction’ group (CVE-I). No significant differences in knowledge acquisition and cognitive load between the groups were found. Also using medical simulations, Kulasegaram et al. (2018) found that discovery learning before direct instruction had no effect on immediate knowledge acquisition but improved transfer performance in further simulation-based training. Schalk et al. (2018) investigated learning transfer (assessed after four weeks) of Swiss ninth graders who worked on a mathematical problem. The students were assigned to either ‘tell and practice’ (T&P; equal to DI-PS) or one of four PS-DI instructional designs (some PS-DI conditions involved prompts). Their results demonstrated that the instructional designs themselves were not crucial for the effectiveness of the problem solving, but the prompts were influential. Students in T&P were disadvantaged in learning transfer compared to students who were prompted to develop general solutions to cases without any contextual background.

We outlined studies here that investigated the effects of instructional designs on problem solving, learning and some of them also cognitive load. However, their investigations were carried out under different conditions like different domains, environments (also paper-pencil setting), or paid less attention to problem-solving processes (process data) and instructional preferences.

2.5. Expected learning processes and performance patterns

Despite the more exploratory nature of the study, based on the above theoretical arguments as well as prior research, and considering the characteristics of the present business simulation, we anticipate distinct learning processes and corresponding performance patterns in the two groups. When learners receive instruction before the problem-solving phase (DI-PS), the instructional video is expected to support the acquisition of procedural knowledge and to reduce the need for exploration This may result in more accurate applications of domain-specific algorithms (e.g., calculating purchase costs), more precise, domain-specific justifications (e.g., articulating arguments for and against potential suppliers) and lower cognitive load during the problem-solving phase. In contrast, beginning with problem solving (PS-DI) is expected to trigger knowledge-gap awareness, deeper semantic processing, and the integration of conceptual relations during the subsequent instruction phase. Accordingly, this sequence may manifest in a more frequent use of exploratory learning resources and higher cognitive load during the problem-solving phase but may also lead to higher conceptual knowledge gains after the consolidation of the initial problem exploration during the direct instruction phase. Based on these considerations, the empirical analyses investigate the relationships between DI-PS and PS-DI, learners' instructional preferences, cognitive load, and problem-solving and knowledge-acquisition outcomes in the computer-based office simulation.

3. Method

3.1. Participants and procedure

A pre-post-test experimental study was conducted involving 81 undergraduate German students of business and economic education (age: M = 22.2 years; SD = 2.8; 71.6 % female; average semester completed = 3.2 semesters; SD = 2.0). The participants were randomly assigned to one of the instructional designs, either DI-PS or PS-DI (see Fig. 1). The study comprised four phases. After a short introduction by the study instructors (phase 1), both groups completed questionnaires (demographics and instructional preferences) and a pre-test to assess prior knowledge of supplier selection (phase 2). In phase 3, all participants worked on a 4-min onboarding to familiarise themselves with the web-based office simulation (Fig. 2). Participants of the DI-PS group first watched a video scenario (DI; see Fig. 3) before they engaged in a problem-solving scenario (PS, see Fig. 4). In reverse, participants of the PS-DI group started with the problem-solving task followed by the video. In phase 4, all participants completed a post-test to assess their knowledge acquisition. Finally, participants completed two additional questionnaires.

All questionnaires and scenarios were integrated into the office simulation LUCA. The LUCA office simulation is a browser-based office simulation that was developed as a learning and assessment environment in business contexts. LUCA provides learners with typical office tools such as an email client, a file system, a PDF viewer, a spreadsheet application, a calculator, a notepad, and a reduced enterprise resource planning (ERP) system (see Fig. 2). An integrated authoring

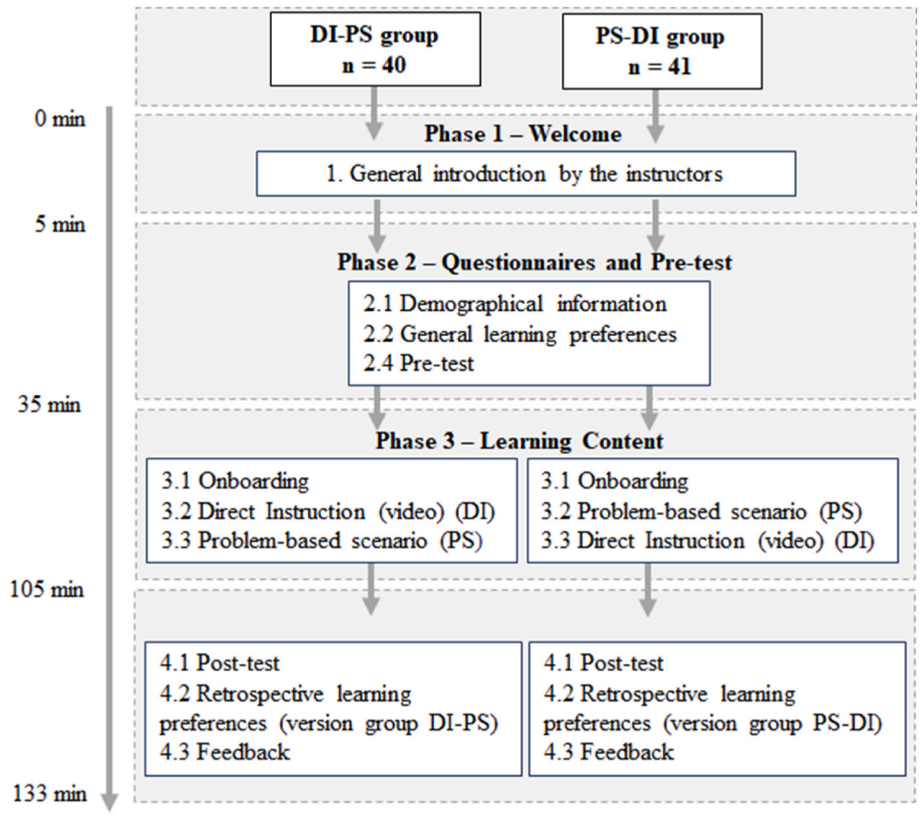


Fig. 1. Study procedure.

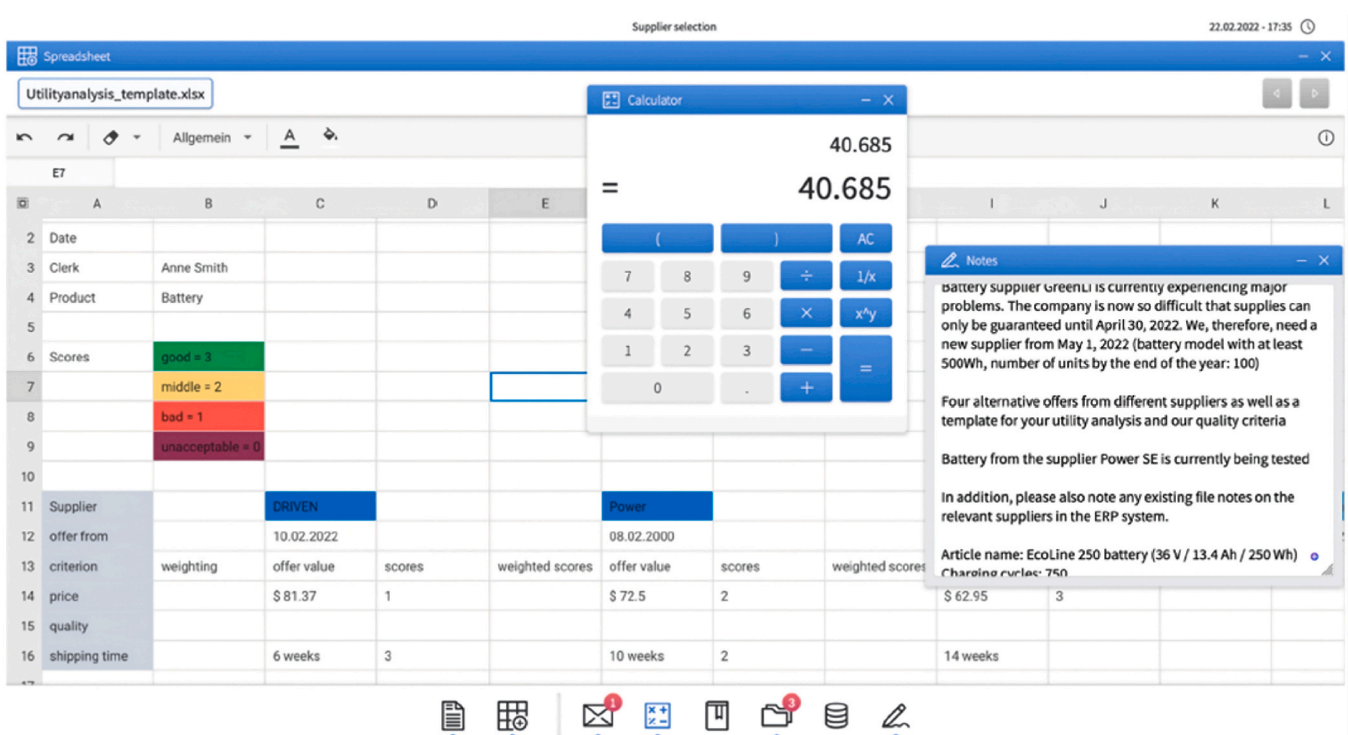


Fig. 2. Screenshot of LUCA Office Simulation showing spreadsheet, notepad, and calculator.

environment allows for developing and modifying authentic work scenarios for learning or assessment purposes. Furthermore, questionnaires and traditional test items were also implemented. To ensure compliance, the students were placed under individual observation. They could not influence the order of the phases and were explicitly instructed about permitted and non-permitted actions.

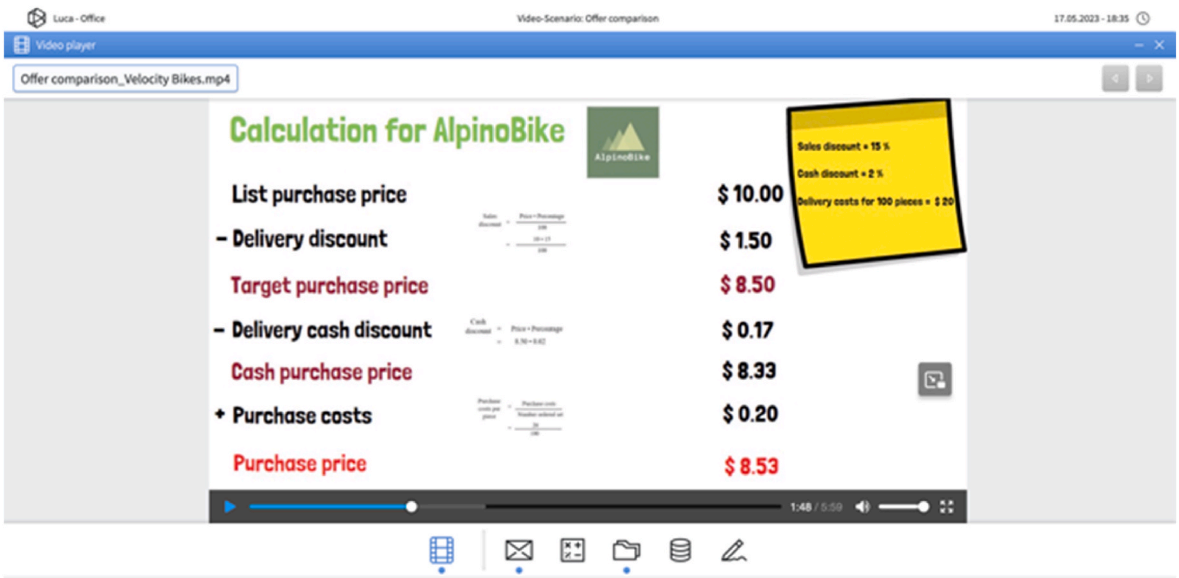


Fig. 3. Screenshot of the direct instruction (DI; video scenario) presenting quantitative and qualitative criteria.

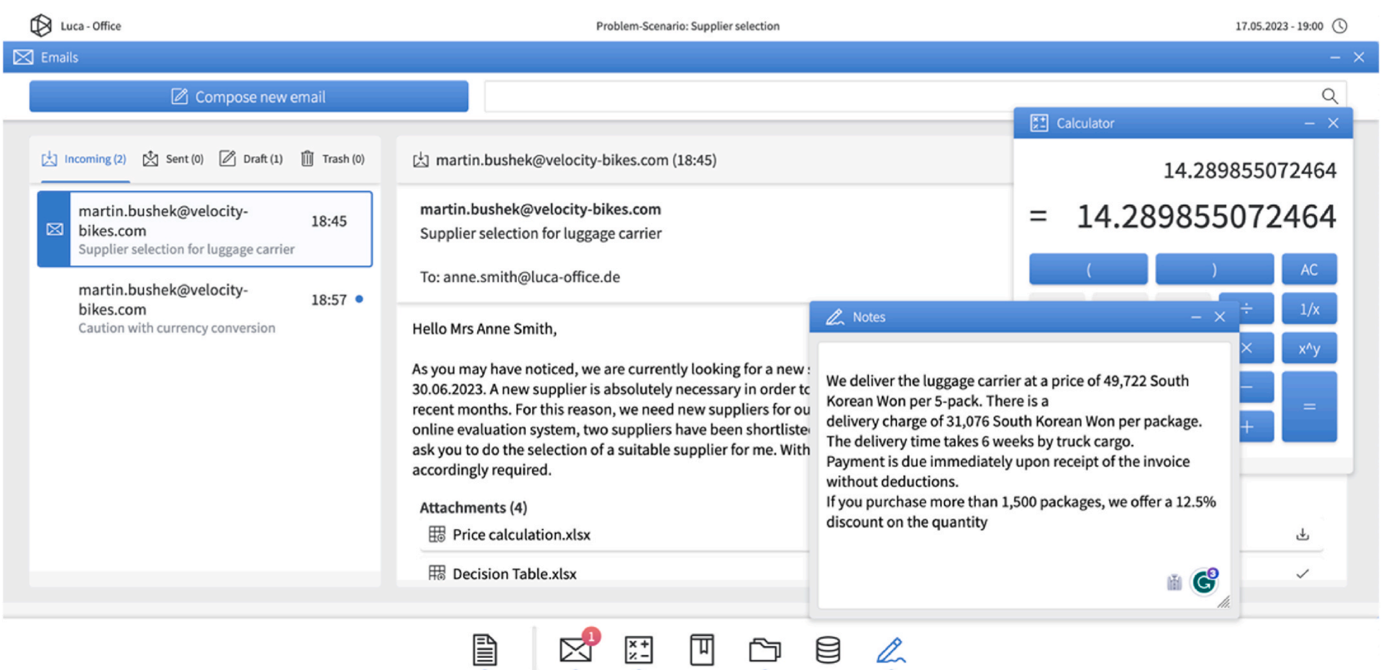


Fig. 4. Screenshot of the office simulation and problem-solving scenario (PS) showing the email client with the task assignment, a notepad, and a calculator.

3.2. Learning material

3.2.1. Learning context

Supplier selection is a central and typical higher-order task in the business domain. Successful mastery involves identifying relevant information, understanding subject-specific information, and identifying information requirements and action steps. Furthermore, quantitative data must be processed using typical tools (here: spreadsheets) and qualitative data must be quantified using specific procedures (purchase price calculations and multi-attribute utility analysis). Finally, a well-founded decision (here: selection of a supplier) must be made and adequately communicated. The challenges are structurally similar to many other 'big ideas' from the business management domain (e.g., personnel selection). In simulation-based competence assessments, data from supplier selection scenarios and other scenarios could be scaled to overarching dimensions. Supplier selection can therefore be considered a representative problem.

3.2.2. DI phase

The DI phase included a short video scenario (5.36 min) with a step-by-step instruction on supplier selection and a worked example (Fig. 3). The video scenario involved a multi-attribute utility analysis to integrate quantitative criteria (i.e., purchase prices) and qualitative criteria (e.g., delivery time, product quality). Students could also pause the video and could rewind and fast-forward.

3.2.3. PS phase

During the PS phase, the students worked in an office simulation that provides typical office tools such as an email client, file system, PDF viewer, spreadsheet program, calculator, notepad, and a reference book/wiki (Fig. 4). In the 40-min problem-based scenario, the participants first received an email with the task of conducting a supplier selection. To solve the problem, participants had to read two supplier offers and additional information regarding the product quality and delivery sustainability. Based on the information provided in the offers, they calculated purchase prices for the suppliers and evaluated the suppliers in a decision table (multi-attribute utility analysis). Additional information regarding currency exchange or a reference book with an explanation of supplier procedure was provided. The participants ware requested to come up with a well-founded decision in favour of one of the two suppliers and provide an argumentation in an e-mail reply. The scenario ends when the decision is sent via email, regardless of whether the choice was correct. Participants were not provided with the correct solution.

3.2.4. Measures

Problem-solving performance. In assessing *problem-solving performance (RQ 1a)*, the students' responses and calculations in the problem-based scenario (PS) were evaluated by two trained raters based on a domain-specific competence model with four scoring criteria and eleven sub-categories (see Table 1).

The problem-solving performance evaluation achieved an average *Intraclass Correlation Coefficient (ICC)* of .83 (see Appendix A for further details on coding). Regarding the content of the direct instruction (video scenario), we expected positive effects on the identification of necessary action steps (1.1), the application of algorithms (2.2), the correct decision (3.1) and correct reasoning (3.2, 3.3). Furthermore, positive effects were expected for the reflection of the procedure (3.5) but no effects were expected on the opening of documents (2.1), the use of tools (2.3), the estimated quality of one's own action (3.4) or communication (4.1, 4.2) since none of these performances were addressed in the video. However, to provide the full picture, t-tests were calculated for each sub-category despite an alpha error accumulation. More important, Cohen's d was calculated to assess effect sizes, with the thresholds of .2 for small, .5 for medium, and .8 for large effects (Cohen, 1988, 1992).

Conceptual and procedural knowledge acquisition. To investigate *conceptual and procedural knowledge acquisition* (RQ 1b), differences between pre- and post-test scores (gain scores; Δ) were calculated. Each test consisted of 12 items with a binary scoring (0/1; see Appendix B for exemplary items). Seven items aimed at measuring conceptual and five items aimed at procedural knowledge. The Kuder Richardson 20 score is .57 for the pre-test and .50 for the post-test. The items in the pre-test and post-test were (almost) the same but presented in a different order. Furthermore, in the case of arithmetic tasks, they were based on different numbers and in case of multiple-choice tasks contained response options in a varying order. This approach made it possible to use gain scores as indicators for knowledge acquisition. However, this involved risks regarding possible memory effects. To mitigate these effects, no feedback regarding correct or incorrect answers was provided to participants. Still, possible downsides of our approach to measuring knowledge acquisition are discussed in the final section of the paper.

Instructional preference and retrospective satisfaction with the instructional sequencing. The *instructional preference* was measured in the initial questionnaire with a 13-items scale. The scale shows an acceptable Cronbach's alpha of .74 (Nunnally & Bernstein, 1994). Overall, it can be stated that both groups show an instructional preference towards DI-PS ($M_{DI\text{-}PS} = 2.61$, $SD_{DI\text{-}PS} = .35$; $M_{PS\text{-}DI} = 2.72$, $SD_{DI\text{-}PS} = .38$, $p = .31$, Cohen's $d = .11$; 1 = Do not agree at all and 4 = Fully agree; see Appendix C for exemplary items). Retrospective satisfaction with the instructional sequencing was measured in the final questionnaire with one item asking students whether they appreciated the chosen sequencing (DI-PS or PS-DI), whether they would have preferred the opposite order, or whether the sequence did not matter to them. In line with the instructional preferences surveyed at the beginning, 80 % of the DI-PS group and 83 % of the PS-DI group stated that they (would have) preferred DI-PS.

Cognitive load. To examine *cognitive load (RQ2a)*, participants responded to a scenario-embedded experience sampling (EES, see Appendix D.1) during the scenario and answered questions after the scenarios (5 items, 4-point Likert scale). We measured two types of intrinsic cognitive load: (1) intrinsic in-scenario cognitive load (IICL) and (2) intrinsic retrospective cognitive load (IRCL). The ICL was measured based on a one-item Embedded Experience Sampling (EES) situated within the problem-solving phase; Cronbach's alpha for the three-item scale of RCL with .75 is classified as acceptable. Moreover, we measured extraneous cognitive load based on one-item measure (see Appendix D.2). The formulation of the items is similar to those used in other instruments for measuring cognitive load (Klepsch et al., 2017; Krieglstein et al., 2023).

3.2.5. Data analysis

For answering RQ 1, mixed ANOVAs were used to assess the impact of both instructional designs (DI-PS vs. PS-DI; categorical and independent variable) and pre- and post-tests (repeated measures) on conceptual and procedural knowledge gains (dependent variables), allowing for the examination of main effects and interactions between factors. Moreover, an analysis of covariance (ANCOVA) serves to assess whether there are any significant differences in the learning outcomes (dependent variable) between the instructional design conditions (independent variable), while statistically adjusting for the influence of the instructional preference (covariate). This helps to minimise the impact of potential confounding variables and provides a more accurate assessment of the relationship between the independent variable and the dependent variable. Requirements for the ANCOVA (such as variance homogeneity and independence of the predictor from the covariate) were tested. For the ANCOVA, we used the R packages *car* (Fox et al.,

Table 1
Scoring criteria and sub-categories used to assess problem-solving performance.

Scoring Criteria	Sub categories	Description
1. Planning	1.1 Identifiable action steps	Necessary problem-solving steps are recognisable (regardless of their correctness)
2. Application	2.1 Information processing: Open documents	Relevant documents have been opened at least once.
	2.2 Application of algorithms	The necessary calculations were performed correctly (e.g., purchase price calculations).
	2.3 Use of tools	Useful functions were utilised (e.g. cell expansion in the spreadsheet).
3. Decision-making	3.1 Correct decision for supplier	The supplier with the highest utility score was selected.
	3.2 Correct reasoning (Supplier 1: Frenos X-Treme)	The correct advantages and disadvantages for this supplier were determined on the basis of the documents available.
	3.3 Correct reasoning (Supplier 2: Dishan Yibei Bikes)	The correct advantages and disadvantages for this supplier were determined on the basis of the documents available.
	3.4 Reflection regarding Action	Statements about the assessment of the quality of one's own solutions.
	3.5 Reflection regarding Procedure	Statements about the assessment of the quality of the procedure of supplier selection (very rare).
4. Communication	4.1 Politeness and appropriate communication standards	Coding based on standard etiquette in internal company communications.
	4.2 Absence of formal errors	Coding based on spelling and grammar rules.

2012) and *rockchalk* (Johnson, 2022). Additionally, t-tests were conducted to determine significant differences between the two groups' means.

To explore *problem-solving processes* (RQ2b), log data (including mouse clicks and keystrokes) were collected during the problem-solving (PS) scenario. Similar subsequent actions were then grouped, and behavioural indicators (e.g., note-taking or reading documents) were derived (see Table 2). In addition, we conducted regression analyses to examine the relationship between behavioural indicators and performance (RQ3).

4. Results

4.1. Effects of instructional designs on problem-solving performance (RQ1a)

Table 3 shows that the DI-PS group obtained a higher average problem-solving performance in the problem-solving scenario than the PS-DI group ($M_{\text{DI-PS}} = 22.13$ vs. $M_{\text{PS-DI}} = 19.24$). However, the *t*-test revealed no significance ($p = .08$). Having a closer look, the DI-PS group outperformed the PS-DI group in 10 out of 12 problem-solving performance evaluation criteria (see mean scores). However, the t-tests revealed that the instructional designs only significantly impacted problem-solving performance for two items: for both evaluation criteria '3.2 correct reasoning supplier 1' and '3.3 correct reasoning Supplier 2'. We found significant differences ($p = .01$ and $p = .01$, respectively) and large effect sizes ($d = .60$ and $d = .64$ respectively). This means that the DI-PS group significantly outperformed the PS-DI in arguing for or against the specific suppliers. Moreover, we found medium-sized effects for '2.2 Application of algorithms' and '3.5 Reflection regarding procedure' ($d = .42$ and $d = .49$) meaning that DI-PS outperformed the PS-DI group in spreadsheet calculations and in reflecting their problem-solving procedure. There are no differences in the scoring criteria '1. Planning' and '4. Communication'.

4.2. Effects of instructional designs on knowledge acquisition (RQ1b)

Descriptive analyses reveal that the Group DI-PS exhibited lower overall scores in both the pre-test ($M_{pre} = 5.38$; $SD_{pre} = 2.20$) and post-test ($M_{post} = 5.70$; $SD_{post} = 2.34$) compared to the Group PS-DI ($m_{pre} = 5.70$ and $m_{post} = 7.71$). Hence, the differences (Δ) between the pre- and post-test scores indicate higher knowledge acquisition for PS-DI in total ($\Delta M_{DI\text{-}PS} = 1.74$; $\Delta M_{PS\text{-}DI} = 2.01$). However, there were no significant differences. For the two different knowledge types, conceptual knowledge (Fig. 5) and procedural knowledge (Fig. 6), measured with the pre- and post-test, show differences between pre- and post-test scores across both groups in the conceptual knowledge acquisition (Δ) ($\Delta M_{Post\text{-}Pre\ DI\text{-}PS} = .92$, $\Delta SD_{Post\text{-}Pre\ DI\text{-}PS} = 1.61$ and $\Delta M_{Post\text{-}Pre\ PS\text{-}DI} = 1.54$, $\Delta SD_{Post\text{-}Pre\ PS\text{-}DI} = 1.7$) and differences in procedural knowledge acquisition (Δ) ($\Delta M_{\text{Post-PreDI-PS}} = .78$, $\Delta_{Post\text{-}Pre}SD_{DI\text{-}PS}=1.01$ and $\Delta M_{Post\text{-}Pre\ PS\text{-}DI} = .29$, $\Delta SD_{Post\text{-}Pre\ PS\text{-}DI} = 1.16$). T-tests did not show significant differences in both knowledge types between the groups.

A mixed ANOVA reveals significant differences between pre- and post-tests *within* the groups meaning that both groups have a positive learning outcome through the applied treatments (learning material). Moreover, significant procedural knowledge differences between pre- and post-tests can also be found *within* the groups meaning that both groups have a positive learning outcome through the applied treatments (learning material). An interaction effect was also present but was not significant, meaning that both groups do not differ over time (here: pre- to post-test) (see Table 4).

An ANCOVA was conducted to assess the impact of instructional design on problem-solving performance with intrinsic cognitive load during the scenario serving as the covariate (Table 5). The analysis only showed a significantly negative effect of cognitive load on problem solving competence.

Moreover, two ANCOVAs were conducted to assess the impact of instructional design on conceptual knowledge acquisition (Table 6) and procedural knowledge acquisition (Table 7) with instructional preferences serving as the covariate. The analysis showed a significant effect of the combined variables on conceptual knowledge acquisition. The beta-coefficient of .19 indicates a negative relationship, meaning that the DI-PS group may be associated with a decrease in conceptual knowledge compared to the PS-DI group. The negative beta-coefficient of −.21 suggests that preferences towards the implemented DI-PS instructional design might negatively affect the conceptual knowledge gain. The positive beta-coefficient of .09 indicates that the direction of the instructional design's impact on conceptual knowledge gain is minimally affected by the preferences.

An ANCOVA was conducted to assess the impact of instructional design on procedural knowledge acquisition with instructional preferences being the covariate (Table 7). The analysis showed a significant effect for the combined variables. This suggests that the impact of instructional design on procedural knowledge acquisition was controlled by the individual's preferences. The positive beta-coefficient of .03 implies that procedural knowledge acquisition increases when receiving the preferred instructional design.

4.3. Effect of instructional designs on cognitive load and on problem-solving processes (RQ2)

To investigate the *effects of instructional designs on cognitive load (RQ2a),* students' intrinsic in-scenario cognitive load (IICL), intrinsic retrospective cognitive load (IRCL), and extraneous cognitive load (ECL) were reported for both conditions (see Table 8). The t-tests showed that the groups differ significantly in IICL and ECL. The PS-DI rated the task during the PS scenario itself as more difficult and, hence, showed significantly higher IICL than DI-PS. A moderate effect size for IICL can be reported ($d = .53$). Moreover, the DI-PS group found the scenario significantly more useful ($M = 1.69$) than PS-DI ($M = 2.00$).

In general, there were not many differences in problem-solving processes between both groups (RQ2b; Table 9). They only significantly differ in their activities regarding reference books that contain a document showing the procedure of supplier selection ($M_{DI\text{-}PS} = 1.3$; $SD_{DI\text{-}PS} = 2.89$ and $M_{PS\text{-}DI} = 4.51$; $SD_{PS\text{-}DI} = 4.67$, $d = .83$). We extended our analysis to include behaviour sequences by constructing bi-grams (pairs of two subsequent events or behaviors). However, no additional information or significance emerged from this analysis.

4.4. Predictors for problem-solving performance and knowledge acquisition (RQ3)

Several multiple regression analyses were conducted to test the predictors (behaviour indicators introduced in Table 1 of problem-solving performance and the knowledge acquisition in each instructional design group. The results of the regression analysis revealed that the multitude of predictors could explain neither conceptual (Tables 10

Table 2
Measure of problem-solving processes – Behaviour indicators derived from log data.

Behaviour indicators	Description
Email task	Frequency of opening and reading the relevant email with the task assignment
PDF-Documents	Frequency of opening and reading four relevant documents
Spreadsheet 1	Frequency of opening and calculating in the spreadsheet 1
Spreadsheet 2	Frequency of opening and calculating in the spreadsheet 2
Note taking	Frequency of taking and reading notes
Calculator	Frequency of calculating in the calculator
Reference Book	Frequency of opening and reading the reference book (wiki)
Time on task	Total time for the problem solving task

Table 3
Problem-based performance based on the problem-based scenario and the competence model.

Scoring Criteria	Sub categories	Score/Credits (maximum)	DI-PS M (SD)	PS-DI M (SD)	p-value	Effect size (Cohen's d)
1. Planning	1.1 Identifiable action steps	3	2.73 (0.55)	2.56 (0.67)	.42	.28
2. Application	2.1 Information processing: Open documents	6	5.63 (0.63)	5.51 (0.66)	.82	.19
	2.2 Application of algorithms	3	2.35 (1.14)	1.83 (1.36)	.13	.42
	2.3 Use of tools	2	1.35 (0.95)	1.51 (0.85)	.38	.18
3. Decision making	3.1 Correct decision for supplier	1	0.88 (0.33)	0.85 (0.33)	.93	.09
	3.2 Correct reasoning (Supplier 1)	4	**2.45 (1.36)**	1.68 (1.22)	.01*	**.60**
	3.3 Correct reasoning (Supplier 2)	4	**2.18 (1.22)**	1.39 (1.26)	.01**	**.64**
	3.4 Reflection regarding action	1	0.10 (0.30)	0.07 (0.27)	.94	.11
	3.5 Reflection regarding procedure	1	0.25 (0.44)	0.07 (0.27)	**.03***	**.49**
4. Communication	4.1 Politeness	4	2.93 (1.62)	2.51 (1.57)	.35	.26
	4.2 Absence of formal errors	2	1.30 (0.77)	1.24 (0.75)	.81	.08
Total		**31**	**22.13 (8.59)**	**19.24 (5.93)**	.08	.39

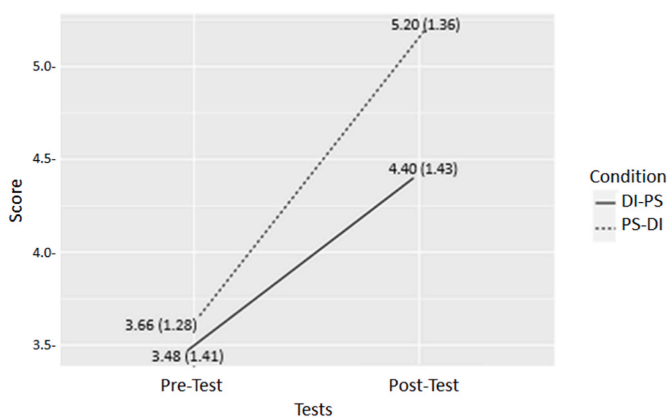


Fig. 5. The effect of the groups (DI-PS vs. PS-DI) on students' conceptual knowledge acquisition (Δ) - pre- and post-test scores across both group.
Note: x-Axis: Pre- and post-test (conceptual knowledge only); y-Axis: Mean Scores achieved in the tests by the groups DI-PS (1) and PS-DI (2).

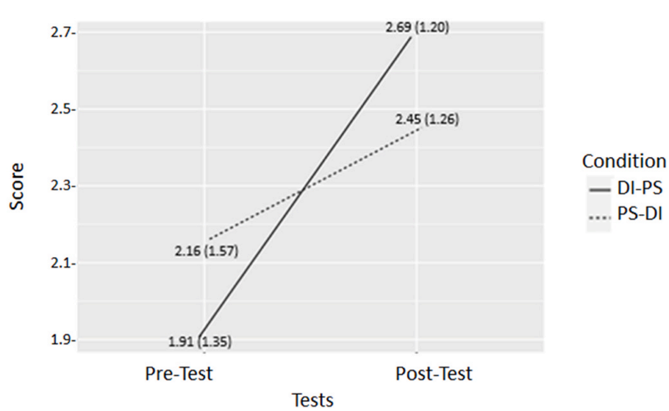


Fig. 6. The effect of the groups (DI-PS vs. PS-DI) on students' procedural knowledge acquisition (Δ) - pre- and post-test scores across both group.
Note: x-Axis: Pre- and post-test (procedural knowledge only); y-Axis: Mean Scores achieved in the tests by the groups DI-PS (1) and PS-DI (2).

and 11) nor procedural knowledge acquisition (Tables 12 and 13). Regarding conceptual knowledge, no predictor had a significant effect on problem solving or the learning outcomes in both groups. Regarding procedural knowledge, 'spreadsheet price calculation' was the only predictor for group PS-DI. Hence, for both knowledge types and for both groups the predictors hardly explained the variance in the learning outcomes (see Tables 12 and 13).

However, this does not hold true for problem-solving performance. A

Table 4
Mixed ANOVA – Conceptual and procedural knowledge acquisition.

	DFn	DFd	F	*p*	partial eta-squared
Conceptual knowledge gain (Δ)					
Group (DI-PS vs PS-DI) (between)	1	79	4.07	.05	.049
Pre-test/post-test (within)	1	79	44.67	.000***	.361
Group x pre-test/post-test (Interaction)	1	79	2.76	.1	.034
Procedural knowledge gain (Δ)					
Group (DI-PS vs PS-DI) (between)	1	61	.001	.976	.000
Pre-test/post-test (within)	1	61	15.34	.000***	.201
Group x pre-test/post-test (Interaction)	1	61	3.22	.078	.050

Note: n = 61; '***' p < .001; '**' p < .01; '*' p < .05; '.' p < .1.

Table 5
ANCOVA for the effects of instructional designs, controlling for cognitive load on problem-solving performance (RQ1b).

	DF	Sum Sq	Mean Sq	F	*p*	beta-coefficient
Instructional Design (DI-PS = 1 vs PS-DI = 0)	1	2.59	2.59	2.95	.09.	.11
Intrinsic cognitive load during scenario	1	6.14	6.14	7.00	.01**	−.29
Instructional Design (DI-PS vs PS-DI condition) x In-scenario Intrinsic Cognitive load (EES)	1	3.32	3.32	3.80	.06.	−.22

Note: n = 73; '***' p < .001; '**' p < .01; '*' p < .05; '.' p < .1.

Table 6
ANCOVA result for conceptual knowledge acquisition (standardised).

	DF	Sum Sq	Mean Sq	F	*p*	beta-coefficient
Instructional Design (DI-PS = 1 vs PS-DI = 0)	1	2.7	2.67	2.85	.09.	−.19
Instructional preferences for DI-PS	1	3.74	3.74	3.96	.05.	−.21
Instructional Design (DI-PS vs PS-DI condition) x Instructional preferences	1	.69	.69	.73	.04*	.09

Note: n = 77, residual standard error = .972; '***' p < .001; '**' p < .01; '*' p < .05.

Table 7
ANCOVA result for procedural knowledge acquisition (standardised).

	DF	Sum Sq	Mean Sq	F	*p*	beta-coefficient
Instructional Design (DI-PS = 1 vs PS-DI = 0)	1	2.29	2.29	2.56	.12	.18
Instructional preferences for DI-PS	1	.56	.56	.63	.43	−.06
Instructional Design (DI-PS vs. PS-DI condition) x instructional design preferences	1	6.34	6.336	7.08	.01*	.03

Notes: n = *59; Residual Standard Error* = *.946;* '***' p < .001; '**' p < .01; '*' p < .05; '.' p < .1.

Table 8
Effects of DI-PS and PS-DI on cognitive load.

	DI-PS M (SD)	PS-DI M (SD)	*p*	Effect size Cohen's *d*
In-scenario Intrinsic cognitive load (Embedded Experience Sampling) (IICL)	2.50 (.80)	2.90 (.79)	.03*	.53
Retrospective intrinsic cognitive load (3-items scale) (RICL)	2.68 (.66)	2.79 (.74)	.59	.16
Extraneous cognitive load (ECL)	1.69 (.80)	2.00 (.73)	.01**	.40

few exhibited statistical significance in relation to problem-solving performance. Notably, the factors 'note-taking' (p = .049) and calculations in both spreadsheets 'Price calculation' and 'Decision Table' (p = .026 and p = .001) emerged as statistically significant predictors, indicating their noteworthy influence on problem-solving for group DI-PS (see Appendix E).

The correlation tables for all variables used in the study are provided in Appendix E. Table E.1 shows the results of the correlation analysis for the DI-PS group. Moderate to large positive correlations were found between the two spreadsheets (price calculation and decision spreadsheet) and problem solving performance. In contrast, moderate negative correlations were found between the two measures of intrinsic cognitive load and problem-solving performance. Table E.2 shows the results of the correlation analysis for the PS-DI group. Moderate to large positive correlations were found between the two spreadsheets (price calculation and decision spreadsheet) and procedural knowledge acquisition. There are also large positive correlations between the spreadsheet decision table and problem-solving performance.

5. Discussion

The present study addressed the effects of instructional designs and instructional preferences on problem-solving performance, knowledge acquisition, cognitive load, and problem-solving processes, in an office simulation.

Regarding the instructional preferences, students expressed a general preference toward DI-PS in the initial questionnaire, which was also apparent in their retrospective indication of satisfaction with the instructional sequencing. However, as is known from research, following students' instructional preferences does not necessarily lead to higher learning outcomes.

Regarding the effect of instructional designs on *problem-solving performance* (*RQ1a),* the DI-PS group showed higher (but not statistically

Table 10
Regression analysis on conceptual knowledge (Group DI-PS).

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	.01	.01	1.09	.28
Spreadsheet Decision table	.00	.01	.12	.90
Note-taking	−.00	.02	−.19	.85
Documents (supplier offers)	−.07	.04	−1.71	.09†
Calculator	.01	.02	.64	.53
Task assignment (email)	.14	.26	.54	.59
Reference Book (Procedure of a supplier selection)	.03	.09	0.3	.77

Note: $R^2 = .19$; adj. $R^2 = .01$ ($df = 32$, $F = 1.07$, $p < .40$); † <.10; * < .05; ** < .01.

Table 11
Regression analysis on conceptual knowledge acquisition (Group PS-DI).

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	−.01	.01	−1.28	.21
Spreadsheet Decision table	−.02	.015	−1.20	.24
Note-taking	−.01	.01	−0.62	.54
Documents (supplier offers)	.06	.04	1.35	.19
Calculator	−.001	.02	−.36	.72
Task assignment (email)	.16	.32	.49	.63
Reference Book (Procedure of a supplier selection)	.05	.06	.75	.46

Note: $R^2 = .20$; adj. $R^2 = .03$ ($df = 33$, $F = 1.2$, $p < .33$); † <.10; * < .05; ** < .01.

Table 12
Regression analysis on procedural knowledge acquisition (Group DI-PS).

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	.001	.01	.20	.85
Spreadsheet Decision table	−.01	0.01	−1.11	.28
Note-taking	.00	.01	.45	.66
Documents (supplier offers)	.01	.03	.30	.77
Calculator	.02	.02	1.16	.25
Task assignment (email)	.03	.18	.15	.88
Reference Book (Procedure of supplier selection)	−.07	.06	−1.09	.28

Note: $R^2 = .10$; adj. $R^2 = -.10$ ($df = 32$, $F = .52$, $p < .82$); † <.10; * < .05; ** < .01.

Table 9
Effects of instructional designs on problem-solving processes (Behaviour indicators).

Behaviour indicators	DI-PS (n = 40) *M (SD)*	PS-DI (n = 41) *M (SD)*	*P* (*p*.adj)	Effect size Cohen's *d*
Spreadsheet Price calculation	65.18 (31.09)	63.10 (45.04)	.862 (.933)	.05
Spreadsheet Decision table	39.15 (21.69)	37.51 (24.60)	.602 (.933)	.07
Note-taking	24.48 (17.37)	23.95 (20.41)	.901 (.933)	.03
Relevant Documents (supplier offers)	17.01 (7.68)	15.76 (7.69)	.451 (.933)	.16
Calculator	13.18 (11.35)	13.39 (11.78)	.933 (.933)	.20
Task assignment (email)	4.29 (1.97)	4.22 (.94)	.719 (.933)	.45
Reference Book (Procedure of a supplier selection)	1.3 (2.89)	4.51 (4.67)	.000 (.003)**	.83
Time on task (in seconds)	2206.78 (291.84)	2284.54 (264.66)	.21 (.933)	.28

Note: Behaviour indicators are in frequency; exception ‚time on task' is measured in seconds.

Table 13
Regression analysis on procedural knowledge acquisition (Group PS-DI).

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	.01	.00	3.17	.00**
Spreadsheet Decision table	.00	.01	.47	.64
Note-taking	−.00	.01	−.051	.96
Documents (supplier offers)	.01	.03	.28	.78
Calculator	.00	.02	.02	.98
Task assignment (email)	−.39	.22	−1.78	.08.
Reference Book (Procedure of a supplier selection)	.07	.04	1.66	.11

Note: $R^2 = .40$; adj. $R^2 = .27$ ($df = 33$, $F = 3.11.$, $p < .01$); . <.10; * < .05; ** < .01.

significant) problem-solving performance than the PS-DI group. According to the CLT (Sweller, 1994), using the provided video as a DI before the problem solving (DI-PS) could have resulted in higher prior knowledge, which may have led to reduced cognitive load and higher problem-solving performance.

Regarding the effect of instructional designs on *knowledge acquisition (RQ1b),* the PS-DI group showed marginally superior knowledge acquisition overall, but this may also indicate a memory effect. This is particularly noteworthy as the post-test was embedded immediately after the instructional video. In more detail, the PS-DI group acquired more conceptual knowledge, which corresponds to the findings of Kapur (2014) or Loibl and Rummel (2014a). In contrast, the DI-PS group showed higher procedural knowledge, which is in line with findings reported by Loibl et al. (2020). Hence, it can be concluded that findings from other domains are generalisable and applicable to the business domain. Another noteworthy point is that the DI-PS group in our sample reported higher satisfaction with the instructional design. This aligns with the existing literature and can be explained by German schools' predominant teaching method and learning culture (Meyer, 2014). Additionally, the more likely it was that students would receive their preferred instructional design, the more likely it was that students would gain procedural and conceptual knowledge. This suggests considering student preferences when designing instructional designs.

Regarding the effect on cognitive load (RQ2a), both groups perceived the scenario as rather difficult and experienced high intrinsic cognitive load. The PS-DI group, in particular, scored significantly higher on in-scenario cognitive load than the DI-PS group. Van Gog et al. (2011) also showed that in a similar study using worked examples (WE) as a form of more direct instruction, WE-PS showed significantly lower cognitive load. However, despite the perceived high cognitive load experienced by PS-DI, they achieved higher conceptual knowledge. This aligns with Kapur (2014), who found that students in the productive failure condition (similar to PS-DI) achieved higher conceptual knowledge despite high cognitive load. Likourezos and Kalyuga (2017) also showed that learners who were given worked examples experienced less cognitive load; however, this low cognitive load did not lead to higher learning outcomes in total. This is also shown in the present study. Based on these findings in the business field, instructional designers can reduce intrinsic cognitive load by manipulating the level of task difficulty in office simulations (Klepsch & Seufert, 2020). According to CLT, problems with high element interactivity can lead to working memory overload in the PS-DI condition (Ashman et al., 2020). Also, extraneous load was significantly lower for DI-PS. This is in line with Costley et al. (2023) showing the "explicit instruction first" and "supported problem-solving first" groups showed lower ECL, but not with the findings by Leppink et al. (2014), who could not find any differences in ECL between the groups 'worked example-problem (WE-P)' and 'problem-worked example (P-WE)'.

Moreover, the effect of instructional designs on problem solving processes (RQ2b) have not been investigated. Therefore, this study provides new insights: The DI-PS group showed slightly higher total activity frequency than the PS-DI group. This is in line with prior replication studies (Ludwig & Rausch, 2023; Ludwig et al., 2024) that revealed that better problem solvers showed a higher total activity. However, a noticeable difference between the two groups is that the PS-DI group reads more frequently the "Procedure of a supplier selection" document in the reference book. This shows that they sought information since DI was not provided before. This implies that the text instruction given in the problem-based scenario was a comparably helpful source compared to the video-based tutorial.

In research on problem-solving behaviour as predictors of knowledge acquisition and problem-solving performance (RQ3), the use of certain problem-solving strategies is assumed to have a strong influence on problem solving performance. In particular, strategies such as note-taking or spreadsheet use are good predictors for problem solving performance as shown in our prior replication studies (Ludwig & Rausch, 2023; Ludwig et al., 2024).

5.1. Limitations and future research

The present study has certain limitations. Firstly, the method used to measure knowledge acquisition should be considered: Despite the pre-/post-test design for valid measurement of knowledge acquisition, the use of partly identical questions in both the pre- and post-tests may have resulted in a 'test effect,' potentially influencing the results as students might recall the questions or might have remember from the first test (though there was no feedback on test results). Additionally, the low internal consistency of the knowledge test presents another limitation. Secondly, it would be beneficial to include a broader range of problem-based scenarios in future research. This would allow for a more comprehensive assessment of problem-solving performance and a transfer to other problems. Exploring further measures, such as adding another problem-solving scenario or utilising a transfer scale, as suggested by Bego, Chastain, and DeCaro (2023, as cited in DeCaro et al., 2023), could broaden the scope and depth of the evaluation. Moreover, it would be very insightful in future research to compare different types of DI (e.g., video tutorials vs. textbooks vs. worked examples), different types of PS (e.g., simulation-based learning, case studies, role plays), and instructional designs with more than two phases (e.g., DI-PS-DI vs. PS-DI-PS; e.g., Costley et al., 2024). Thirdly, prior experience with active learning environments can influence outcomes. Hence, an onboarding was provided to familiarise our participants with the simulation. Nevertheless, students who have not previously worked with tools like spreadsheets might find the tasks more challenging. Moreover, further research could investigate whether students with higher prior experience (e.g., graduates) or experts in this domain would also prefer DI-PS. This variation in experience can affect the results and should be carefully considered in future analyses. Fourthly, the current findings are based on short-term observations. Future studies should examine long-term learning effects and knowledge transfer. Fifthly, the EES appears to be an effective tool for measuring in-scenario cognitive load. However, further measurements of cognitive load (e.g., physiological measurements) would be valuable for follow-up studies. Lastly, it remains to be investigated how other cultural groups in the study evaluate the two conditions.

To address the weaknesses discussed above, a follow-up study would need to consider the following: (1) The repeated measurement of knowledge in the pre-test and the post-test should be based on a collection of test items that meets the assumptions of Item Response Theory (IRT; Embretson & Reise, 2013). This would allow for using only a few anchor items in both tests but vary most of the test items and thus, mitigating memory effects. However, the development of such tests is very demanding and requires large samples. Furthermore, to control for memory effects, a Solomon four-group design, including additional groups without pre-test, is suggested (Dimitrov & Rumrill, 2003). Here, too, a much larger sample size is required to obtain statistical significance. If possible, performance assessments should also be used to assess the transfer of knowledge in addition to knowledge acquisition. (2) In a follow-up study, it is recommended to consider standard instruments for measuring cognitive load (e.g. Klepsch et al., 2017; Krieglstein et al., 2023). However, regarding the in-scenario measurement based on embedded experience sampling, a one-item measure still seems advisable in order to reduce participant disruption and maintain 'embeddedness'. (3) To gain more insight into the mental processes and current goals of the participants in the different conditions (DI vs. PS) think-aloud protocols of a subsample could provide revealing results as a supplement to the log data-based behavioral data (Fan et al., 2023). (4) Finally, for a follow-up study, the improved study design should be implemented in a different context and possibly using a different simulation environment to better assess the generalizability of the findings reported here.

6. Conclusion

This study provides valuable insights into the effectiveness of different instructional designs in enhancing learning within office simulations. While the traditional approach (DI-PS), which starts with theories and concepts, was preferred due to its perceived lower intrinsic cognitive load during the scenario and extraneous cognitive load, it does not always yield the best learning outcomes. Our findings reveal that learning material in the computer-based office simulation facilitates learning on the domain-specific topic regardless of the instructional design, making it challenging to definitively recommend either DI-PS or PS-DI in the business domain. Moreover, the findings highlight the synergy between the individual preferences and instructional designs. Furthermore, the study found that problem-solving processes did not significantly differ between the groups. Interestingly, while problem-solving processes predicted problem-solving performance, they had little impact on knowledge acquisition. To conclude, further research is needed to establish a balance concerning effectiveness, learner preference, and cognitive load in instructional designs. This will guide the development of more effective teaching strategies. However, further research is needed, particularly to investigate the generalizability of the findings beyond the context of the study.

CRediT authorship contribution statement

Sabrina Ludwig: Writing – original draft, Validation, Resources, Methodology, Formal analysis, Conceptualization, Visualization, Software, Project administration, Investigation, Data curation. **Andreas Rausch:** Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Data curation, Writing – review & editing, Validation, Software, Project administration, Investigation, Formal analysis, Conceptualization. **Michelle Taub:** Supervision, Conceptualization, Writing – review & editing, Project administration.

Disclosure

During the preparation of this work the author(s) used ChatGPT in order to improve the language style. 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 work was supported by the German Federal Ministry of Education and Research (BMBF) within the research initiative "Technology-based Assessment of Skills and Competences in VET (ASCOT+)" (Grant number 21AP008A).

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.

APPENDIX

Appendix A). Domain-specific competence model with scoring criteria and sub-categories

Scoring Criteria	Sub categories	Items	Max. score achievable
1. Planning	1.1 Identifiable action steps	1 point possible for each: - A reply mail was formulated (content irrelevant) - Changes have been made to the spreadsheet (calculations and/or fill in the blanks; content irrelevant). - Changes have been made to the weighting table (scoring and/or ranking; content irrelevant)	3
2. Application	2.1 Information processing: Open documents	- Offer Dishan Yibei Bikes.pdf - Offer Frenos X-Treme.pdf - Decision Table.xslx - Price calculation.xslx - Additional information on international suppliers.pdf - Current exchange rate.pdf	6
	2.2 Application of algorithms	- Offer Supplier 1 (Frenos X-Treme) has a higher ranking than offer Dishan Yibei Bikes (2 p.) - Offer Supplier 1 (Frenos X-Treme) has a significantly higher score (min. 5 p. more) than offer Dishan Yibei Bikes (1 p.) - Offer Dishan Yibei Bikes has a higher ranking and score than offer Frenos X-Treme (0 p.) - No rating was made (0 p.)	3
	2.3 Use of tools	- Formulas were used in the Spreadsheet.xslx file. - Formulas were used in the Weighting Table.xslx file.	2
3. Decision-making	3.1 Correct decision for supplier	- Decision for Frenos X-Treme (1P.) - Decision against Dishan Yibei Bikes (0 p.) - No selection was made (0 p.)	1
	3.2 Correct reasoning (Supplier 1: Frenos X-Treme)	Criteria Offer Frenos X-Treme (1 point each; min. 0 and max. 3): - Purchase price is lower or similar to $ 11.10 (2 credits)	4

(continued on next page)

(*continued*)

Scoring Criteria	Sub categories	Items	Max. score achievable
		- Price, sustainability, social aspects particularly good - After qualitative analysis the higher ranking	
	3.3 Correct reasoning (Supplier 2: Dishan Yibei Bikes)	Criteria offer Dishan Yibei Bikes (1 point each; min. 0 and max. 3): - Purchase price is higher or similar to $ 12.00 (2 credits) - Delivery time, sustainability, social less good - After qualitative analysis the lower ranking	4
	3.4 Reflection regarding Action	- At least one reference to it (1 p.) - No indication (0 p.)	1
	3.5 Reflection regarding Procedure	- At least one reference to it (1 p.) - No indication (0 p.)	1
4. Communication	4.1 Politeness and appropriate communication standards	- Objective, friendly tone/general politeness of the e-mail (2P.) - Appropriate greeting (Hello Mr …, Dear Mr …) (1 p.) - Appropriate closing formula (Yours sincerely, Many greetings, Best regards …) (1 p.) - No mail written (or only very short fragments; few words) (0 p.)	4
	4.2 Absence of formal errors	- No mistakes (2 p.) - A few minor errors (1 p.) - Many major errors (0 p.) - No mail was written (0 p.)	2
	Total		**31**

Appendix B). Pretest and posttest items

Pretest (exemplary items).

What is a supplier selection and what is its benefit for a company? **(Conceptual Knowledge)** a **Offers from different suppliers are compared quantitatively and qualitatively so that a company can make sensible and profitable decisions. (correct)** b Offers from different suppliers are compared quantitatively and qualitatively so that a company becomes the market leader. c Offers from different suppliers are compared in terms of price so that a company can choose the cheapest one. d Offers from different suppliers from different countries are compared with each other so that a company can get the best quality.
A supplier sells a luggage carrier for $ 8.00 gross (VAT 12 %) and gives you a 12 % discount. Calculate with the net price and deduct the discount. What is the price minus the discount? **(Procedural Knowledge)** a **$ 6.29 (correct)** b $ 6.20 c $ 6.08 d $ 6.38

Posttest (exemplary items).

Name three other characteristics of qualitative supplier selection. (Conceptual Knowledge) e.g. Quality of Products, Sustainability, Customer Service (correct answers)
A supplier sells a luggage carrier for $ 12.00 gross (VAT 12 %) and gives you a 12 % discount. Calculate with the net price and deduct the discount. What is the price minus the discount? **(Procedural Knowledge)** a **$ 9.43 (correct)** b $ 9.29 c $ 9.12 d $ 9.57

Appendix C. Instructional preference (Phase II) (exemplary items)

When I want to learn something new or understand a new topic I like to watch instructional videos first. a Strongly disagree. b Rather/Somewhat disagree. c Somewhat agree. d Strongly agree.
I start with a concrete task (“learning by doing”) when I learn something new or want to understand a new topic. a Strongly disagree.

(*continued on next page*)

(*continued*)

b Somewhat disagree.
c Somewhat agree.
d Strongly agree.

Note: Self-developed questions.

Questionnaire: Retrospective satisfaction with the instructional sequencing (Phase IV).

Group DI-PS

I would have understood the topic better if I had worked on the supplier selection scenario/case study first and then watched the instructional video.

a) Yes, the other way around (first the supplier selection scenario/case study, then the instructional video) would have been more helpful.
b) No, the order was helpful as it was.
c) The order did not matter.

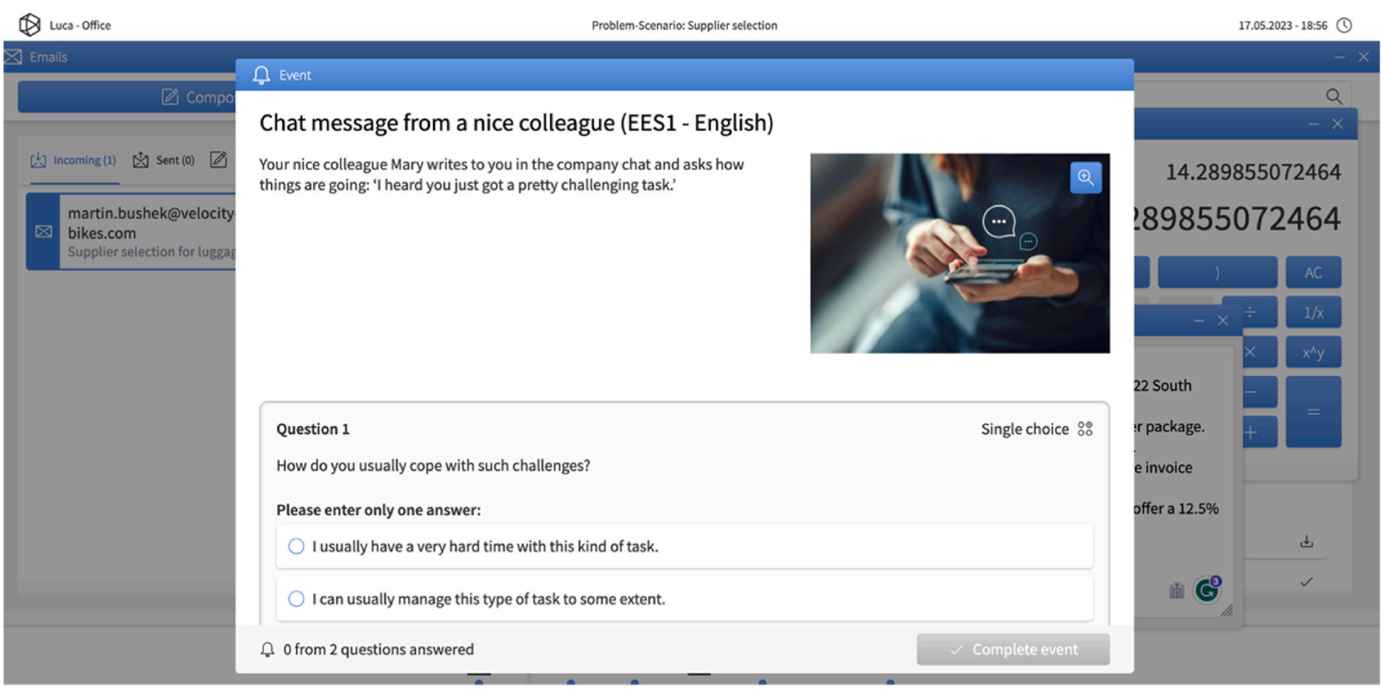
Group (PS-DI)

I would have understood the topic better if I had watched the instructional video first and then worked on the supplier selection scenario/case study.

a) Yes, the other way around (first the supplier selection scenario/case study, then the instructional video would have been more helpful.
b) No, the order was helpful as it was.
c) The order did not matter.

Appendix D. Items to measure Cognitive Load

Appendix D.1. Scenario-embedded experience sampling (EES) as a social interaction to measure learning experience (after 7 min)



Appendix D.2. ***Intrinsic in-scenario cognitive load (IICL) and intrinsic retrospective cognitive load (IRCL)***

We also measured two types of intrinsic cognitive load (ICL): (1) intrinsic in-scenario cognitive load (IICL) and (2) intrinsic retrospective cognitive load (IRCL). The IICL was measured based on a one-item Embedded Experience Sampling (EES) situated within the problem solving phase. The three-item scale for IRCL was measured in the feedback questionnaire and showed Cronbach's alpha of .75 is classified as acceptable. Moreover, extraneous cognitive load was also measured retrospectively with a one-item measure.

Intrinsic in-scenario cognitive load (Embedded Experience Sampling; EES)	How are you coping with the current task? 1 = I am coping with the task very well. 4 = I am finding the task very difficult.
Intrinsic retrospective cognitive load (3-items scale)	I think completing the supplier selection scenario/case study was … 1 = very easy; 4 = very difficult (inverted version) The scenario overwhelmed me in terms of content. 1 = Do not agree at all. 4 = I agree. The scenario underwhelmed me in terms of content. 1 = I agree; 4 = Do not agree at all (inverted version)
Extraneous cognitive load	I think the case study on supplier selection was … 1 = very useful for learning 4 = useless for learning

Appendix E. Regressions analysis

Overall, all predictors accounted for 50 % of the variance in problem-solving competence ($R^2 = .59$; adj. $R^2 = .50$; $df = 32$, $F = 6.50$, $p < .000$). In the group PS-DI, only 'spreadsheet decision table' was the strongest predictor for problem-solving performance. Only 19 % of the predictors could explain the variance in problem-solving competence ($R^2 = .34$; adj. $R^2 = .19$ ($df = 33$, $F = 2.42$, $p < .04$).

Table E.1
Regression analysis on problem-solving performance (Group DI-PS)

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	.07	.03	2.33	.026*
Spreadsheet Decision table	.15	.04	3.50	.001**
Note taking	.10	.05	2.04	.049*
Documents (supplier offers)	.05	.13	.41	.686
Calculator	.01	.07	.18	.861
Task assignment (email)	.68	.76	.89	.377
Reference Book (Procedure of supplier selection)	.17	.28	.62	.539

Note: $R^2 = .59$; adj. $R^2 = .50$ ($df = 32$, $F = 6.50$, $p < .001$); † <.10; * < .05; ** < .01.

Table E.2
Regression analysis on problem-solving performance (Group PS-DI)

Predictors	Estimate	Std. Error	*t*	*p*
Spreadsheet Price calculation	.02	.02	1.04	.31
Spreadsheet Decision table	.14	.05	3.12	.004 **
Note taking	.08	.05	1.58	.12
Documents (supplier offers)	−.05	.13	−.39	.7
Calculator	.05	.07	.7	.49
Task assignment (email)	.41	1.02	.40	.69
Reference Book (Procedure of supplier selection)	.13	.20	.66	.52

Note: $R^2 = .34$; adj. $R^2 = .19$ ($df = 33$, $F = 2.42$, $p < .04$); † <.10; * < .05; ** < .0.

Appendix F. Correlation Tables

Table F.1
Correlation table for the DI-PS group

DI-PS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Spreadsheet Price calculation	1.00	.38	.04	−.23	−.09	.16	.05	.11	**−.46**	−.30	−.18	.05	**.46**	.27	−.09
Spreadsheet Decision table		1.00	.01	.25	.21	.13	.01	−.24	−.25	−.39	**−.46**	−.39	**.61**	−.01	−.30
Note-taking			1.00	−.32	.09	−.02	−.12	.20	−.20	.00	−.30	−.15	.27	.03	.08
Relevant Documents (supplier offers)				1.00	.11	.26	.30	**−.43**	**.46**	.01	−.01	−.04	.19	−.27	.08
Calculator					1.00	−.02	.04	.29	.26	.09	−.18	−.11	.10	.07	.12
Task assignment (email)						1.00	.14	−.19	−.08	−.07	−.20	.00	.24	.06	−.04
Reference Book (Procedure of a supplier selection)							1.00	.03	.17	−.03	−.01	.09	.11	.02	−.12
Time on task								1.00	.04	.01	.02	.22	−.25	−.09	.20
Intrinsic in-scenario cognitive load (IICL)									1.00	**.62**	.05	.27	**−.44**	−.06	.14
Intrinsic retrospective cognitive load (IRCL)										1.00	.19	.24	**−.42**	.24	.06
Extraneous cognitive load (ECL)											1.00	.23	−.30	−.11	.02
Instructional design preference												1.00	−.32	−.20	.34
Problem-solving performance													1.00	.08	−.26
Conceptual knowledge acquisition.														1.00	−.03
Procedural knowledge acquisition.															1.00

X.

Table F.2
Correlation table for the PS-DI group

PS-DI	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Spreadsheet Price calculation	1.00	.37	−.23	.19	.04	−.05	.05	.10	.05	−.22	−.03	−.18	.27	−.30	**.55**
Spreadsheet Decision table		1.00	**−.41**	**.45**	−.05	−.21	−.03	.07	−.16	−.11	.05	−.12	**.55**	−.13	**.41**
Note-taking			1.00	−.26	.00	−.13	−.20	−.01	−.13	−.04	.25	−.18	−.12	−.16	−.21
Relevant Documents (supplier offers)				1.00	.17	.19	.25	.31	.14	.09	.11	−.24	.21	.16	.14
Calculator					1.00	.04	.08	.23	.14	.07	.13	−.21	.00	−.06	.00
Task assignment (email)						1.00	.31	.26	.19	.15	.03	−.11	−.12	.25	−.25
Reference Book (Procedure of a supplier selection)							1.00	.14	.39	**.42**	−.07	.08	.03	.20	.13
Time on task								1.00	.00	.11	.17	−.31	−.23	.10	.22
Intrinsic in-scenario cognitive load (IICL)									1.00	**.71**	−.20	−.04	.03	.04	.23
Intrinsic retrospective cognitive load (IRCL)										1.00	−.10	.03	−.09	.13	.11
Extraneous cognitive load (ECL)											1.00	−.20	.04	−.24	.13
Instructional design preference												1.00	−.06	−.26	−.24
Problem-solving performance													1.00	.00	.15
Conceptual knowledge acquisition.														1.00	.05
Procedural knowledge acquisition.															1.00

Data availability

The data that has been used is confidential.

References

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist, 48*(1), 35–44.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed). Longman.

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review, 32*, 229–247. https://doi.org/10.1007/s10648-019-09500-5

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207–210.

Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In *Handbook of self-regulation of learning and performance* (pp. 254–270). Routledge.

Bego, C. R., Chastain, R. J., & DeCaro, M. S. (2023). Designing novel activities before instruction: Use of contrasting cases and a rich dataset. *British Journal of Educational Psychology, 93*(1), 299–317. https://doi.org/10.7551/mitpress/4561.003.0011

Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. *European Journal of Psychology of Education, 35*(3), 607–624. https://doi.org/10.1007/s10212-019-00445-5

Chen, O., Retnowati, E., & Kalyuga, S. (2019). Effects of worked examples on step performance in solving complex problems. *Educational Psychology, 39*, 188–202. https://doi.org/10.1080/01443410.2018.1515891

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 Padagogische Psychologie*. https://doi.org/10.1024/1010-0652/a000357. online first.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin, 112*, 1155–1159.

Costley, J., Gorbunova, A., Courtney, M., Chen, O., & Lange, C. (2023). Problem-solving support and instructional sequence: Impact on cognitive load and student performance. *European Journal of Psychology of Education*, 1–24. https://doi.org/10.1007/s10212-023-00757-7

Costley, J., Gorbunova, A., Savelyev, A., Shcheglova, I., & Lange, C. (2024). Using worked examples to support problem solving of legal cases in an inductive instructional sequence. *Innovations in Education and Teaching International*, 1–15.

DeCaro, M. S., Isaacs, R. A., Bego, C. R., & Chastain, R. J. (2023). Bringing exploratory learning online: Problem-solving before instruction improves remote undergraduate physics learning. In *Frontiers in education* (Vol. 8)Frontiers Media SA, Article 1215975. https://doi.org/10.3389/feduc.2023.1215975.

Deslauriers, L., McCarty, L. S., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences, 116*(39), 19251–19257.

Dimitrov, D. M., & Rumrill, P. D., Jr. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159–165.

Duncker, K. (1945). On problem-solving (L. S. Lees, trans.). *Psychological Monographs, 58* (5). https://doi.org/10.1037/h0093599

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41* (10), 1040.

Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors, 53*(6), 749–770. https://doi.org/10.1177/0018720811419154

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press.

Fan, Y., Rakovic, M., Van Der Graaf, J., Lim, L., Singh, S., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2023). Towards a fuller picture: Triangulation and integration of the measurement of self-regulated learning based on trace and think aloud data. *Journal of Computer Assisted Learning, 39*(4), 1303–1324. https://doi.org/10.1111/jcal.12801

Fässler, C., Sinha, T., Schmied, C. M., Goldhahn, J., & Kapur, M. (2022). Problem-solving in virtual environment simulations prior to direct instruction for differential diagnosis in medical education: An experimental study. *MedEdPublish, 12*. https://doi.org/10.12688/mep.19348.1

Fischer, F., Bauer, E., Seidel, T., Schmidmaier, R., Radkowitsch, A., Neuhaus, B. J., Hofer, S. I., Sommerhoff, D., Ufer, S., Kuhn, J., Küchemann, S., Sailer, M., Koenen, J., Gartmeier, M., Berberat, P., Frenzel, A., Heitzmann, N., Holzberger, D., Pfeffer, J., … Fischer, M. R. (2022). Representational scaffolding in digital simulations – Learning professional practices in higher education. *Information and Learning Sciences, 123* (11/12), 645–665. https://doi.org/10.1108/ILS-06-2022-0076

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., … Heiberger, R. (2012). Package 'car'. *Vienna: R Foundation for Statistical Computing, 16*(332), 333.

Gorbunova, A., van Merrienboer, J. J., & Costley, J. (2023). Are inductive teaching methods compatible with cognitive load theory? *Educational Psychology Review, 35* (4), 111. https://doi.org/10.1007/s10648-023-09828-z

Greene, J., Moos, D., & Azevedo, R. (2011). Self-regulation of learning with computer-based learning environments. *New directions for teaching and learning.* Publicado en línea en Wiley Online Library.

Hahnel, C., & Stemmann, J. (2023). Entwicklung eines Fragebogens zur ökonomischen Erfassung technikbezogener Einstellungen und Selbstwahrnehmungen. *Diagnostica*. https://doi.org/10.1026/0012-1924/a000323

Hartmann, C., van Gog, T., & Rummel, N. (2021). Preparatory effects of problem solving versus studying examples prior to instruction. *Instructional Science, 49*(1), 1–21. https://doi.org/10.1007/s11251-020-09528-z

Johnson, P. E. (2022). *Package 'rockchalk'*.

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

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28*, 831–852. https://doi.org/10.1007/s10648-015-9352-0

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*, 651–672. https://doi.org/10.1007/s11251-012-9209-6

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*, 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences, 21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery,

problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41* (2), 75–86. https://doi.org/10.1207/s15326985ep4102_1
Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997
Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science, 48*(1), 45–77. https://doi.org/10.1007/s11251-020-09502-9
Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review, 35*(1), 9. https://doi.org/10.1007/s10648-023-09738-0
Kulasegaram, K., Axelrod, D., Ringsted, C., & Brydges, R. (2018). Do one then see one: Sequencing discovery learning and direct instruction for simulation-based technical skills training. *Academic Medicine, 93*(11S), S37–S44. https://doi.org/10.1097/ACM.0000000000002378
Kulgemeyer, C., & Geelan, D. (2024). Towards a constructivist view of instructional explanations as a core practice of science teachers. *Science Education*. https://doi.org/10.1002/sce.21863
Leppink, J., Paas, F., Van Gog, T., van Der Vleuten, C. P., & van Merriënboer, 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
Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science, 45*, 195–219. https://doi.org/10.1007/s11251-016-9399-4
Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*, 693–715. https://doi.org/10.1007/s10648-016-9379-x
Loibl, K., & Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*, 305–326. https://doi.org/10.1007/s11251-013-9282-5
Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004
Loibl, K., Tillema, M., Rummel, N., & van Gog, T. (2020). The effect of contrasting cases during problem solving prior to and after instruction. *Instructional Science, 48*(2), 115–136. https://doi.org/10.1007/s11251-020-09504-7
Ludwig, S., & Rausch, A. (2023). The relationship between problem-solving behaviour and performance–Analysing tool use and information retrieval in a computer-based office simulation. *Journal of Computer Assisted Learning, 39*(2), 617–643.
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.
McCormick, R. (1997). Conceptual and procedural knowledge. *International Journal of Technology and Design Education, 7*, 141–159. https://doi.org/10.1023/A:1008819912213
Meyer, E. (2014). *The culture map: Breaking through the invisible boundaries of global business*. Public Affairs.
Nunnally, J. C., & Bernstein, I. C. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
Plass, J. L., & Schwartz, R. N. (2014). Multimedia learning with simulations and microworlds. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 729–761). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.036.
Prince, M., & Felder, R. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education, 95*, 123–137. https://doi.org/10.1002/j.2168-9830.2006.tb00884.x
Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346–362. https://doi.org/10.1037/0022-0663.93.2.346
Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulation-based learning experiences: An integrative review. *Nurse Education Today, 99*, Article 104815. https://doi.org/10.1016/j.nedt.2021.104815
Saba, J., Kapur, M., & Roll, I. (2023). The development of multivariable causality strategy: Instruction or simulation first?. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Eds.), *Artificial intelligence in education. AIED 2023. Lecture notes in computer science( )* (Vol 13916) Cham: Springer. https://doi.org/10.1007/978-3-031-36272-9_4.
Schalk, L., Schumacher, R., Barth, A., & Stern, E. (2018). When problem-solving followed by instruction is superior to the traditional tell-and-practice sequence. *Journal of Educational Psychology, 110*(4), 596. https://doi.org/10.1037/edu0000234
Sinha, T., & Kapur, M. (2021). When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. *Review of Educational Research, 91*(5), 761–798. https://doi.org/10.3102/00346543211019105.
Smith, W., Hermida, J., & Güss, C. D. (2022). "Oh no, the forest is burning!" cultural differences in the complex problem-solving process only under high uncertainty. *Frontiers in Psychology, 13*, Article 965623. https://doi.org/10.3389/fpsyg.2022.965623
Sternberg, R. J. (1997). Intelligence and lifelong learning: What's new and how can we use it? *American Psychologist, 52*(10), 1134. https://doi.org/10.1037/0003-066X.52.10.1134
Strohschneider, S., & Güss, D. (1998). Planning and problem solving: Differences between Brazilian and German students. *Journal of Cross-Cultural Psychology, 29*(6), 695–716. https://doi.org/10.1177/0022022198296002
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5
Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5
Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/A:1022193728205
Taub, M., & Azevedo, R. (2019). How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? *International Journal of Artificial Intelligence in Education, 29*(1), 1–28. https://doi.org/10.1007/s40593-018-0165-4.
Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments*? Computers in Human Behavior, 39*, 356–367. https://doi.org/10.1016/j.chb.2014.07.01.
Thibaut, L., Ceuppens, S., De Loof, H., De Meester, J., Goovaerts, L., Struyf, A., Boeve-de Pauw, J., Dehaene, W., Deprez, J., De Cock, M., Hellinckx, L., Knipprath, H., Langie, G., Struyven, K., Van de Velde, D., Van Petegem, P., & Depaepe, F. (2018). Integrated STEM education: A systematic review of instructional practices in secondary education. *European Journal of STEM Education, 3*(1), 2. https://doi.org/10.20897/ejsteme/85525
Tobias, S., & Duffy, T. M. (2009). The success or failure of constructivist instruction: An introduction. In S. Tobias, & T. M. Duffy (Eds.), *Constructivist instruction: Success or failure?* (pp. 15–22). Routledge.
Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212–218. https://doi.org/10.1016/j.cedpsych.2010.10.004
Van Merriënboer, J. J., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5–13. https://doi.org/10.1207/S15326985EP3801_2
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
Winne, P. H., & Azevedo, R. (2022). Metacognition and self-regulated learning. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 93–113). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108888295.007.
Winne, P., & Hadwin, A. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Erlbaum.
Zhang, J., Andres, J. M. A. L., Hutt, S., Baker, R. S., Ocumpaugh, J., Mills, C., … Young, T. (2022). Detecting SMART model cognitive operations in mathematical problem-solving process. *International Educational Data Mining Society*. https://doi.org/10.5281/zenodo.6853161
Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13–39). Academic press. https://doi.org/10.1016/B978-012109890-2/50031-7.

Sabrina Ludwig is a research assistant of Economics and Business Education, Mannheim Business School, University of Mannheim. Her research focuses on students' problem-solving behavior and learning in simulation-based environments.

Andreas Rausch is a full professor of Economics and Business Education, Mannheim Business School, University of Mannheim. His research focuses on simulation-based learning, workplace learning, problem solving, and competence assessment.

Michelle Taub is an assistant professor in the Department of Learning Sciences and Educational Research, University of Central Florida. Her research focuses on multimodal data, metacognition and self-regulation of learning.