Received: 2 May 2022 | Revised: 28 November 2022 | Accepted: 4 December 2022

DOI: 10.1111/jcal.12770

ARTICLE

Journal of Computer Assisted Learning WILEY

The relationship between problem-solving behaviour and performance – Analysing tool use and information retrieval in a computer-based office simulation

Sabrina Ludwig | Andreas Rausch

Chair of Economic and Business Education – Workplace Learning, University of Mannheim, Mannheim, Germany

Correspondence
Sabrina Ludwig, University of Mannheim, Chair of Economic and Business Education – Workplace Learning, University of Mannheim, L 4, 1, 68161 Mannheim, Germany.
Email: sabrina.ludwig@uni-mannheim.de

Funding information
Federal Ministry of Education and Research, Grant/Award Number: 01DB1119-23

Abstract

Background: Problem-solving competences have evolved into key skills for professionals. Computer-based business simulations enable the analysis of problem-solving processes beyond end results. An important aspect of successful problem-solving is to systematically use built-in tools and process the provided information efficiently.

Objectives: This investigation explores the relationship between students' tool use, information retrieval, and problem-solving success in a computerized problem-solving office simulation.

Methods: Around 30,000 recorded behavioural log data points of 432 German vocational students were analysed.

Results and Conclusion: Distinct user groups are identified and cognitive problem-solving competences are assessed to draw a link between behaviour and performance. An explorative cluster analysis based on student behaviour revealed four clusters. Significant results support the use of two cognitive tools that lead to success. One successful behaviour is using a notepad, a domain-general and voluntary tool. Another successful problem-solving behaviour is the use of a domain-specific and solution-relevant spreadsheet program. Note-taking organizes information and mental processes while the spreadsheet leads to efficient computing. In line with other studies, students with higher problem-solving competences tend to access tools and documents providing information more frequently.

Takeaways: Domain-general tool use differs from domain-specific tool use over time. There are two different successful behaviour patterns in complex problem solving. Instructional and simulation designers should provide specific tools to support students as well as tackle problems.

KEYWORDS

21st-Century Competences, Adult Learning, Cluster Analysis, Cognitive Tools, Educational Data Mining, Log Data, Problem-Solving Competences

1 | INTRODUCTION

Over the past two decades, there has been a paradigm shift towards complex open-ended learning environments (OELE) and computer-based assessments emphasizing the relevance of problem-solving skills for both academic and professional work environments (Csapó & Funke, 2017; Krkovic et al., 2018; World Economic Forum, 2020). Given the relevance of problem-solving competences and their hard-to-measure nature, the question of how these skills can be accurately assessed has arisen. In 2000, and for the first time, students' problem-solving competences were assessed using the paper-based 'Programme for International Student Assessment' (PISA) (Wirth & Klieme, 2003). A decade later, for the sake of authenticity, the PISA study 2013 was digitized and distributed internationally. A similar assessment, the 'Programme for the International Assessment of Adult Competencies (PIAAC)', designed for adults, aimed to measure the extent to which adults could cope with problems in technology-rich environments (PIAAC, 2009). Since then, computer-based assessments have been further developed and diversified. Advanced learning and assessment technologies, such as Intelligent Tutoring Systems enable new opportunities to deploy authentic scenarios (e.g., Bouchet et al., 2012). This often allows responsible and role-based participation in real-life decision making within a realistic environment (e.g., 'microworlds'). Accomplishing authentic scenarios requires but also promotes problem-solving competences such as the application of problem-solving strategies. These explorative problem-solving processes necessitate a high degree of self-regulation (Bosch et al., 2021; Greene & Azevedo, 2010). Learners are provided with autonomy, flexibility, and choices while they plan, as well as monitor and actively evaluate their own solving path (Kostons et al., 2012; Winne, 2001). Built-in cognitive tools might assist problem-solvers to reinforce their problem-solving processes and relieve their cognitive load by providing information as well as organizing and processing embedded content (Jonassen, 1996; Liu & Bera, 2005). Depending on the underlying problem the learner faces, cognitive tools might be either domain-specific or domain-general and either mandatory or voluntary.

Typically, to draw inferences about his or her abilities, a learner's proposed solution to the problem is assessed. However, to understand and assess problem solving, it is a promising approach to also look at the problem-solving processes instead of just looking at the product (He, Borgonovi et al., 2019). Problem-solving processes can be investigated by analysing log data. Using computerized systems over traditional assessment methods offers the advantage of being able to log every interaction between the problem solver and the available cognitive tools, the provided information and real-world problem situations (Frensch & Funke, 1995; Li & Tsai, 2017; Mislevy et al., 2017; Molkenthin et al., 2008). As digitally recorded and observable behavioural log data within computerized simulation tasks provides evidence for learners' traits and performance (Mislevy et al., 2017), problem-solving processes that formerly remained a black box become transparent (di Mitri et al., 2018; Dörner & Funke, 2017; Mayr et al., 2011). However, opening the black box requires appropriate theory-driven or data-driven approaches. Specifically, explorative cluster analyses enable the identification of problem-solving patterns based on process indicators (e.g., frequency or duration of performed activities), as well as the deduction of valid interpretations beyond end results such as test scores (Eichmann et al., 2020; Li & Tsai, 2017; Nisbet et al., 2009). Clustering has received wide attention. However, existing studies differ in domain, content, degree of complexity, tools, and analysis methods. Therefore, findings of existing studies might not be generalizable and transferable. So, evidence is lacking for the relationships between problem-solving behaviour and performance in a digital office simulation designed for Vocational Education and Training (VET). To address this research desideratum, we conduct a cluster analysis based on students' problem-solving behaviour because (1) from a problem-solving perspective, it reveals ways in which learners approach domain-specific problems. (2) From a technical perspective, it supports the design of learning environments and digital assessments, and (3) from an instructional perspective, identifying different problem-solving approaches enables targeted support. Against this background, this study aims to identify different user patterns based on cognitive tool use and information retrieval in a problem-based office simulation for industrial business management assistant.

The following underlying research questions will be investigated:

> Research Question 1: Can student clusters be identified based on the log data of tool use and information retrieval?
>
> Research Question 2: How do these student clusters differ in their tool use and information retrieval over time (process perspective) and between scenarios (context perspective)?
>
> Research Question 3: How do these student clusters differ regarding their problem-solving performance (product perspective)?

2 | LOG DATA ANALYSIS OF PROBLEM-SOLVING BEHAVIOUR IN COMPUTER-BASED ASSESSMENTS

2.1 | Definition of complex problem-solving behaviour

A 'problem' exists when a living being in general, or a specific person, wants to shift the confronted initial problem state to a desired and still unknown state due to lack of knowledge (Duncker, 1935). The unknown goal state can be reached by applying strategies in the problem's space (Newell & Simon, 1972). Here, Jonassen (2000a) distinguishes between two kinds of problems. The first is well-structured problems, such as simple puzzles, characterized by the manipulation of just a few operators and a clearly defined solution. The second kind is ill-structured problems, such as dilemmas, characterized by a high number of interrelated variables and having many possible solution paths, resulting in high complexity (Funke, 1991). According to Funke (2012) and Dörner and Funke et al. (2017), a complex problem situation is characterized by (1) *complexity* due to the underlying number of variables, (2) *connectivity* regarding their interplay and dependency on each other; (3) *dynamics*, such as (limited) time, and as the problem situation is under ongoing development; (4) *intransparency* due to multiple variables and functions, and (5) *polytely* as multiple (*poly*) goals (*telos*) are pursued, which may be in conflict with each other. For example, project managers can be confronted with complex problems since several stakeholders with different interests are involved, there are changes in available resources, and the outcome is uncertain and unpredictable in advance.

Working through these complex problems requires a comprehensive orchestration of cognitive, metacognitive and non-cognitive steps. For the sake of simplification and to reduce the problem's complexity, problem solvers strive to divide the overall problem goal into sub-goals or a "series of core steps" (Newell, 1979; Huang et al., 2017; PIAAC, 2009, p. 15). This is done by (1) constructing an 'internal mental representation' of the problem (e.g., thinking or procedural knowledge), followed by (2) a search for solutions and (3) implementation of a solution as well as externalization of physical representation through active manipulation of the problem's variables (Jonassen, 1997, 2000a, p. 65; Mayer & Wittrock, 2006). The sequence of these phases can be questioned as it depends on the complexity and type of problem (PIAAC, 2009; Jonassen, 2000a; Rausch et al., 2017).

Depending on the individual situation, problem solvers' approaches to problems can be classified as heuristics or algorithms. Heuristics are domain-generic strategies whose application based on experience leads to quick conclusions in any domain. Typical heuristics are rules-of-thumb or trial-and-error (Winne & Nesbit, 2010). However, as knowledge might be incomplete in complex problems and unknown environments, heuristics might be more error-prone and lead to systematic errors or bias. To address this problem, another strategy that is more reliable, although domain-related, is the use of algorithms (Newell & Simon, 1972). The latter needs more time and effort as it poses an iterative procedure; however, the likelihood of finding a solution is higher. Therefore, Van Merrienboer (2013) categorizes algorithms as strong methods and heuristics as weak methods. On a continuum, Van Merriënboer (2013, p. 153f) differentiates between four methods: (1) 'weak problem-solving methods' (such as note-taking); (2) 'knowledge-based problem-solving methods' (such as information reasoning); (3) 'strong problem-solving methods' (such as using domain-specific algorithms in a spreadsheet); and (4) a combination of 'strong' and 'knowledge-based methods'. Weak problem-solving methods are characterized by their domain-generic applicability to problem situations. Therefore, for example, the use of a notepad (such as retrieving or inserting notes) can be aligned to weak methods since it does not necessarily solve a problem. It rather serves as a support or reminder and can be applied in any domain. In contrast, strong and knowledge-based methods can be classified as domain-specific which includes, for example, correct spreadsheet use that can support solving a domain-specific problem.

2.2 | Computer-based simulations, cognitive tools and information sources

Computer-based simulations are valid environments to assess problem-solving behaviour depending on the extent of the action scope. According to Behrens et al. (2012), the action scope and complexity of simulations depend on the following four spaces: (1) problem space, (2) tool space, (3) solution space, and (4) response space. Since the scope of these spaces can vary from narrow to broad, the degree of task complexity can be adapted. One example for a narrow problem space is the most widely known computer-based and large-scale-assessment PISA. In PISA 2012, students worldwide were asked to solve ill-structured and non-transparent everyday problem-based tasks (OECD, 2014). The so-called "MicroDYN" tasks, based on a linear equation system, comprise up to three input and output variables. The task was divided into two parts. First, students acquired knowledge about the relations between variables while exploring the dynamically changing problem. In doing so, they manipulated the independent variables by moving the slider of the input variables. Secondly, they were provided with the correct model and were asked to apply their acquired knowledge to achieve certain goals by manipulating the input variables adequately in only a few clicks. The "vary-one-thing-at-a-time strategy" (VOTAT) proved the best strategy since students could see the isolated effect of the manipulation of only one input variable on the output variable while all other variables were held constant (Greiff et al., 2016).

The other psychometric approach "MicroFIN" functions as a finite state machine. Participants work with the system (e.g., an MP3-player) by shifting from one state to another. During this transition, certain actions are triggered and the subsequent state results from the current state and an external event (e.g., pressing the 'play' button; Greiff et al., 2013). As both tasks ("MicroDYN" and "MicroFIN") demonstrate low complexity, problem-solving behaviour could be assessed in a more accurate way ("reliability-validity dilemma"; Seifried et al., 2020). An assessment with a similar purpose and narrow space was designed for adults aged 16–64 years, proposed by the PIAAC (PIAAC, 2009). Hereby, the study's goal was to measure adults' reading, mathematical and problem-solving competences in technology-rich environments (PS-TRE). Problem-solving competences, in particular, embrace evaluating several information sources, which problem solvers base their decisions on. Decision-making is a key competence for successful information processing required in the modern 'information society' (Levy, 2010; Nesbit & Winne, 2008). To solve the information problem, computer-based tools (such as spreadsheets or graphical tools) that are not available without technology-based assessments amplify assessments with new perspectives (PIAAC, 2009). Typically, clear variables and less complex problem situations encourage problem solvers to act in minimal complex systems that open the door to reliable assessments. However, at the same time, minimal complex systems do not act as proper frames to validly assess complex problem solving (Funke et al., 2017, 2018). As it is often discussed that real-world problems are characterized by a higher degree of complexity and dynamics (see 'microworlds'; Brehmer & Dörner, 1993), more sophisticated instructional design software is key to designing more comprehensive problems with dynamic and real-world characteristics.

Complex computerized systems with a broader problem space, such as Intelligent Tutoring Systems (Bouchet et al., 2012), help simulate real-world scenarios with a safe scope of action (see, for example, medical interventions or flight simulators; Elger et al., 2018). Furthermore, as a 'student-centered approach', complex computerized systems allow experiences to be conveyed with a sense of autonomy. Besides being a goal-directed learning environment, simulations in particular can be a promising platform for measuring high-ordered competences such as critical thinking, decision making or problem solving ('simulation-based assessment', Levy, 2013; Caruso, 2019).

Depending on the design of these real-life scenarios and action scope of the simulations ('free flow') (Behrens et al., 2012), solving scenarios in computerized environments requires a high degree of self-regulative behaviour (Nesbit et al., 2007). Deciding which information is relevant and which is irrelevant in the given problem space and deciding on the procedure and duration they want to engage with given materials (Azevedo, 2005; Behrens et al., 2012; Kostons et al., 2012) is a challenging task. Similar challenges are discussed in research on 'Information retrieval' and 'Multiple Document Comprehension (MDC)' (Hahnel et al., 2021; Mahlow et al., 2020).

A variety of different tools supports learners in coping with complexity. To address this issue, Jonassen (2000b) calls for the provision of a full range of cognitive tools that support necessary skills and guide the problem solver through the task. In this context, Lajoie (1993, p. 261) categorizes four cognitive tools that trigger the acquisition and application of knowledge to achieve sub-goals, and, thus, enhance the externalized representation of inner cognitive processes (Liu & Bera, 2005). The author distinguishes between tools that (1) support cognitive processes (such as notepads or calculators), (2) share cognitive load (such as information sources, e.g., databases or files that contain multiple documents), (3) enable engagement in cognitive activities that are usually out of their reach (e.g., domain-specific tools such as spreadsheets or a voltmeter that are generally not available within the classroom), and (4) enhance the generation of hypotheses and their testing (e.g., requesting feedback on partial solutions). The design and function of embedded cognitive tools are often derived from the real world. Therefore, a distinction between domain-general and -specific tools should be evident in the following. For example, a spreadsheet program in the case of a business problem would be a specific workplace tool for employees to support them in their algorithmic thinking (Marriott, 2004). Geiger et al. (2015) confirm this positive impact of spreadsheets by arguing that students can think flexibly and critically due to individual data input. In contrast, a voltmeter, for example, would be an adequate tool when solving a problem from the domain of electricity. Moreover, some tools are necessary to solve tasks and, thus, less easily replaced. Accordingly, Trafton and Trickett (2001) highlight the difference between using mandatory and voluntary tools. Whereas the former tools represent an indispensable means to solve problematic tasks, the latter function as optional devices and are instead supportive during the problem-solving process. A notepad is domain-general and voluntary, while enterprise resource planning (ERP) software is domain-specific and mandatory if the present problem can only be solved with the ERP software.

2.3 | Log file analysis of tool use and information retrieval in computer-based simulations

Patterns of tool use and information retrieval can be identified through log file analysis. Every interaction between the computerized system and the user, such as mouse clicks, keyboard strokes or page visits, is collected as time-stamped log data and stored in log files. These action logs – also known as system logs, process data, or trace data – provide valuable information about users' activity patterns. Thus, they enable insights into student behaviour and open up new opportunities in educational research that previously remained sealed. Linking a relationship between behaviour patterns, (meta-)cognitive approaches and performance adds extended information and a better understanding of underlying mechanisms (Cheng & Chau, 2016; Li & Tsai, 2017; Tóth et al., 2017).

Several studies (e.g., He, Borgonovi et al., 2019; Liu & Bera, 2005; Tóth et al., 2014) showed that indicators based on process data, like the frequency of activities ('engagement level') or the time that students spend on information material and respective tools, proved themselves to be widely used and promising metrics to measure problem-solving behaviour. An example is provided by Goldhammer et al. (2014), who examine problem-solving success in a large-scale computer-based assessment. They revealed a positive effect between task duration and correct responses as controlled processing takes time. Furthermore, based on time-stamped process data, different problem-solving patterns among successful and less successful problem solvers can be identified (He, Liao et al., 2019). Questions regarding whether individuals might have remained consistent in their systematic approach over time and whether it leads to a better performance can be answered. Moreover, the design of the assessment can be customized to the target group and, thus, optimized (He, Liao et al., 2019) so that different stakeholders, such as teachers or students can profit (Shute & Ventura, 2013). In sum, log data provide meaningful insights. However, to shed light on digitally recorded problem-solving behaviour, appropriate methods such as clustering is required (Romero & Ventura, 2010). Clustering encompasses the opportunity to find hidden action patterns in unstructured big data (Romero & Ventura, 2010; Shin & Shim, 2020). By means of these data-driven approaches, distinct successful action patterns can be identified and linked to performance. An overview of relevant studies is given in the following.

Liu and Bera (2005) investigated problem-solving patterns and tool use within the problem-based learning (PBL) software *Alien Rescue*. Sixth graders learn and gain domain-specific knowledge about the solar system and approaches that scientists typically apply (Liu et al., 2002). To support learners in solving the problem of this scientific topic, the environment provides different authentic cognitive tools such as several databases with information about aliens or the solar system, and a notebook for taking notes or solution forms to test hypotheses (Liu & Bera, 2005). Based on user behaviour and a cluster analysis on the above-mentioned cognitive tools in the problem-based hypermedia environment, Liu and Bera (2005) examine tool use over time. They found that "tools supporting cognitive processing and sharing cognitive load played a more central role early in the problem-solving process, whereas tools supporting cognitive activities that would be out of students' reach otherwise, and hypothesis generation and testing were used more in the later stages of problem-solving" (Liu & Bera, 2005, p. 5). Furthermore, they found that high-performing problem solvers exhaust tools more productively than less successful students. In sum, the tool use patterns provided evidence that the tools support the students during their problem-solving process as they are able to express their mental problem approach externally (Jonassen, 2003). In another research study on *Alien Rescue*, three clusters (high-, average- and low-access cluster) based on the frequency of tool use were identified (Bera & Liu, 2006). Interestingly, the analysis revealed that the tool access frequency rate relates negatively to the clusters' test scores. However, students with the lowest frequency spent the most time in the first half. Other researchers, like Dumdumaya et al. (2018, p. 283), investigated persistence during the problem-solving process based on clustering. The analysis of inferred process indicators like time on task ('Time on Tutored Problems'), 'Time on Resources' and frequency of hint requests identified two clusters: 'more persistent students' and 'less persistent students'. Other researchers also investigate students' persistence (DiCerbo, 2014). The indicator 'total time spent on a task-relevant event' or 'time spent for a solved problem' might be evidence for persistence, which turned out to be crucial for successful complex problem-solving (DiCerbo, 2014, p. 18).

In another study, in *Betty's Brain*, a learning-by-teaching environment developed by Biswas et al. (2005), students were asked to construct the brain of the fictive agent *Betty*, which compromises cause-and-effect relationships in the field of science such as the Greenhouse effect. They were provided with information resources and a notepad, and could evaluate their construction by letting Betty take quizzes. The authors aimed to investigate cognitive and metacognitive processes in self-regulative learning (SRL). By means of combining sequence mining techniques they were able to distinguish the frequency patterns of high and low achievers (Kinnebrew et al., 2013). The high-achievers typically tended to consider quizzes with relevant readings or queries, whereas the low-performers showed a tendency towards quizzes with unrelated reading (Kinnebrew et al., 2013). In their study, Sabourin et al. (2012) reveal that students with a higher SRL score exploit resources (such as reading books and posters) in the game-based learning environment *Crystal Island* more and also take notes in a more intensive manner than students with a low SRL score while dealing with problems in the domain of 'microbiology'.

Within the project 'Domain-specific Problem-Solving Competence of Industrial Clerks', an office simulation was developed with the aim of assessing Vocational Education and Training (VET) students' problem-solving competences. The design of the office simulation, including authentic cognitive tools and real-world scenarios comprising a stack of documents, was undertaken and derived from a comprehensive domain analysis in the business field of 'controlling' (Eigenmann et al., 2015; Rausch & Kögler, 2016), resulting in the integration of three complex scenarios into a model company (a bicycle manufacturer). Based on descriptive analyses regarding the tool use and information retrieval, they revealed that the group of more successful problem solvers show a higher number of activities and tool use dependent on the present problem scenario and performance groups.

Until now, many problem-based environments have been developed and corresponding explorative data-driven approaches applied to investigate students' problem-based behaviour. However, one cannot generalize the findings of other domains on problem-solving behaviour since problem solving is highly context-specific. Furthermore, simulation environments are difficult to compare because the integrated tools are also domain-specific. Indeed, simulations are widely spread in the field of VET. Nevertheless, to the authors' knowledge, no study involving a data-driven approach exists that focuses on profiling students in VET based on their information retrieval and tool use while completing complex problem-based scenarios within the domain of controlling. The scarcity of research on behaviour and timing regarding information retrieval in digital environments (Li & Tsai, 2017; Molkenthin et al., 2008; Reich, 2015) and a lack of scholarly consensus regarding the effectiveness of tools (Clarebout & Elen, 2006; Liu & Bera, 2005) pose a need for further explorative investigations. Therefore, in our article we rely on a data-driven approach to investigate the research questions posed above: (1) Can student clusters be identified based on tool use and information retrieval? (2) How is tool use and information retrieval behaviour distributed over time? (3) How is this tool use behaviour and information retrieval related to their problem-solving performance?

3 | METHOD

Our log data analysis is based on a dataset from the research project "Domain-Specific Problem-Solving Competence of Industrial Clerks" which was funded by the German Federal Ministry of Education and Research (BMBF). The underlying office simulation was developed to assess the domain-specific problem-solving competences of students in VET (Rausch et al., 2016). Earlier process analyses relied on descriptive approaches such as heat maps (Rausch et al., 2017). The present paper uses cluster analyses to explore the data further.

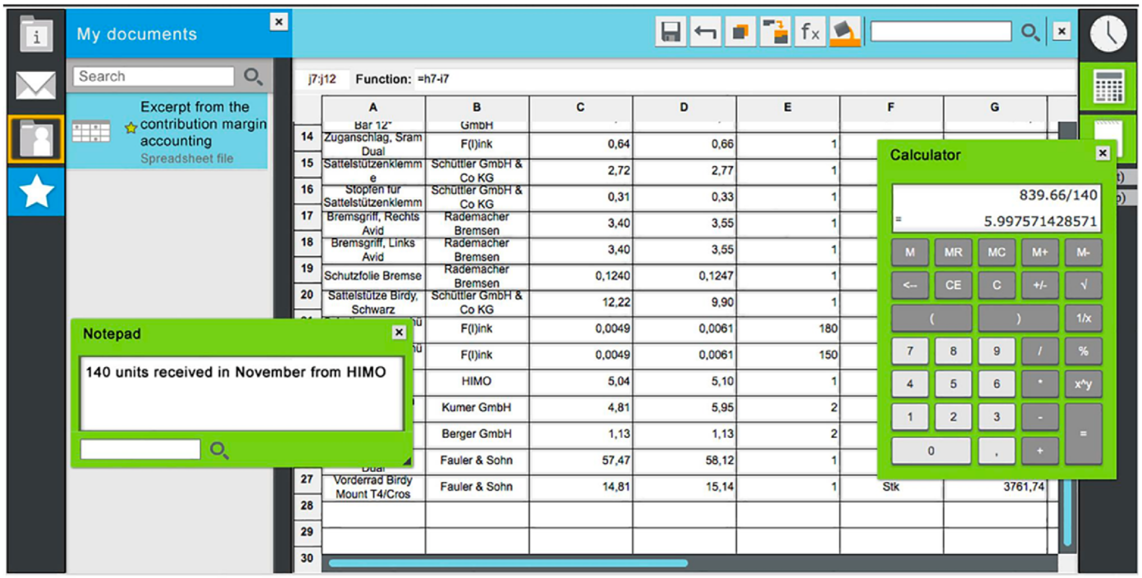


FIGURE 1 Screenshot of the office simulation (translated from German, Rausch et al., 2016, p. 8)

3.1 | Participants

In the present study, we analyzed log data points of 432 German students in VET who participated in this study. The German VET system is characterized by a combination of vocational school instruction and workplace learning within a training company. The participants were in their second or third year of a 3-year apprenticeship program and showed a typical age distribution ($M = 21.3$ years; $SD = 2.69$; $min = 17$; $max = 44$). They were enrolled in three similar commercial apprenticeship programs (industrial business management assistants, IT-system management assistants, and merchants in wholesale and foreign trade; see Rausch et al., 2016).

3.2 | Office simulation and procedure

Data were collected in computer-equipped classrooms in vocational schools. The researchers introduced the project, provided information about data protection, and emphasized the voluntary nature of participation. All participants provided written, informed consent. After completing a questionnaire on biographical data, the participants were introduced to the fictive model company and the features of the office simulation in a 15-minute tutorial (Rausch et al., 2016). Afterwards, they worked continuously, with only one short break, through three scenarios from the domain of financial control. Participants had 30 min to complete each scenario. The office simulation provided typical tools and features of an office workplace such as an e-mail client, a file system, spreadsheet software, a calculator, and a notepad (see Figure 1). Altogether, the simulation is characterized by its broad tools and solution space according to Behrens et al. (2012), which allows students to work autonomously within the time constraint without being given any further assistance or hints. Participants should have some prior knowledge on the content and tools. This was ensured by comprehensive domain and curriculum analyses.

Each work scenario included algorithmic, information and decision-making problems. Solving these work scenarios required the application of domain-specific quantitative algorithms (e.g., calculating acquisition prices, carrying out a cost variance analysis) as well as weighing up the relevance and credibility of qualitative information (e.g., relevance of delivery times, effects of outsourcing). Therefore, some of the provided tools were mandatory. In particular, carrying out calculations in the spreadsheet application or composing an email reply were indispensable to successfully process the scenario. Other tools like the calculator and the notepad were voluntary. The information provided in the form of documents was either relevant or irrelevant. The available tools and information sources can be classified according to the definitions of cognitive tools introduced in the theoretical section (Table 1).

In scenario 1, the participants conducted a cost variance analysis to compute target costs as well as the variances between target and actual costs within a given spreadsheet table (application of algorithms). Then they had to interpret the quantitative results on the basis of a variety of business documents such as invoices, business letters or file notes (information resources). Based on that information, the participant needed to communicate possible consequences and recommended actions (decision-making) to his or her supervisor as an elaborated email response (Rausch et al., 2016). In scenario 2, students selected an appropriate supplier based on a benefit analysis (application of algorithms) considering quantitative and qualitative characteristics (information resources) and proposed their solution in an email (decision-making). In scenario 3, the student decided between in-house production or purchasing from an external supplier (make-or-buy decision). These scenarios differed in their sequences of required activities and the extent of relevant information sources.

TABLE 1 Categorisation of information sources and tools in the office simulation at hand

Rausch et al. (2016)	Trafton & Tricket (2001)	Van Merriënboer (2013)	Lajoie (1993)
Information sources			
Client for relevant Emails	mandatory use	strong method	out-of-reach-tool
Client for irrelevant Emails	voluntary use	strong method	out-of-reach-tool
Relevant documents (PDF-viewer)	mandatory use	knowledge-based method	sharing tool
Irrelevant documents (PDF-viewer)	voluntary use	knowledge-based method	sharing tool
Tools			
Relevant spreadsheet program	mandatory use	strong method	out-of-reach-tool
Irrelevant spreadsheet program	voluntary use	strong method	out-of-reach-tool
Notepad	voluntary use	weak method	supportive tool
Calculator	voluntary use	weak method	supportive tool

[Correction added on 19 January 2023, after first online publication: The heading for column 1 in Table 1 has been updated in this version.]

TABLE 2 Partial credits for each subdimension of problem-solving performance

		Maximum achievable score		
Subdimensions of problem-solving performance	**Description of the underlying performance**	**Scenario 1**	**Scenario 2**	**Scenario 3**
A1. Identification of needs for action and sources of information	The participant identifies relevant sources of information.	4	4	4
A2. Process of information	The participant applies domain-specific processing algorithms (e.g., correct calculations) and domain-specific tools (e.g., functions of the spreadsheet application).	6	4	5
A3. Making well-founded decisions	The participant comes to a plausible decision by weighing up and integrating the results of the quantitative analyses and qualitative information.	6	4	5
A4. Communication of decisions in an adequate way	The participant communicates the solution appropriately by using correct technical language and complying with typical communication standards	3	3	3

Therefore, we expected scenario-specific patterns of tool use and information retrieval.

3.3 | Performance measure ('product perspective)

The participants' email replies (m = 150 words) and further calculations within the spreadsheet served as the basis for performance assessment (please refer to Table A.1 in Appendix A.1 for exemplary coding schemes). The problem-solving performance for each scenario was assessed on four dimensions according to the theoretical framework by Rausch et al. (2016) (Table 2). The procedure comprises the following three steps (Seifried et al., 2020): First, a fine-grained analysis of each proposed solution was conducted based on a comprehensive coding scheme with scoring rubrics. This analysis resulted in a huge variety of solution patterns. Second, in an iterative process, experts assigned partial credit scores for each of these response patterns. Third, a multidimensional model was calibrated based on these partial credits to receive generalized competence scores. However, in order to investigate the relation of tool use patterns and problem-solving success, we used the scenario-specific partial credits (see Appendix A.2 for the corresponding Figure A.2 outlining the three-step procedure, and Appendix A.3 with Table A.3 for the descriptive statistics on problem-solving performance).

3.4 | Log data analysis ('process perspective')

To be included in the analysis, log data and performance data for each student and for all scenarios have to be complete. The log file containing all time-stamped student interactions executed in the office simulation were processed and assigned to a dataset. In a first step, all mouse clicks and keystrokes that were made in direct succession in a particular tool were combined into one activity (e.g., several keystrokes in a notebook resulted in one notebook activity, successive entries in a spreadsheet resulted in one spreadsheet activity, etc.). The end of each activity led into a new activity (e.g., finishing a notebook entry and beginning with a calculation in the spreadsheet). This resulted in a dataset of approximately 30,000 activities by 432 participants in three

TABLE 3 Variables selected from the dataset with corresponding description

Variables	Description
Relevant Emails freq Relevant Emails dura	Absolute frequency or duration of viewing or writing an email with relevant content such as the task assignment within the domain-specific tool email client (mandatory use)
Irrelevant Email freq Irrelevant Email dura	Absolute frequency or duration of viewing an email with irrelevant content within the domain-specific tool email client (voluntary use)
Relevant documents freq Relevant documents dura	Absolute frequency or duration of viewing relevant pdf documents (with information about the model company, its purchases and expenses as well as technical and economic information) within the domain-specific document viewer (mandatory use)
Irrelevant documents freq Irrelevant documents dura	Absolute frequency or duration of viewing an irrelevant pdf document or images (with information about the model company, its purchases and expenses as well as technical and economic information) within the domain-specific document viewer (voluntary use)
Relevant spreadsheet freq Relevant spreadsheet dura	Absolute frequency or duration of working in or viewing a relevant spreadsheet within the domain-specific spreadsheet program (mandatory use)
Irrelevant spreadsheet freq Irrelevant spreadsheet dura	Absolute frequency or duration of working with or viewing an irrelevant spreadsheet within the domain-specific spreadsheet program (voluntary use)
Notepad freq Notepad dura	Absolute frequency or duration of using the notepad (voluntary use)
Calculator freq Calculator dura	Absolute frequency or duration of using the calculator (voluntary use)

Note: Freq stands abbreviated for frequency; dura stands abbreviated for duration.

scenarios, that is, an average of 23 activities from each participant in each scenario. In a second step, to investigate patterns of tool use and information retrieval (RQ 1), we computed variables for the frequency and duration of each participant's activity for each scenario (Table 3). In the third step, we divided the processing time into three equal intervals to gain a deeper understanding of how tool use and information retrieval were distributed over time within a scenario (RQ2). Each interval lasts 10 min in all three scenarios.

Descriptive statistics of the data (Table 4) reveal that the participants spent the majority of their time on the Relevant Spreadsheets in all three scenarios. Moreover, the average frequency of using the Relevant Spreadsheets in two scenarios is at the top compared to the other variables. The second most time was devoted to the use of Relevant Emails.

3.5 | Cluster analysis of log data

Cluster analysis is an exploratory method of multivariate data analysis and an unsupervised data mining method that provides great potential for profiling students' navigation (Bera & Liu, 2006; Kotsiantis et al., 2013; Lawless & Kulikowich, 1996). As an extensively used clustering algorithm (Bharara et al., 2018), k-means appears to be most appropriate for exploratory identification of students' behaviour (e.g., Khare et al., 2018; Kotsiantis et al., 2013). Numerous studies with similar research questions have chosen a similar approach (e.g., Angeli & Valanides, 2013; Bouchet et al., 2012; Tóth et al., 2014) and demonstrated successful analyses.

The clustering algorithm 'k-means' groups similar data (called clusters) so that participants or objects in the same cluster are homogeneous. At the same time, there is heterogeneity across clusters (Angeli & Valanides, 2013). K-means offers several advantages. First, it is highly efficient compared to the hierarchical algorithm since it does not compute all possible distances but reassigns cases to clusters repeatedly. That means the same case can move from cluster to cluster during analysis (Antonenko et al., 2012; Xing et al., 2014). Second, the cases are exclusively assigned to one cluster (Chandrasekaran et al., 2022), so no overlapping exists compared to soft clustering.

The participants were clustered according to the behaviour-related variables presented in Table 1 (RQ1). The optimal number of clusters results from the Elbow method (Marutho et al., 2018; Figure B in Appendix B). In order to investigate whether groups' behaviour in general differs significantly in tool use over time and information retrieval (RQ2) and performance (RQ3), and, if so, which cluster differs from another cluster, a mean values comparison and Kruskal–Wallis tests with corresponding pairwise post-hoc tests were conducted. The alpha level was 5%. Kruskal-Wallis was preferred to an ANOVA because the clusters differ in size and the considered variables are not always normally distributed according to Shapiro–Wilk tests. Data analyses were conducted in the statistical R studio (R Core Team, 2021). Essential packages were "tidyverse" (R 4.1.0; R Core Team, 2021; "stringr" (R 4.1.0; R Core Team, 2021), "stringi" (R 4.1.1, Gagolewski, 2021) and "stats" (R 4.1.1; R Core Team, 2021).

4 | RESULTS

4.1 | Clusters of problem-solving behaviour regarding frequency and duration (RQ 1)

Tool use and document retrieval allow student clusters to be identified. The k-means algorithm provides first indications of a possible cluster solution. The Elbow criterion suggests a four-cluster-solution. Tables 5, 6 and 7 illustrate the cluster solution and give the corresponding mean values per cluster for each behaviour-related variable. This allows a first insight into the unique characteristics of each cluster.

TABLE 4 Descriptive statistics of tool use and information retrieval according to frequency and duration

	Scenario 1				Scenario 2				Scenario 3			
Variable	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
Relevant Emails freq	6.53	2.60	2.00	17.00	8.29	3.42	1.00	22.00	5.90	2.31	2.00	15.00
Relevant Emails dura	342.96	184.42	60.58	1065.52	348.07	179.01	47.76	1314.94	357.31	209.52	54.59	1263.58
Irrelevant Emails freq	1.41	0.88	0.00	6.00	na	na	na	na	1.13	0.66	0.00	4.00
Irrelevant Emails dura	30.37	30.55	0.00	350.24	na	na	na	na	41.58	71.48	0.00	545.5
Relevant documents freq	1.29	2.88	0.00	30.00	**15.91**	8.42	0.00	48.00	7.23	4.74	0.00	23.00
Relevant documents dura	15.47	29.84	0.00	204.60	173.5	98.93	0.00	586.7	135.74	92.30	0.00	616.22
Irrelevant documents freq	8.66	9.29	0.00	55.00	1.12	1.46	0.00	7.00	**8.24**	6.58	0.00	55.00
Irrelevant documents dura	56.12	75.23	0.00	357.50	11.91	25.18	0.00	217.52	149.57	105.23	0.00	557.82
Relevant spreadsheets freq	**8.84**	4.07	0.00	23.00	13.29	6.49	0.00	32.00	**11.88**	7.98	0.00	49.00
Relevant spreadsheets dura	**518.3**	252.84	0.00	1337.7	**538.0**	259.11	0.00	1191.3	**409.9**	247.68	0.00	1169.6
Irrelevant spreadsheets freq	na	na	na	na	0.33	1.04	0.00	14.00	na	na	na	na
Irrelevant spreadsheets dura	na	na	na	na	6.34	28.62	0.00	398.73	na	na	na	na
Notepad freq	3.45	4.59	0.00	32.00	7.44	6.62	0.00	32.00	4.29	5.19	0.00	30.00
Notepad dura	105.30	164.27	0.00	971.94	273.2	266.93	0.00	1135.30	161.51	226.89	0.00	1275.71
Calculator freq	3.79	2.81	0.00	21.00	3.28	3.08	0.00	21.00	1.95	1.87	0.00	14.00
Calculator dura	287.23	271.89	0.00	1172.90	113.81	132.99	0.00	760.94	65.6	141.86	0.00	970.732

Note: 'na' means that this variable or cognitive tool was not available in the respective scenario; duration is given in seconds. Bold numbers indicate the highest mean value for the given variable across the clusters.

Based on their characteristics, we identified four student groups as represented in the four profiles: (1) 'Note Takers', (2) 'Spreadsheet Users', (3) 'Calculator Users', and (4) 'Email Assignment Readers'. As shown in Tables 5–7, participants from cluster 1 ($n = 115$ students) show the highest number of activities ($M_{total\_freq} = 137.72$). The students' behaviour from cluster 1 differs significantly from the other clusters in all three scenarios, showing the highest use of the notepad regarding frequency ($M_{sc1freq} = 5.71$; $M_{sc2freq} = 13.83$; $M_{sc3freq} = 8.57$) and average time of notepad use ($M_{sc1dura} = 197.05$ s; $M_{sc2dura} = 585.6$ s; $M_{sc3dura} = 365.5$ s). Therefore, we label cluster 1 as the 'Note Takers'. Moreover, looking at the access of relevant ($M_{sc1freq} = 1.96$) and irrelevant ($M_{sc1freq} = 11.81$) documents embedded in the office simulation, students from cluster 1 ('Note Takers') with the highest access in scenario 1, changed their behavioural pattern in scenarios 2 and 3. Cluster 2 ($n = 152$ students) overtook cluster 1 in viewing documents and became the top viewer of relevant documents ($M_{sc2freq} = 1.23$ and $M_{sc3freq} = 8.59$) and irrelevant documents ($M_{sc2freq} = 8.84$ and $M_{sc3freq} = 8.59$). Cluster 2 consists of students with a moderate engagement regarding the frequency of activities ($M_{total\_freq} = 133.37$) and spent the most time within the scenarios in comparison to the other clusters ($M_{total\_time} = 4265.73$ s). In this cluster, students demonstrate the most intensive use of the spreadsheet program in all three scenarios regarding frequency ($M_{sc1freq} = 9.55$; $M_{sc2freq} = 16.96$; $M_{sc3freq} = 16.25$) and time spent ($M_{sc1dura} = 635.70$ s; $M_{sc2dura} = 739.38$ s; $M_{sc3dura} = 584.3$ s). Based on this unique behaviour profile, cluster 2 consists of Spreadsheet Users. The students of clusters 3 ($n = 84$ students) and 4 ($n = 81$ students) are not as active as those in clusters 1 and 2. Cluster 3 demonstrates high mean values of duration of applying the calculation tool (for example $M_{sc1dura} = 691.73$ s; $M_{sc2dura} = 193.19$ s). Cluster 4 showed the highest total overall frequency ($M_{total\_freq} = 21.77$) as well as highest total time spent ($M_{total\_dura} = 1527$ s) on reading emails with the task instruction among others. Due to their tool use behaviour, we labelled cluster 3 as 'Calculator Users' and cluster 4 as 'Email Assignment Readers'.

4.2 | Tool use and information retrieval distributed over time (RQ2)

In order to investigate tool use and information retrieval over time (RQ2), students' behaviour was examined at three different time intervals (I–III). A students' dropout was noticed over time (please refer to Table C in Appendix C).

As expected, tool use and information retrieval also varied over time. However, it also turns out that behaviour patterns can be both strongly scenario-dependent and scenario-independent.[1]

Focusing on voluntary *notepad* use, over all scenarios, this supportive tool was used the least at the end of the scenarios (Figure 2a).

TABLE 5 Average frequency and mean duration (in seconds) of tool use and information retrieval, Kruskal-Wallis-test and corresponding pairwise comparisons in scenario 1

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests					
Number of students	$n = 115$	$n = 152$	$n = 84$	$n = 81$							
Scenario 1	M	M	M	M	p						
Relevant Email freq	6.30	**7.20**	5.69	6.43	0.000***	cluster 1 < cluster 2*			cluster 2 > cluster 3***	cluster 2 > cluster 4 *	
Relevant Email dura	319.23	365.06	264.36	**416.67**	0.000***	cluster 1 < cluster2***	cluster 1 > cluster 3***	cluster 1 < cluster 4***	cluster 2 > cluster 3***	cluster 2 < cluster 4 *	cluster 3 < cluster 4***
Irrelevant Email freq	1.41	**1.50**	1.32	1.33	0.325						
Irrelevant Email dura	24.66	29.62	29.55	**40.75**	0.102						
Relevant documents freq	**1.96**	1.61	0.51	0.56	0.000***		cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4**	
Relevant documents dura	**22.00**	18.93	6.30	9.25	0.000***		cluster 1 > cluster 3***	cluster 1 > cluster 4**	cluster 2 > cluster 3***	cluster 2 > cluster 4 **	
Irrelevant documents freq	**11.81**	9.48	5.04	6.40	0.000***	cluster 1 > cluster 2**	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4 *	
Irrelevant documents dura	63.59	**66.83**	31.38	51.08	0.000***	cluster 1 < cluster 2**	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4 *	
Relevant spreadsheet freq	9.52	**9.55**	7.95	7.43	0.000***		cluster 1 > cluster 3*	cluster 1 > cluster 4**	cluster 2 > cluster 3**	cluster 2 > cluster 4 **	
Relevant spreadsheet dura	490.63	**635.70**	341.26	520.95	0.000***	cluster 1 < cluster 2***	cluster 1 > cluster 3***		cluster 2 > cluster 3***	cluster 2 > cluster 4 **	cluster 3 < cluster 4***
Irrelevant spreadsheet freq	na	na	na	na	na	na	na	na	na	na	na
Irrelevant spreadsheet dura	na	na	na	na	na	na	na	na	na	na	na
Notepad freq	**5.71**	3.05	2.52	1.94	0.000***	cluster 1 > cluster 2**	cluster 1 > cluster 3***	cluster 1 > cluster 4**	cluster 2 > cluster 3***	cluster 2 > cluster 4 *	
Notepad dura	**197.05**	84.33	71.47	49.46	0.000***	cluster 1 > cluster 2***	cluster 1 > cluster 3***	cluster 1 > cluster 4***		cluster 2 > cluster 4 *	

TABLE 5 (Continued)

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests					
Calculator freq	3.77	3.02	**6.27**	2.72	0.000***	cluster 1 > cluster 2**	cluster 1 < cluster 3***	cluster 1 > cluster 4**	cluster 2 < cluster 3***		cluster 3 > cluster 4***
Calculator dura	223.28	175.34	**691.73**	168.52	0.000***		cluster 1 < cluster 3***		cluster 2 < cluster 3***		cluster 3 > cluster 4***
Total frequency of activities (Scenario 1)	**40.48**	**35.41**	**29.31**	**26.80**	0.000***	cluster 1 > cluster 2*	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3**	cluster 2 > cluster 4*	cluster 3 > cluster 4
Average time spent in seconds (Scenario 1)	**1340.43**	**1375.81**	**1436.05**	**1256.69**	0.163	cluster 1 < cluster 2	cluster 1 < cluster 3	cluster 1 > cluster 4**	cluster 2 < cluster 3**	cluster 2 > cluster 4 **	cluster 2 < cluster 3**

Note: 'na' means that this variable or cognitive tool was not available in the respective scenario; duration is given in seconds. Bold numbers indicate the highest mean value for the given variable across the clusters.

TABLE 6 Average frequency and average duration (in seconds) of tool use and information retrieval, Kruskal-Wallis-test and corresponding pairwise comparisons in scenario 2

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests					
Scenario 2	**M**	**M**	**M**	**M**	**p**						
Relevant Email freq	**8.97**	7.67	8.13	8.62	0.024*	cluster 1 > cluster 2*					
Relevant Email dura	280.65	312.37	360.79	**497.58**	0.000**		cluster 1 < cluster 3***	cluster 1 < cluster 4***	cluster 2 < cluster 3*	cluster 2 < cluster 4***	cluster 3 < cluster 4***
Irrelevant Email freq	na	na	na	na	na	na	na	na	na	na	na
Irrelevant Email dura	na	na	na	na	na	na	na	na	na	na	na
Relevant Documents freq	12.84	**19.11**	1.65	13.64	0.000***	cluster 1 < cluster2***	cluster 1 > cluster 3**		cluster 2 > cluster 3*	cluster 2 > cluster 4***	cluster 3 < cluster 4*
Relevant documents dura	117.14	**217.87**	185	158.26	0.000**	cluster 1 < cluster 2***	cluster 1 > cluster 3***	cluster 1 < cluster 4*	cluster 2 > cluster 3*	cluster 2 > cluster 4***	

(Continues)

TABLE 6 (Continued)

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests					
Irrelevant documents freq	1.04	**1.23**	1.11	1.02	0.502						
Irrelevant documents dura	9.99	**13.12**	11	12.92	0.516						
Relevant spreadsheet freq	10.82	**16.96**	12.29	9.90	0.000***	cluster 1 < cluster 2***	cluster 1 < cluster 3**		cluster 2 > cluster 3***	cluster 2 > cluster 4***	cluster 3 > cluster 4**
Relevant spreadsheet dura	391.83	**739.38**	500	406.94	0.000**	cluster 1 < cluster 2***	cluster 1 < cluster 3***				
Irrelevant spreadsheet freq	0.40	0.33	0.29	0.27	0.582						
Irrelevant spreadsheet dura	7.15	5.77	**7.88**	4.66	0.568		cluster 1 > cluster 3***				
Notepad freq	**13.83**	4.44	6.68	4.80	0.000***	cluster 1 > cluster 2***	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 < cluster 3**		cluster 3 > cluster 4*
Notepad dura	**585.6**	124.43	224.44	159.61	0.000**	cluster 1 > cluster 2***	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 < cluster 3**		cluster 3 > cluster 4*
Calculator freq	3.34	2.78	**4.90**	2.46	0.000***		cluster 1 < cluster 3***	cluster 1 > cluster 4*	cluster 2 < cluster 3***		cluster 3 > cluster 4***
Calculator dura	117.21	89.14	**193.19**	72.94	0.000**		cluster 1 < cluster 3***	cluster 1 > cluster 4*	cluster 2 < cluster 3***		cluster 3 > cluster 4***
Total frequency of activities (Scenario 2)	**51.24**	**52.51**	**35.05**	**40.72**	**0.059**		cluster 1 > cluster 3***		cluster 2 > cluster 3***		cluster 3 > cluster 4**
Average time spent in seconds (Scenario 2)	**1509.56**	**1502.08**	**1482.49**	**1312.91**	**0.002 ****		cluster 1 > cluster 3***	cluster 1 > cluster 4	cluster 2 > cluster 3***	cluster 2 > cluster 4	cluster 3 > cluster 4***

Note: 'na' means that this variable or cognitive tool was not available in the respective scenario; duration is given in seconds. Bold numbers indicate the highest mean value for the given variable across the clusters.

TABLE 7 Average frequency and average duration (in seconds) of tool use and information retrieval, Kruskal-Wallis-test and corresponding pairwise comparisons in scenario 3

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests					
Scenario 3											
Relevant Email freq	6.15	5.51	5.49	**6.72**	0.000***	cluster 1 > cluster 2**	cluster 1 > cluster 3*			cluster 2 < cluster 4*	cluster 3 < cluster 4*
Relevant Email dura	280.5	287.7	342.1	**612.7**	0.000**	cluster 1 < cluster 2*	cluster 1 < cluster 3***	cluster 1 < cluster 4***	cluster 2 < cluster 3 **	cluster 2 < cluster 4***	cluster 3 < cluster 4***
Irrelevant Email freq	**1.21**	1.13	1.05	1.12	0.348						
Irrelevant Email dura	27.9	32.8	54.8	**63.7**	0.348						
Relevant documents freq	8.14	**8.59**	6.07	4.62	0.000***		cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4***	cluster 3 > cluster 4*
Relevant documents dura	127.2	**163.0**	129.9	102.8	0.000**	cluster 1 < cluster 2**		cluster 1 > cluster 4**	cluster 2 > cluster 3 ***	cluster 2 > cluster 4***	cluster 3 > cluster 4*
Irrelevant documents freq	8.45	**8.84**	8.19	6.85	0.006**			cluster 1 > cluster 4*		cluster 2 > cluster 4**	
Irrelevant documents dura	145.1	147.6	**179.5**	128.5	0.048*						cluster 3 > cluster 4*
Relevant spreadsheet freq	11.26	**16.25**	10.56	5.90	0.000***	cluster 1 < cluster 2***		cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4***	cluster 3 > cluster 4***
Relevant spreadsheet dura	359.7	**584.3**	383.2	181.7	0.000**	cluster 1 < cluster 2***		cluster 1 > cluster 4***	cluster 2 > cluster 3 ***	cluster 2 > cluster 4***	cluster 3 > cluster 4***
Irrelevant spreadsheet freq	na	na	na	na	na	na	na	na	na	na	na
Irrelevant spreadsheet dura	na	na	na	na	na	na	na	na	na	na	na
Notepad freq	**8.57**	3.28	2.56	1.94	0.000***	cluster 1 > cluster 2***	cluster 1 > cluster 3***	cluster 1 > cluster 4***		cluster 2 > cluster 4*	
Notepad dura	**365.5**	109.3	82.0	52.3	0.000**	cluster 1 > cluster 2***	cluster 1 > cluster 3***	cluster 1 > cluster 4 ***		cluster 2 > cluster 4**	
Calculator freq	**2.23**	1.86	2.15	1.49	0.046*			cluster 1 > cluster 4*		cluster 2 > cluster 4*	cluster 3 > cluster 4*

(Continues)

TABLE 7 (Continued)

	Note-takers (cluster 1)	Spreadsheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post-hoc tests
Calculator dura	82.7	63.1	**85.1**	25.7	0.118	
Total average frequency of activities (Scenario 3)	**46.00**	45.45	36.07	28.64	0.000***	cluster 1 > cluster 3***; cluster 1 > cluster 4***; cluster 2 > cluster 3***; cluster 2 > cluster 4***; cluster 3 > cluster 4**
Average time spent in seconds (Scenario 3)	**1388.72**	1387.83	1256.48	1167.54	0.000***	cluster 1 > cluster 3***; cluster 1 > cluster 4***; cluster 2 > cluster 3***; cluster 2 > cluster 4***
Total average frequency of activities in (scenario 1–3)	**137.72**	133.37	100.43	96.16	0.000***	cluster 1 > cluster 3***; cluster 1 > cluster 4***; cluster 2 > cluster 3***; cluster 2 > cluster 4**; cluster 3 > cluster 4**
Average time spent in seconds (scenario 1–3)	4238.71	**4265.73**	4175.01	3737.14	0.001**	cluster 1 > cluster 3**; cluster 2 > cluster 3***; cluster 3 > cluster 4***

Note: 'na' means that this variable or cognitive tool was not available in the respective scenario; duration is given in seconds. Bold numbers indicate the highest mean value for the certain variable across the clusters.

A closer look at the clusters for each scenario reveals additional insights (Figures 2b–d). When solving scenario 1, students in clusters 1 and 2 used the notepad especially at the midpoint of the allotted time while students in clusters 3 and 4 used it intensively in the first ten min (Figures 2b). However, in scenarios 2 and 3 the clusters show a different behaviour. In scenario 2, for all clusters, notepad use occurred especially at the beginning and then decreased over time (Figure 2c), whereas in scenario 3, in all clusters, the highest use occurred in the middle of the time interval (Figure 2d).

Regarding the knowledge-based method, the use of the mandatory tool *document viewer* also appears to be scenario-specific (Figure 3a). Having a closer look at the clusters for each scenario gives additional insights. Focusing on scenario 1, in clusters 1 and 2, activities increase steadily over time (Figure 3b). However, in clusters 3 and 4, activities increase slightly and stay constant more or less for the remaining time. A different picture emerges in scenarios 2 and 3 (Figure 3c and d). Retrieval of the relevant content in the mandatory cognitive tool *document viewer* decreases equally over time across all clusters.

In contrast, use of the mandatory *spreadsheet program* (strong method) shows that cognitive tool use is not always scenario-specific (Figure 4a–d). Going in-depth over all clusters, the tendency to use the *spreadsheet program* is at its highest at the second time interval in all scenarios. Interestingly, use of another strong method (*email client*) shows the same pattern over time in all scenarios and clusters (see Figure D in Appendix D).

Against this background, how the distribution of behaviour over time is related to student performance will be investigated in the following.

4.2.1 | Differences in terms of problem-solving performance (RQ3)

The four clusters differ significantly in their problem-solving performance (Table 8). The Note Takers (cluster 1) achieved the highest mean score ($M_{total} = 21.72$), closely followed by the Spreadsheet Users (cluster 2; $M_{total} = 21.36$). The two top-performing clusters do not show significant performance differences, indicating that both behaviour patterns can lead to high performance. The Calculator Users (cluster 3) gained a moderate achievement level ($M_{total} = 17.24$). The Email Assignment Readers (cluster 4) received the lowest scores among the scenarios ($M_{total} = 15.06$).

The Kruskal-Wallis tests reported 10 out of 12 significant differences among the clusters regarding the four subdimensions of problem-solving competence A1-A4 (see Table 8). In all three scenarios, the Note Takers (cluster 1) achieved the highest partial credit score of the performance subdimension A1. Students might take actions by making notes after identifying and viewing relevant resource documents. Students from cluster 2 (Spreadsheet Users) met the requirement of applying domain-specific processing algorithms such as correctly applied calculation schemes within the domain-specific spreadsheet tool (performance subdimension A2) the most and gained the highest score of A2 among the clusters.

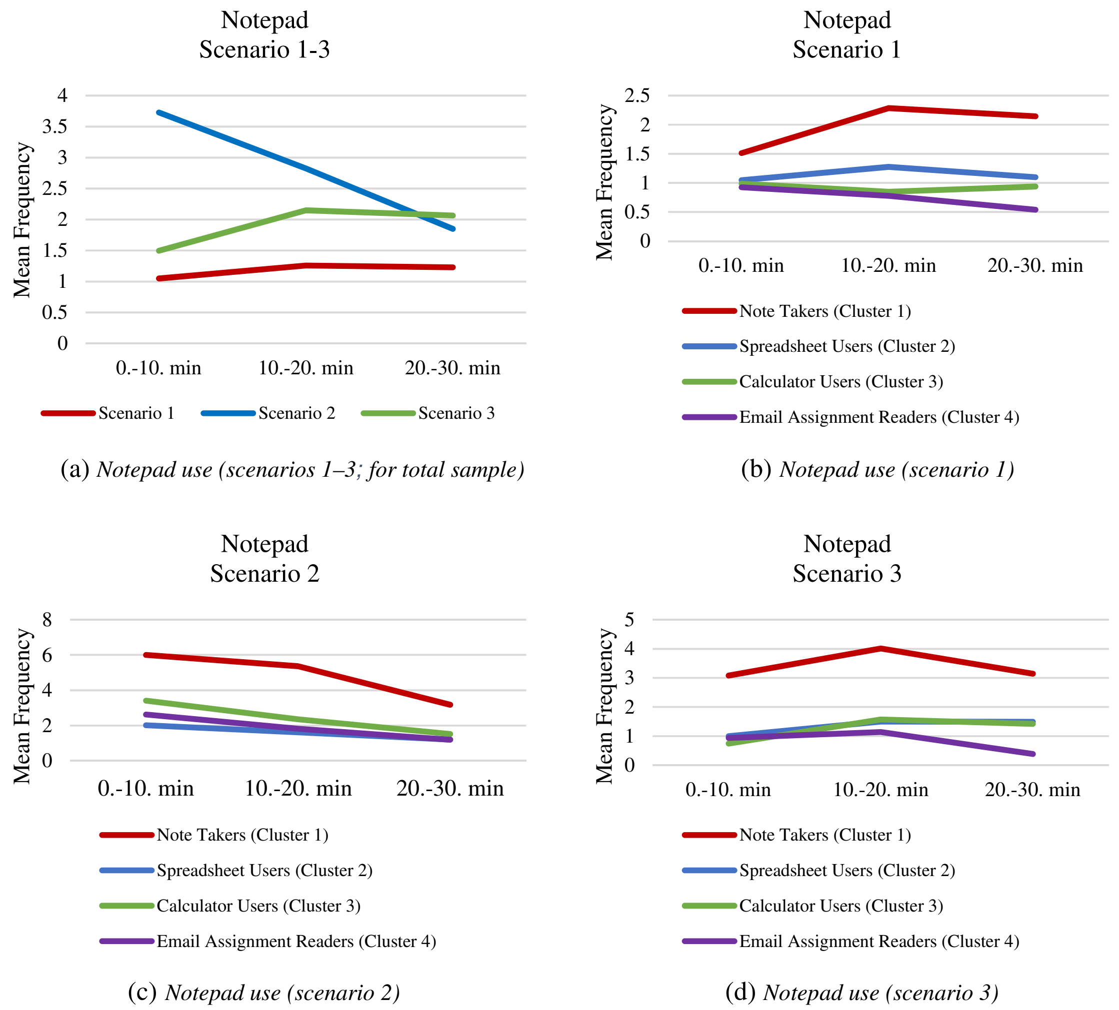
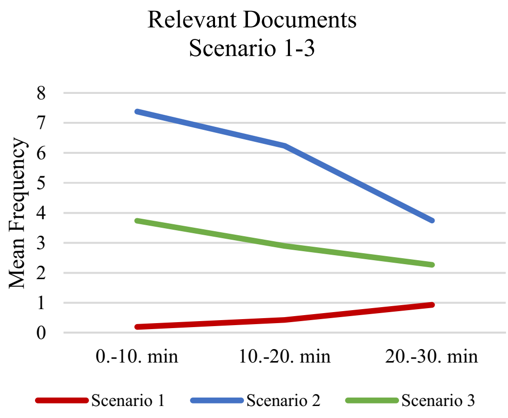


FIGURE 2 Notepad use distributed over time. (a) Notepad use (scenarios 1–3), (b) Notepad use (scenario 1), (c) Notepad use (scenario 2) and (d) Notepad use (scenario 3)

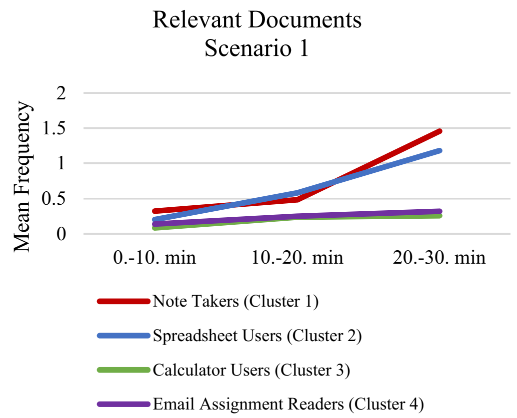
Because cluster 1 made well-founded decisions in two scenarios, they achieved the highest partial credit scores in subdimension A3. Regarding A4 (adequate communication), none of the clusters showed striking or unique performance characteristics in any scenario, indicating that this subdimension is somewhat detached from the other subdimensions.

5 | DISCUSSION

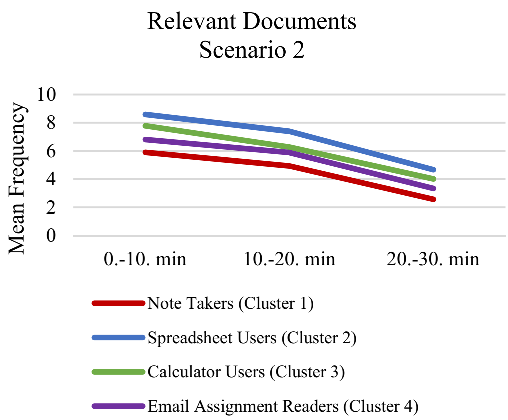
The cluster analysis based on behavioural data revealed significant findings that will be discussed by framing them in context with the above literature. The identification of four clusters indicates that students follow unique problem-solving behaviours and methods (RQ1). Both process indicators, *frequency* as well as *duration* of tool use and information retrieval, give insights into different tendencies towards different types of cognitive tools and information sources. Some students prefer domain-specific tools (e.g., *spreadsheet program*) while some solve their tasks by means of domain-general and weak methods (e.g., note taking). In addition, the high provision of relevant documents was embraced by certain students. In line with Sabourin et al. (2012), for example, empirical evidence for different patterns of self-regulated behaviour exists when approaching domain-specific problems. The researchers confirmed that there are students that indeed prefer an intensive use of resources (e.g., books and posters) and the notebook compared with other students. This also implies that the process indicators, *frequency* as well as *duration* of tool use and information retrieval, give insights into different engagement levels of performed activity across the clusters. These findings are similar, for example, to the results of Bera and Liu (2006). They identified three student groups with an average, high and low tool access rate based on the same process indicators used within the hypermedia learning environment *Alien Rescue*. In line with the findings of Lust and colleagues (2011), accessing those tools that contain information material is the most frequently performed activity in computerized environments. In our study, the document viewer and especially the central mandatory tool, the spreadsheet program – both containing information – recorded high use frequency. Problem solvers reach for different tools, seek information from several sources, and apply different problem-solving strategies since (1) problems might have more



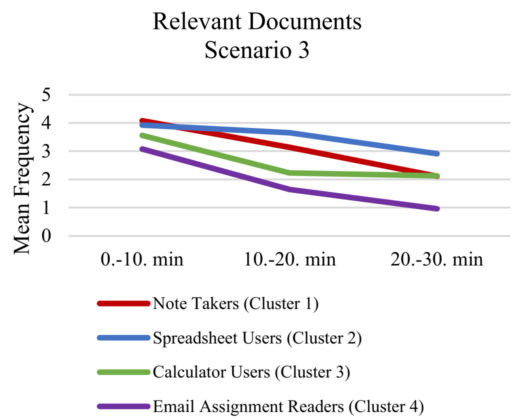
(a) *Relevant Document (scenario 1–3; for total sample)*



(b) *Relevant Document (scenario 1)*



(c) *Relevant Document (scenario 2)*



(d) *Relevant Document (scenario 3)*

FIGURE 3 Relevant document retrieval distributed over time. (a) Relevant document (scenario 1–3), (b) Relevant document (scenario 1), (c) Relevant document (scenario 2) and (d) Relevant document (scenario 3)

than one possible solution path (*polytely*) and (2) problem solvers may have limited knowledge.

Findings regarding distributed tool use over time and information retrieval (RQ2) are also similar to some findings revealed by Bera and Liu (2006). Supportive cognitive tools, such as the *notepad* and *calculator*, as well as sharing cognitive tools, such as the *document viewer* (with relevant content), were used in the earlier and middle problem-solving stages. Collecting and organizing information in the first and second third of the proposed time might be an indicator for orientation and exploration phases, respectively. However, because this finding does not count for the first scenario, the cognitive procedure of how problems are solved is scenario specific. This finding is in line with Behrens et al. (2012) who argue that cognitive processes depend on the task design. Moreover, although some calculating functions are feasible in both the voluntary and mandatory calculation tools, small cannibalistic effects can be detected in the scenarios (see Appendix E). It might be assumed that using the simpler calculator intensively in the first 10 min followed by an increased use of the more complex spreadsheet mirrors a typical domain-specific problem-solving procedure in the workplace. When using the calculator first for simple calculations due to its simplicity, problem solvers realized that a spreadsheet might be more helpful and time-efficient due to its extensive functions. Moreover, cluster 4 shows a notable behaviour since they mainly read the emails that contained the task assignment. This behaviour could be a sign that they did not get beyond the initial phases such as the problem definition and planning phase.

Tool-use behaviour and information retrieval influence problem-solving performance (RQ3). Similar to other researchers (e.g., Hung & Crooks, 2009, Hung & Zhang, 2008), we found significant relations between problem-solving behaviour and problem-solving performance. Our findings indicate that more intense interactions between the problem solver and the business simulation led to a better overall performance in general. In their study, Li and Tsai (2017), who identified three clusters labelled as 'consistent use students', 'slide intensive use students' and 'less use students', show evidence that the two clusters with a higher activity

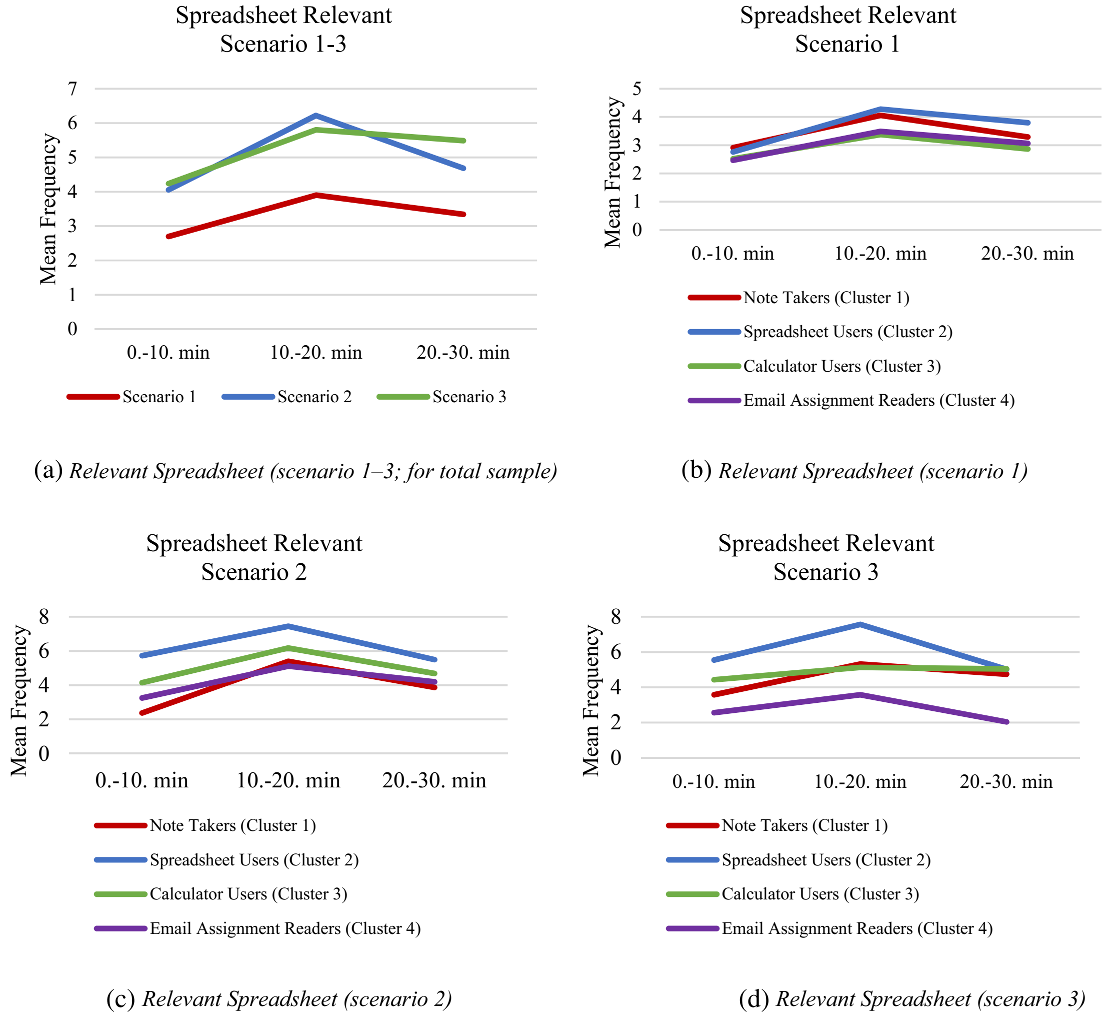


FIGURE 4 Relevant spreadsheet use distributed over time. (a) Relevant spreadsheet (scenario 1–3), (b) Relevant spreadsheet (scenario 1), (c) Relevant spreadsheet (scenario 2) and (d) Relevant spreadsheet (scenario 3)

rate show higher performances than the 'less use students'. Similar findings were revealed by Sabourin et al. (2012) in their study that found high SRL students to demonstrate higher resource retrieval and note taking than lower SRL students. From an educator perspective, teachers should encourage their learners to interact more often with the material. Moreover, DiCerbo (2014, S. 18) and Dumdumaya et al. (2018) argue that the total time spent on task-relevant events or 'time spent for a solved problem' are indicators of student persistence, which might also have positive effects on problem solving. Our study's results confirm this hypothesis as the two top performing clusters showed higher time consumption on the problem than weaker problem solvers. Having a closer look at the tendency towards the different types of cognitive tools and its impact on performance scores, we found similar results compared to other researchers. In line with Trafton and Trickett (2001), tools might be beneficial and supportive but not really necessary to successfully solve a problem. Instead, in our study, note taking and the calculator are assigned to voluntary use and are not relevant to the solution. Voluntary note taking, however, proves itself to be a helpful strategy for successful cognitive and meta-cognitive processing (Nesbit et al., 2007; Trafton & Trickett, 2001) as it supports the problem solvers in reducing cognitive burden (Moos, 2009). Within their study, Trafton and Trickett (2001) found many competitive arguments for notepad use including the advantage that participants who used the notepad solved the problems in a more outright manner.

To what extent the calculator still helped the weaker cluster 3 (Calculator Users) in our study, or prevented them from using the more efficient spreadsheet program, is questionable (and should be researched further). This potential cannibalism effect was also touched upon by Zydney (2010) who found that tools with similar functions could negatively influence behaviour. The 'Calculator Users' have an idea of what to do, but might not be able to master the central tool, the spreadsheet application. Interestingly, Marasigan (2018, p. 166) concluded that calculators facilitate students' learning in the mathematics domain because the "calculators allow students to spend less time on tedious calculations and more time on understanding and solving problems. It helps students develop better number sense and

TABLE 8 Mean scores of problem-solving performances across clusters

	Note-takers (cluster 1)	Spread-sheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post hoc test				
Number and ratio of students	n = 115	n = 152	n = 84	n = 81						
Scenario 1										
A1	**1.88**	1.85	1.60	1.22	0.000**		cluster 1 > cluster 4***		cluster 2 > cluster 4 ***	cluster 3 > cluster 4*
A2	2.84	**2.99**	2.70	2.49	0.045*				cluster 2 > cluster 4*	
A3	**0.97**	0.96	0.51	0.54	0.007**	cluster 1 > cluster 3*	cluster 1 > cluster 4*	cluster 2 > cluster 3*	cluster 2 > cluster 4*	
A4	2.03	**2.04**	1.64	1.73	0.005**	cluster 1 > cluster 3*	cluster 1 > cluster 4*			
Partial credit score (Scenario 1)	**7.73**	**7.84**	**6.45**	**5.99**	0.000***	cluster 1 > cluster 3*	cluster 1 > cluster 4**	cluster 2 > cluster 3***	cluster 2 > cluster 4***	
Scenario 2										
A1	**2.42**	2.35	1.71	1.52	0.000**	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3**	cluster 2 > cluster 4 ***	
A2	1.45	**1.70**	0.98	0.65	0.000**	cluster 1 > cluster 3**	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4 ***	
A3	1.58	1.69	**1.77**	1.52	0.665					
A4	2.15	2.16	**2.21**	1.86	0.165					
Partial credit score (Scenario 2)	**7.60**	**7.90**	**6.68**	**5.56**	0.000***	cluster 1 > cluster 3*	cluster 1 > cluster 4	cluster 2 > cluster 3**		cluster 3 > cluster 4*
Scenario 3										
A1	**1.83**	1.39	0.71	0.43	0.000**	cluster 1 > cluster 3***	cluster 1 > cluster 4***	cluster 2 > cluster 3***	cluster 2 > cluster 4 ***	
A2	0.76	**0.81**	0.46	0.11	0.000**	cluster 1 > cluster 3*	cluster 1 > cluster 4***	cluster 2 > cluster 3**	cluster 2 > cluster 4 ***	cluster 3 > cluster 4***

TABLE 8 (Continued)

	Note-takers (cluster 1)	Spread-sheet users (cluster 2)	Calculator users (cluster 3)	Email assignment readers (cluster 4)	Kruskal-Wallis-test	Post hoc test			
A3	**1.67**	1.43	1.12	1.48	0.047*	cluster 1 > cluster 3*			
A4	**2.14**	1.99	1.81	1.49	0.000**	cluster 1 > cluster 3*	cluster 1 > cluster 4***		cluster 2 > cluster 4**
Partial credit score (Scenario 3)	**6.39**	**5.62**	**4.11**	**3.52**	0.000***	cluster 1 > cluster 3***	cluster 1 > cluster 4**	cluster 2 > cluster 3**	cluster 2 > cluster 4*
Overall Mean	**21.72**	**21.36**	**17.24**	**15.06**	0.000***	cluster 1 > cluster 3***	cluster 1 > cluster 4**	cluster 2 > cluster 3***	cluster 2 > cluster 4**

Note: A1: Identification of needs for action and sources of information; A2: Process of information; A3: Making well-founded decisions; A4: Communication of decisions adequately; Bold numbers indicate the highest mean value for the given variable across the clusters.

allow students to study mathematical concepts. It simplifies tasks while helping students determine the best". However, unlike our results, Marasigan (2018) did not embed and consider the more sophisticated spreadsheet in her study. According to Marriott (2004), spreadsheet use fosters a holistic understanding of the business problem, which was also observed for cluster 2 of this study, who solved the problem successfully. As already formulated by Marriott (2004), educators should encourage the use of domain-specific algorithms in spreadsheets.

When problems have more than one potential solution path and problem solvers have limited knowledge of the problem, problem solvers tend to seek and base reasoning on information from several sources (Barzilai & Strømsø, 2018). Behaviour of the second highest performing cluster 2 is – next to its high application of domain-specific algorithms within the relevant spreadsheet – characterized by a high level of document review of relevant information. In contrast, cluster 1 accessed the documents less frequently than cluster 2 and might compensate for this lower frequency with higher note taking behaviour. Interestingly, cluster 1 as the high-performing cluster visited irrelevant documents the most in scenario 1 compared with cluster 2 who demonstrated discriminate reading. However, due to its very low time investment on these, it might be an indicator of the cluster's consciousness. The frequent but brief retrieval of irrelevant documents in scenario 1 could suggest that these students realize that irrelevant information is embedded within the scenarios and, hence, neglected them in the subsequent two problems.

In sum, given that the two high-performing clusters 1 and 2 show that note taking as well as spreadsheet use leads to successful problem solving, it can be concluded there is no 'one best way' (e.g., Rausch et al., 2017) but rather two distinct methods leading to an accurate problem solution. These findings regarding efficient cognitive tools and information retrieval should be considered by teachers when designing problem tasks in the classroom. In the context of task design, identifying typical errors and difficulties through log data analyses linked to student's performance enable successful individual learning. Therefore, teachers could design automated prompts if the student did not open the notepad or spreadsheet. Furthermore, these findings help to improve the design of simulation environments. Using embedded prompts and understanding the obstacles helps to make the simulations adaptive and automated. Furthermore, for curricular designers the findings might be interesting. The promotion and acquisition of problem solving are anchored as important vocational training objectives in the curriculum. Therefore, understanding problem-solving procedures of the overall steps and where exactly problems lie helps to improve curricular requirements.

Despite these significant findings, the present research design has its limitations. First, it is not evident whether students actively read the provided documents or whether they were mentally absent and unfocused. Second, it is still not clear whether students' behaviour, such as tool use and information retrieval, affects problem-solving performance or vice versa. This causality dilemma ('chicken-and-egg' problem) should be considered. Third, it should be taken into account that the performance partial credit score of competence subdimension A1 (dependent variable) is composed and measured based on the frequency of relevant document retrieval (independent variable).

6 | CONCLUSION

Log data analysis has evolved to be promising in problem solving by bringing clarity to problem-solving processes that formerly remained a black box. In particular, cluster analysis enables ways to analyze and diagnose each unique problem-solving process. The given investigation of specific process indicators takes advantage by finding noteworthy patterns regarding (1) domain-specific and -general tool use, (2) the retrieval of relevant and irrelevant information, and (3) different problem-solving methods in an office simulation (RQ1). In addition, the results provide evidence regarding the pathway that students follow over time to solve problems. Problem solving is scenario specific (RQ2) as different problems require different activities at different times. Most importantly, different behaviours lead to greater performance (RQ3), which provides new insights into students' problem solving in the business domain and the hints that need to be promoted. In a nutshell, although there is not one best way to solve problems within the business domain, we found two effective tool use behaviours. Domain-general, voluntary note taking, and domain-specific, solution-relevant spreadsheet use, have been found to be performance enhancing. Moreover, our findings indicate that higher interactions in the business simulation lead to a better overall performance, which supports the results of many other researchers. This newly created perspective facilitates educators and trainers to adapt and extend their instructional approaches such as encouraging students to take notes at the beginning or preferring the spreadsheet to calculators (He, Liao et al., 2019).

Lastly, future research on domain-specific problem-solving environments should be carried out to minimize the above-mentioned limitations. First, analyses of multi-channel data including eye-tracking, facial expression recognition, or thresholds of mouse movements can enlighten and validate inferred interpretations made from the data (Azevedo et al., 2018; Järvelä et al., 2019). Also, think-aloud studies provide additional accurate results (Cowan, 2019; Ericsson & Simon, 1984). Second, further research should examine the content and structure of notes made within the notepad by text mining techniques for a deeper understanding of the top performers – who proved to be the most frequent note takers. Third, the need for further examination also exists for counterparts showing lower performance. Accounting for almost 20% of all participants in this study, more evidence is required to determine how the weak 'Email Assignment Readers' gain a better understanding of the task and how the 'Calculator Users' can leverage their engagement level for the probably more efficient use of spreadsheets or note taking and, thus, their performance. Therefore, the impact of interceding recommendations and hints in general should be considered in curricular requirements. Also, the use of certain cognitive tools at a specific time should be further investigated. Therefore, for future projects, instructional and software designers should consider the question of how to leverage adaptive learning and how to motivate students (e.g., via error detection that triggers individual real-time feedback or, if necessary, turning off certain tools). Teachers should encourage them to explore the environment efficiently and its repertoire of cognitive tools in its entirety. Fostering students' use of cognitive tools efficiently in educational software or classrooms to achieve their personal best is crucial for their development of life-long, problem-solving competences and for facing future challenges in the workplace.

DECLARATION OF INTEREST

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

ACKNOWLEDGEMENT

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 (ASCOT)" under Grant number [01DB1119-23]. Open Access funding enabled and organized by Projekt DEAL.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/jcal.12770.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://www.forschungsdaten-bildung.de/de/studies/415-modellierung-und-messung-domaenenspezifischer-problemloesekompetenz-bei-industriekaufleuten-domplik

ORCID

Sabrina Ludwig https://orcid.org/0000-0002-2920-7561
Andreas Rausch https://orcid.org/0000-0002-0749-2496

REFERENCES

Angeli, C., & Valanides, N. (2013). Using educational data mining methods to assess field-dependent and field- independent learners' complex problem solving. *Educational Technology Research and Development*, *61*(3) (June 2013), 521–548. https://doi.org/10.1007/s11423-013-9298-1

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist*, *40*, 199–209. https://doi.org/10.1207/s15326985ep4004_2

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. In D. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254–270). Routledge.

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In *Handbook of multiple source use* (1st ed., pp. 99–116).

Behrens, J. T., DiCerbo, K. E., & Ferrara, S. (2012). *Intended and unintended deceptions in the use of simulations*. The K-12 Center at ETS, Educational Testing Service.

Bera, S., & Liu, M. (2006). Cognitive tools, individual differences, and group processing as mediating factors in a hypermedia environment. *Computers in Human Behaviour*, *22*(2), 295–319.

Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data mining for Students' disposition analysis. *Education and Information Technologies*, *23*(2), 957–984.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, *19*(3–4), 363–392.

Bosch, N., Zhang, Y., Paquette, L., Baker, R. S., Ocumpaugh, J., & Biswas, G. (2021). Students' verbalized metacognition during computerized learning. In Proceedings of the 2021 CHI conference on human factors in computing systems (CHI '21).

Bouchet, F., Azevedo, R., Kinnebrew, J., & Biswas, G. (2012). Identifying students' characteristic learning behaviours in an intelligent tutoring system fostering self-regulated learning. In: Proceedings of the 5th international conference on educational data mining (pp. 65–72).

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behaviour*, *9*(2–3), 171–184.

Caruso, J. V. (2019, March). Using business simulations to prepare students to think critically, make better decisions, and solve business problems. In developments in business simulation and experiential learning: Proceedings of the annual ABSEL conference (Vol. 46).

Chandrasekaran, D., Chang, M., & Graf, S. (2022). A learning analytics approach to build learner profiles within the educational game OMEGA+. In *International conference on intelligent tutoring systems* (pp. 139–147). Springer.

Cheng, G., & Chau, J. (2016). Exploring the relationships between learning styles, online participation, learning achievement and course satisfaction: An empirical study of a blended learning course. *British Journal of Educational Technology*, *47*(2), 257–278. https://doi.org/10.1111/bjet.12243

Clarebout, G., & Elen, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behaviour*, *22*(3), 389–411.

Cowan, J. (2019). The potential of cognitive think-aloud protocols for educational action-research. *Active Learning in Higher Education*, *20*(3), 219–232.

Csapó, B., & Funke, J. (2017). The development and assessment of problem solving in 21st-century schools.

di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Educational Technology & Society*, *17*(1), 17–28.

Dumdumaya, C. E., Banawan, M. P., & Rodrigo, M. M. T. (2018, July). Identifying students' persistence profiles in problem solving task. In adjunct publication of the 26th conference on user modeling, adaptation and personalization (pp. 281–286).

Duncker, K. (1935). *Zur Psychologie des produktiven Denkens*. Springer.

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, *8*, 1153.

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning.*, *36*, 933–956.

Eigenmann, R., Siegfried, C., Kögler, K., & Egloffstein, M. (2015). Aufgaben angehender Industriekaufleute im controlling: Ansätze zur Modellierung des Gegenstandsbereichs. [prospective industrial clerks' tasks in controlling: Approaches to domain modeling]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *111*(3), 417–436.

Elger, M., Bertuzi, R., Silva, P., & Göppner, R. (2018). Learning change management by using a business simulation game. *Planspiele-Interaktion gestalten: Über die Vielfalt der Methode*, *10*, 159.

Ericsson, K., & Simon, H. (1984). *Protocol analysis: Verbal reports as data*. The MIT Press.

Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 3–25). Erlbaum.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Stemberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Lawrence Erlbaum Associates.

Funke, J. (2012). Complex problem solving. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 682–685). Springer.

Funke, J., Fischer, A., & Holt, D. V. (2017). When less is less: Solving multiple simple problems is not complex problem solving – A comment on Greiff et al. (2015). *Journal of Intelligence*, *5*(1), 5. https://doi.org/10.3390/jintelligence5010005

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

Gagolewski, M. (2021). Stringi: Fast and portable character string processing in R. *Journal of Statistical Software*, *103*(2), 1–59.

Geiger, V., Goos, M., & Dole, S. (2015). The role of digital technologies in numeracy teaching and learning. *International Journal of Science and Mathematics Education*, *13*(5), 1115–1137.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608–626.

Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4), 203–209.

Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving*, *5*(2), 71–91.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46.

Hahnel, C., Goldhammer, F., Kroehne, U., Mahlow, N., Artelt, C., & Schoor, C. (2021). Automated and controlled processes in comprehending multiple documents. *Studies in Higher Education*, *46*(10), 2074–2086.

He, Q., Borgonovi, F., & Paccagnella, M. (2019). Using process data to understand adults' problem-solving behaviour in the programme for the international assessment of adult competencies (PIAAC): Identifying generalised patterns across multiple tasks with sequence mining.

He, Q., Liao, D., & Jiao, H. (2019). Clustering behavioural patterns using process data in PIAAC problem-solving items. In *Theoretical and practical advances in computer-based educational measurement* (pp. 189–212). Springer.

Huang, S. C., Jin, L., & Zhang, Y. (2017). Step by step: Sub-goals as a source of motivation. *Organizational Behaviour and Human Decision Processes*, *141*, 1–15.

Hung, J. L., & Crooks, S. M. (2009). Examining online learning patterns with data mining techniques in peer-moderated and teacher-moderated courses. *Journal of Educational Computing Research*, *40*(2), 183–210.

Hung, J. L., & Zhang, K. (2008). Revealing online learning behaviours and activity patterns and making predictions with data mining techniques in online teaching. *MERLOT Journal of Online Learning and Teaching*, *4*(4), 426–437.

Jonassen, D. H. (1996). *Computers in the classroom: Mindtools for critical thinking*. Prentice Hall, Inc.

Jonassen, D. H. (1997). Instructional design models for well-structured and III-structured problem-solving learning outcomes. *Educational Technology Research and Development*, *45*(1), 65–94.

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

Jonassen, D. H. (2000b). Revisiting activity theory as a framework for designing student-centered learning environments. *Theoretical Foundations of Learning Environments*, *89*, 121.

Jonassen, D. H. (2003). Using cognitive tools to represent problems. *Journal of Research on Technology in Education*, *35*(3), 362–381.

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process. *Learning and Instruction*, *72*(7), 4.

Khare, K., Lam, H., & Khare, A. (2018). Educational data mining (EDM): Researching impact on online business education. In *On the line* (pp. 37–53). Springer.

Kinnebrew, J. S., Loretz, K. M., & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behaviour patterns. *Journal of Educational Data Mining*, *5*(1), 190–219.

Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, *22*(2), 121–132.

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. *International Journal of Technology Enhanced Learning*, *5*(2), 133–150.

Krkovic, K., Mustafic, M., Wüstenberg, S., & Greiff, S. (2018). Shifts in the assessment of problem solving. In *Assessment and teaching of 21st century skills* (pp. 55–73). Springer.

Lajoie, S. P. (1993). Computer environments as cognitive tools for enhancing learning. In *Computers as Cognitive Tools* (1st ed., pp. 261–288).

Lawless, K. A., & Kulikowich, J. M. (1996). Understanding hypertext navigation through cluster analysis. *Journal of Educational Computing Research*, *14*(4), 385–399.

Levy, F. (2010). How technology changes demands for human skills.

Levy, R. (2013). Psychometric and evidentiary advances, opportunities, and challenges for simulation-based assessment. *Educational Assessment*, *18*(3), 182–207. https://doi.org/10.1080/10627197.2013.814517

Li, L. Y., & Tsai, C. C. (2017). Accessing online learning material: Quantitative behaviour patterns and their effects on motivation and learning performance. *Computers & Education*, *114*, 286–297.

Liu, M., & Bera, S. (2005). An analysis of cognitive tool use patterns in a hypermedia learning environment. *Educational Technology Research and Development*, *53*(1), 5–21.

Liu, M., Williams, D., & Pedersen, S. (2002). Alien rescue: A problem-based hypermedia learning environment for middle school science. *Journal of Educational Technology Systems*, *30*(3), 255–270.

Mahlow, N., Hahnel, C., Kroehne, U., Artelt, C., Goldhammer, F., & Schoor, C. (2020). More than (single) text comprehension?–on university students' understanding of multiple documents. *Frontiers in Psychology*, *11*, 562450.

Marasigan, N. V. (2018). Examining calculator use in the mathematics classroom. *International Journal of Recent Innovations in Academic Research*, *2*(8), 165–170.

Marriott, N. (2004). Using computerized business simulations and spreadsheet models in accounting education: A case study. *Accounting Education*, *13*(sup1), 55–70.

Marutho, D., Handaka, S. H., & Wijaya, E. (2018, September). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 international seminar on application for technology of information and communication, IEEE (pp. 533–538).

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 287–303). Erlbaum.

Mayr, E., Smuc, M., & Risku, H. (2011). Many roads lead to Rome: Mapping users' problem-solving strategies. *Information Visualization*, *10*(3), 232–247.

Mislevy, R. J., Haertel, G., Riconscente, M., Rutstein, D. W., & Ziker, C. (2017). Evidence-centered assessment design. In *Assessing model-based reasoning using evidence-centered design* (pp. 19–24). Springer.

Molkenthin, R., Breuer, K., & Tennyson, R. D. (2008). Real-time diagnostics of problem-solving behaviour for business simulations. *Assessment of Problem Solving Using Simulations* (1st ed., pp. 201–221).

Moos, D. C. (2009). Note-taking while learning hypermedia: Cognitive and motivational considerations. *Computers in Human Behaviour*, *25*(5), 1120–1128.

Nesbit, J. C., & Winne, P. H. (2008). Tools for learning in an information society. In T. Willoughby & E. Wood (Eds.), *Children's learning in a digital world* (pp. 173–195). Blackwell Publishing.

Nesbit, J. C., Zhou, M., Xu, Y., & Winne, P. (2007, August). Advancing log analysis of student interactions with cognitive tools. In 12th biennial conference of the european association for research on learning and instruction (EARLI) (pp. 2-20).

Newell, A. (1979). Reasoning, problem solving and decision processes: The problem space as a fundamental category.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Prentice-hall.

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. Academic Press.

OECD. (2014). PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (volume V), PI SA, OECD Publishing. https://doi.org/10.1787/9789264208070-en.

PIAAC Expert Group in Problem Solving in Technology- Rich Environments. (2009). "PIAAC problem solving in technology-rich environments: A conceptual framework", OECD education working papers, No. 36, OECD Publishing. https://doi.org/10.1787/220262483674.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rausch, A., & Kögler, K. (2016). Authenticity and efficiency in assessing domain-specific problem-solving competence: Conflicting goals in large-scale assessments. In *International Conference on Competence Theory, Research and Practice in Wageningen*, Oct (Vol. 19, No. 21, p. 2016).

Rausch, A., Kögler, K., Frötschl, C., Bergrab, M., & Brandt, S. (2017). Problemlöseprozesse sichtbar machen: Analyse von Logdaten aus einer computerbasierten Bürosimulation. [Uncovering problem-sovling processes: Analysis of log data from a computer-based office simulation]. *Zeitschrift für Berufs-und Wirtschaftspädagogik*, *113*(4), 569–594.

Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Reseach in Vocational Education and Training*, *8*, 1–23.

Reich, J. (2015). Rebooting MOOC research. *Science*, *347*(6217), 34–35.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618.

Sabourin, J., Shores, L. R., Mott, B. W., & Lester, J. C. (2012). Predicting student self-regulation strategies in game-based learning environ-ments. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al. (Eds.), (Hrsg.) *Intelligent tutoring systems (lecture notes in computer science, bd. 7315, S.)* (pp. 141–150). Springer Nature.

Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education*, *7*(1), 1719571.

Sembill, D., Rausch, A., Wuttke, E., Seifried, J., Wolf, K. D., Martens, T., & Brandt, S. (2016). Modellierung und messung domänenspezifischer problemlösekompetenz bei industriekaufleuten (DomPL-IK). Datensatz beim FDZ (Forschungsdatenzentrum) am IQB (Institut zur Qualitätsentwicklung im Bildungswesen), Berlin. https://doi.org/10.5159/IQB_DomPL-IK_v1

Shin, D., & Shim, J. (2020). A systematic review on data mining for mathematics and science education. *International Journal of Science and Mathematics Education*, *19*(4), 1–21.

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in games*. MIT Press. https://doi.org/10.7551/mit-press/9589.001.0001

Trafton, J. G., & Trickett, S. B. (2001). Note-taking for self-explanation and problem solving. *Human-Computer Interaction*, *16*(1), 1–38.

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills.

Tóth, K., Rölke, H., Greiff, S., & Wüstenberg, S. (2014). Discovering students' complex problem solving strategies in educational assessment. In educational data mining 2014.

Van Merrienboer, J. J. (2013). Perspectives on problem-solving and instruction. *Computers & Education*, *64*, 153–160. https://doi.org/10.1016/j.compedu.2012.11.025

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Lawrence Erlbauw.

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology*, *61*, 653–678.

Wirth, J., & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy & Practice*, *10*(3), 329–345. https://doi.org/10.1080/0969594032000148172

World Economic Forum. (2020). Future of jobs accessed on March 18th, 2022, https://www3.weforum.org/docs/WEF_Future_of_Jobs_2020.pdf.

Xing, W., Wadholm, B., & Goggins, S. (2014, March). Learning analytics in CSCL with a focus on assessment: An exploratory study of activity theory-informed cluster analysis. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 59-67).

Zydney, J. M. (2010). The effect of multiple scaffolding tools on students' understanding, consideration of different perspectives, and misconceptions of a complex problem. *Computers & Education*, *54*(2), 360–370 Found out that tools with similar functions could influence the behaviour negatively.

How to cite this article: 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. https://doi.org/10.1111/jcal.12770

APPENDIX A

A.1 | APPENDIX

Table A1(a) and A1(b).

TABLE A1 Exemplary coding schemes and scoring rubrics for cognitive phase A2 (scenario 1) (translated from Sembill et al., 2016)

Cognitive phase 2: Information processin
A2S1I1: Domain-specific processing algorithms

Variable	Description	Value label
A2S1I1a (semi-auto; manual)	Calculation of target costs	In cells I26 to I28 9 = no entries have been made (it still says: "to be calculated") 0 = only incorrect values have been entered 1 = at least one correct value, but not all of them correct 2 = three correct values (I26 = 485.90 AND 2 = three correct values) (I26 = 485.90 AND I27 = 12356.05 AND I28 = 3184.15)
A2S1I1b (partial-auto; manual)	Calculation of absolute deviations	Mainly in cells J7 to J25 99 = no entries were made 00 = only wrong values entered 10 = Differences between actual costs and planned costs (wrong!) calculated (e.g. J7 = 1467,46; J8 = −185.92; J9 = −174.91; J10 = −347.00; J11 = −1455.11), but not necessarily in every cell (some cells empty in the first five or less than five possible) 11 = differences between actual costs and planned costs (wrong!) calculated with negative sign (e. g. B. J7 = −1467,46; J8 = 185,92; J9 = 174,91; J10 = 347.00; J11 = 1455.11), but not necessarily in every cell (some cells empty in the first five or less than five possible) 21 = Differences between actual costs and target costs calculated (correct; see values under code "30"), but not necessarily in every cell (some cells empty empty or also single wrong values in the first five or less than five) 22 = Differences between target costs and actual costs are calculated, so that in comparison to the values listed under code "30" only the sign is wrong e.g. J is wrong (e.g. J7 = −13704.10; J8 = −47.30; J9 = −2.15; J10 = −43.00; J11 = 763.25). 30 = Differences between actual costs and target costs correctly determined throughout (e.g. J7 = 13704.10; J8 = 47.30; J9 = 2.15; J10 = 43.00; J11 = −763,25).
A2S1I1c (partial-auto; manual)	Calculation of relative deviations	Mainly in cells K7 to K25 99 = no entries were made 01 = only implausible values were entered 02 = incorrect calculation with target/actual*100 (e.g. K7 = 83.12; K8 = 96.41; K9 = 99.78; K10 = 98.04; K11 = 125,02) 11 = correct relative deviations on the basis of the incorrectly calculated column "J" (consequential error). 12 = Calculation of a change factor without correction as a quotient [" = actual/target"] (e.g. K7 = 1.20; K8 = 1.04; K9 = 1.00; K10 = 1.02; K11 = 0,80) 13 = Calculation of a change factor without correction in percent [" = actual/target*100"] (e.g. K7 = 120.31; K8 = 103.67; K9 = 100.22; K10 = 102,00; K11 = 79,98) 21 = Relative deviations correctly determined as a simple as a simple quotient (e.g. K7 = 0.203; K8 = 0.0368; K9 = 0.0022; K10 = 0.020; K11 = −0,200) 22 = relative deviations correctly determined throughout in percent (e.g. K7 = 20.31; K8 = 3.68; K9 = 0.22; K10 = 0.0022; K11 = −0.200).

TABLE A2 Exemplary coding schemes and scoring rubrics for cognitive phase A3 (scenario 1) (translated from Sembill et al., 2016)

Cognitive phase 3: *Making a reasoned decision*
A3S1I1: Quality of rationale and recommended consequences

Variable	Description	Value label
A3S1I1a (part-auto)	Mention of causes for the deviations	Mention of e.g. the following four causes: (a) Double painting at Lack & Rahmen, (b) Anniversary discount at Mohnhaupt, (c) rush orders 2011 model/model of seat post at Schr. 99 = no reply mail available or obviously broken off due to the system (after or in the first sentence) 01 = no causes are mentioned at all 02 = wrong causes are mentioned 10 = one plausible cause is mentioned 20 = two plausible causes are mentioned 30 = three plausible causes are mentioned 40 = four plausible causes are mentioned
A3S1I1b (manual)	Mention of possible consequences due to the causes of deviation	Name e.g. the following four consequences: to (a) consider higher costs, if necessary look for another supplier (accepting poorer quality would be quality would be wrong); to (b) lower costs were only one-off, therefore not to be taken into account for the future; to (c) Optimize orders or increase stock; (d) keep old model if possible or consider higher costs for new model 99 = no reply mail available or obviously aborted due to the system (after or in the first sentence) 01 = No consequences are mentioned at all. 02 = The consequences mentioned are not plausible. 10 = one plausible consequence is mentioned 20 = two plausible consequences are mentioned 30 = three plausible consequences are mentioned 40 = four plausible consequences are mentioned

A.2 | APPENDIX

Figure A2.

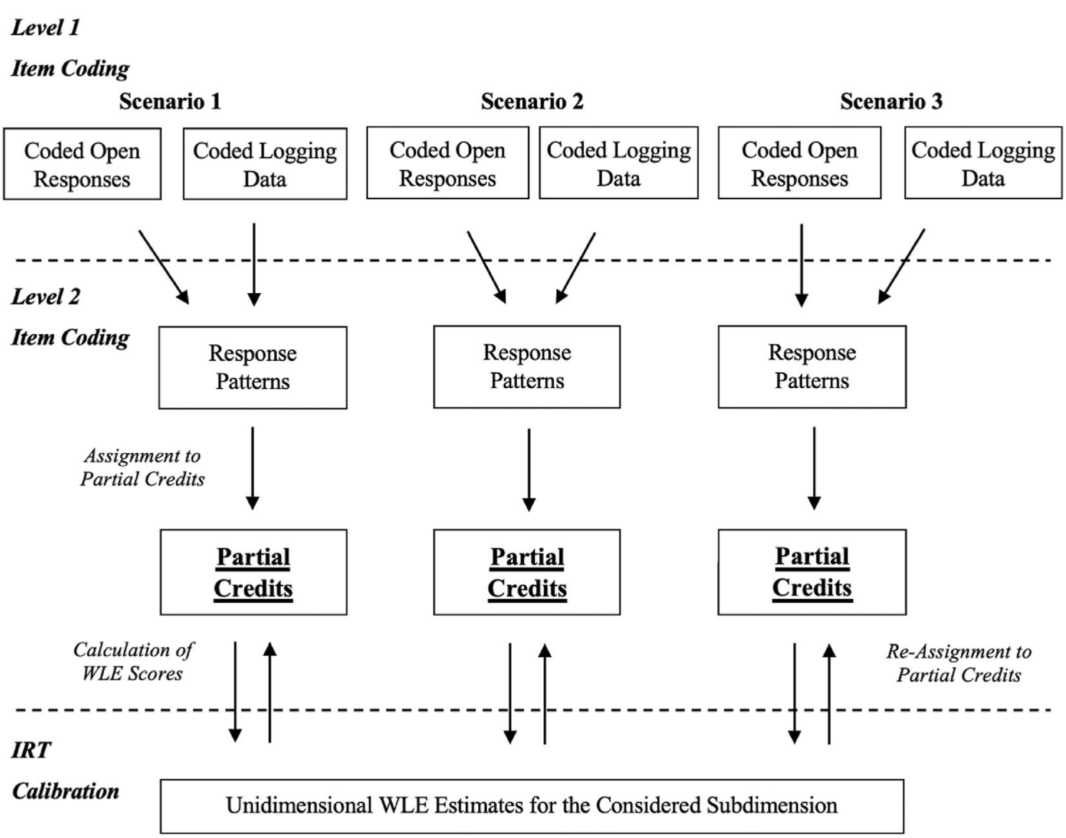


FIGURE A2 Visualization of the three-step procedure from problem-solving performance to problem-solving competence (adapted from Seifried et al., 2020)

A.3 | APPENDIX

Table A3.

TABLE A3 Descriptive statistics of problem-solving performance

Cognitive phase of domain-specific problem solving	Maximum attainable score (in each scenario)	Scenario	Mean	Standard deviation	Min	Max
A1. Identification of needs for action and sources of information	5	1	1.69	1.09	0.0	4.00
		2	2.09	1.34	0.0	4.00
		3	1.19	1.45	0.0	4.00
A2. Process of information	6	1	2.80	1.15	0.0	6.00
		2	1.30	1.20	0.0	4.00
		3	0.59	0.86	0.0	5.00
A3. Making well-founded decisions	7	1	0.80	1.21	0.0	5.00
		2	1.65	1.48	0.0	4.00
		3	1.44	1.51	0.0	5.00
A4. Communication of decisions in an adequate way	3	1	1.90	0.99	0.0	3.00
		2	2.11	1.03	0.0	3.00
		3	1.90	1.04	0.0	3.00

APPENDIX B

Figure B.

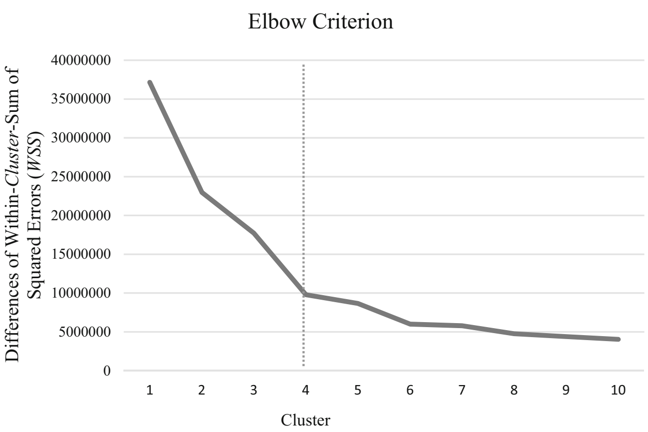


FIGURE B Elbow criterion with a four-cluster-solution. *Note*: The four-cluster solution demonstrated a fair and reasonable compromise of an interpretable number of clusters, i.e., as few clusters as possible, but with fewer clusters, information is lost.

APPENDIX C

Table C.

TABLE C Students' dropout over time (*N* adjusted).

	Scenario 1			Scenario 2			Scenario 3		
	0.- 10. Min	10.- 20. Min	20.- 30. Min	0.- 10. Min	10.- 20. Min	20.- 30. Min	0.- 10. Min	10.- 20. Min	20.- 30. Min
Note takers (Cluster 1)	115	112	92	115	114	91	115	104	70
Spreadsheet users (Cluster 2)	152	145	111	152	146	116	152	136	77
Calculator users (Cluster 3)	84	72	50	84	73	41	84	64	26
Email assignment readers (Cluster 4)	81	80	66	81	81	58	81	70	31

APPENDIX D

Lastly, the clusters share commonality with respect to viewing and working with relevant emails. In all scenarios, emails are read at the beginning (stage 1) and written especially at the end of the scenario across all clusters.

In sum, the third question regarding whether the clusters differ in terms of their tool use over time can be answered. On the one hand, clusters differ in their behaviour over time depending on the scenarios, but also show similar behaviour patterns, for example, calculation repertoire or emails (Figure D).

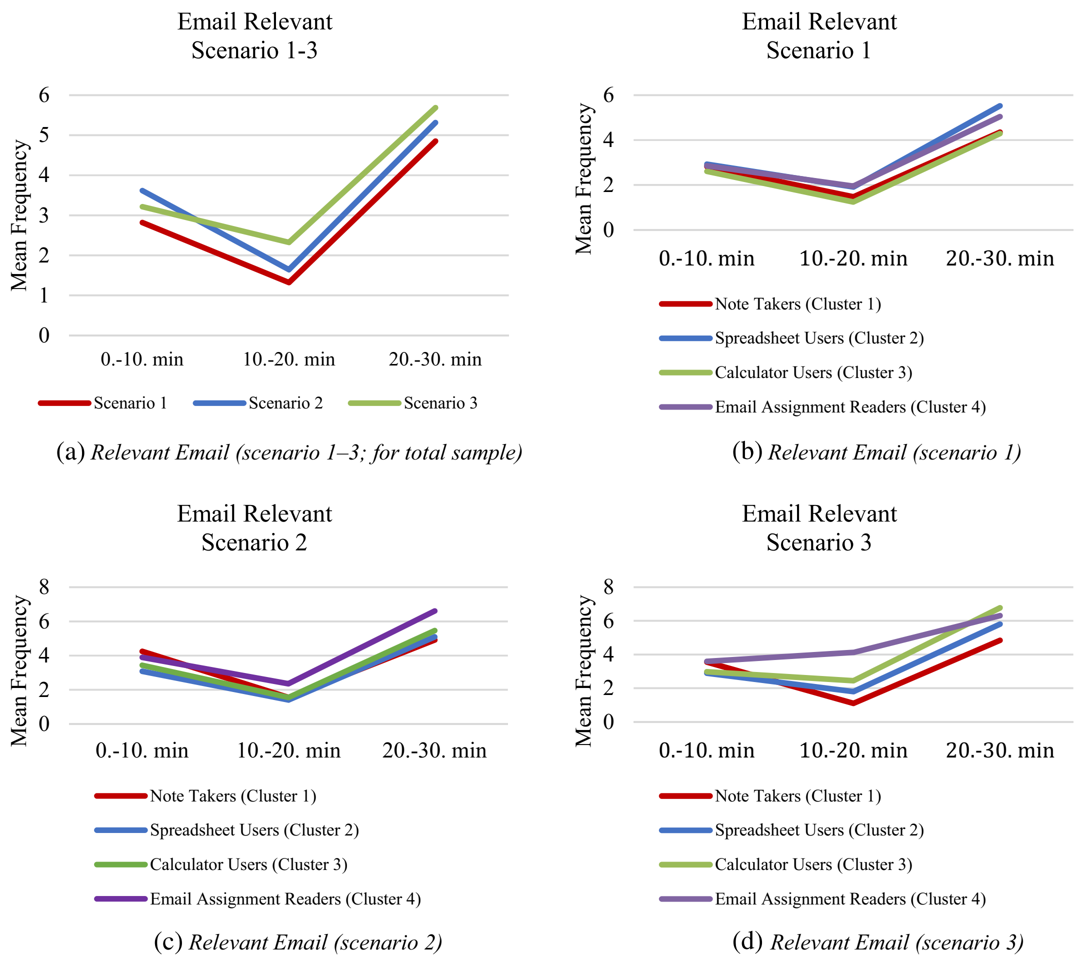
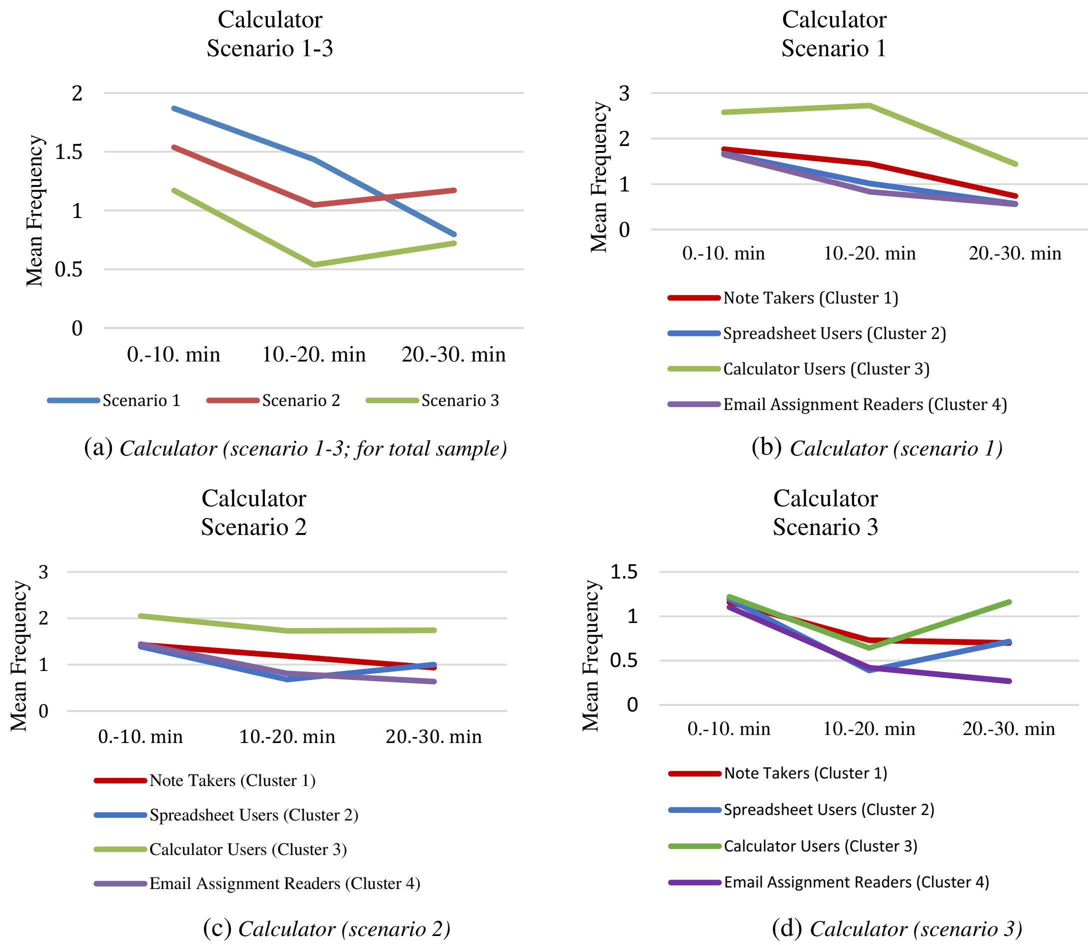


FIGURE D Relevant email use distributed over time (mandatory tool). (a) relevant email (scenario 1–3; for total sample), (b) relevant email (scenario 1), (c) relevant email (scenario 2), (d) relevant email (scenario 3)

APPENDIX E

All clusters showed a high usage of the calculator at the beginning across all scenarios. However, the usage decreases from the first to the second time interval. In scenarios 2 and 3, the calculator use increases again afterward. One reason for the behaviour could be the increased tendency to use the other calculator tool, the spreadsheet program, in the second time interval (Figure E).



(a) *Calculator (scenario 1-3; for total sample)* (b) *Calculator (scenario 1)* (c) *Calculator (scenario 2)* (d) *Calculator (scenario 3)*

FIGURE E Calculator use over time (voluntary tool). (a) calculator (scenario 1–3; for total sample), (b) calculator (scenario 1), (c) calculator (scenario 2) and (d) calculator (scenario 3)