Computers & Education 128 (2019) 1–12

Contents lists available at ScienceDirect

Computers & Education

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

The role of planning in complex problem solving

Beate Eichmann[a,*], Frank Goldhammer[a], Samuel Greiff[b], Liene Pucite[c], Johannes Naumann[d]

[a] German Institute for International Educational Research, Centre for International Student Assessment, Schloßstraße 29, 60486, Frankfurt am Main, Germany
[b] University of Luxembourg, 2, avenue de l'Université, L-4365, Esch-sur-Alzette, Luxembourg
[c] Goethe University Frankfurt, Theodor-W.-Adorno-Platz 1, 60323, Frankfurt am Main, Germany
[d] University of Wuppertal, Gauss Straße 20, 42119, Wuppertal, Germany

A R T I C L E I N F O

Keywords:
Complex problem solving
Planning
Computer-based assessment
Log data
PISA

A B S T R A C T

Complex problem solving (CPS) is a highly transversal competence needed in educational and vocational settings as well as everyday life. The assessment of CPS is often computer-based, and therefore provides data regarding not only the outcome but also the process of CPS. However, research addressing this issue is scarce. In this article we investigated planning activities in the process of complex problem solving. We operationalized planning through three behavioral measures indicating the duration of the longest planning interval, the delay of the longest planning interval and the variance of intervals between each two successive interactions. We found a significant negative average effect for our delay indicator, indicating that early planning in CPS is more beneficial. However, we also found effects depending on task and interaction effects for all three indicators, suggesting that the effects of different planning behaviors on CPS are highly intertwined.

1. Introduction

1.1. Theoretical background

The ability to solve complex problems is an essential competence in both education and everyday life, and is required for active participation in today's society. Ongoing globalization and digitalization confront people with an increasingly complex environment that demands numerous problems to be solved in personal life as well as at the workplace (Fischer, Greiff, & Funke, 2012). Furthermore, problem solving is the basis of many scholastic learning processes and is therefore regarded as a fundamental goal of education (OECD, 2013). From this perspective, the question stands which kind of behavioral engagement with a problem solving task predicts successful task completion. This study investigates the relationship between planning behavior and performance in complex problem solving (CPS) using computer-generated log data from a large-scale assessment.

According to Mayer and Wittrock (2006) a problem occurs when someone wants to achieve a goal and no obvious solution is available. This broad definition refers to many academic and real-world tasks in the field of problem solving. In this paper, we focus on CPS, which is frequently needed in real life. In contrast to non-complex problems, in CPS the given state, the goal state and the barriers between these states are complex, i.e. they can change dynamically and are opaque (not all information is presented at the

* Corresponding author.
E-mail addresses: beate.eichmann@dipf.de (B. Eichmann), goldhammer@dipf.de (F. Goldhammer), samuel.greiff@uni.lu (S. Greiff), pucite@em.uni-frankfurt.de (L. Pucite), j.naumann@uni-wuppertal.de (J. Naumann).

https://doi.org/10.1016/j.compedu.2018.08.004
Received 14 December 2017; Received in revised form 31 July 2018; Accepted 4 August 2018
Available online 06 August 2018

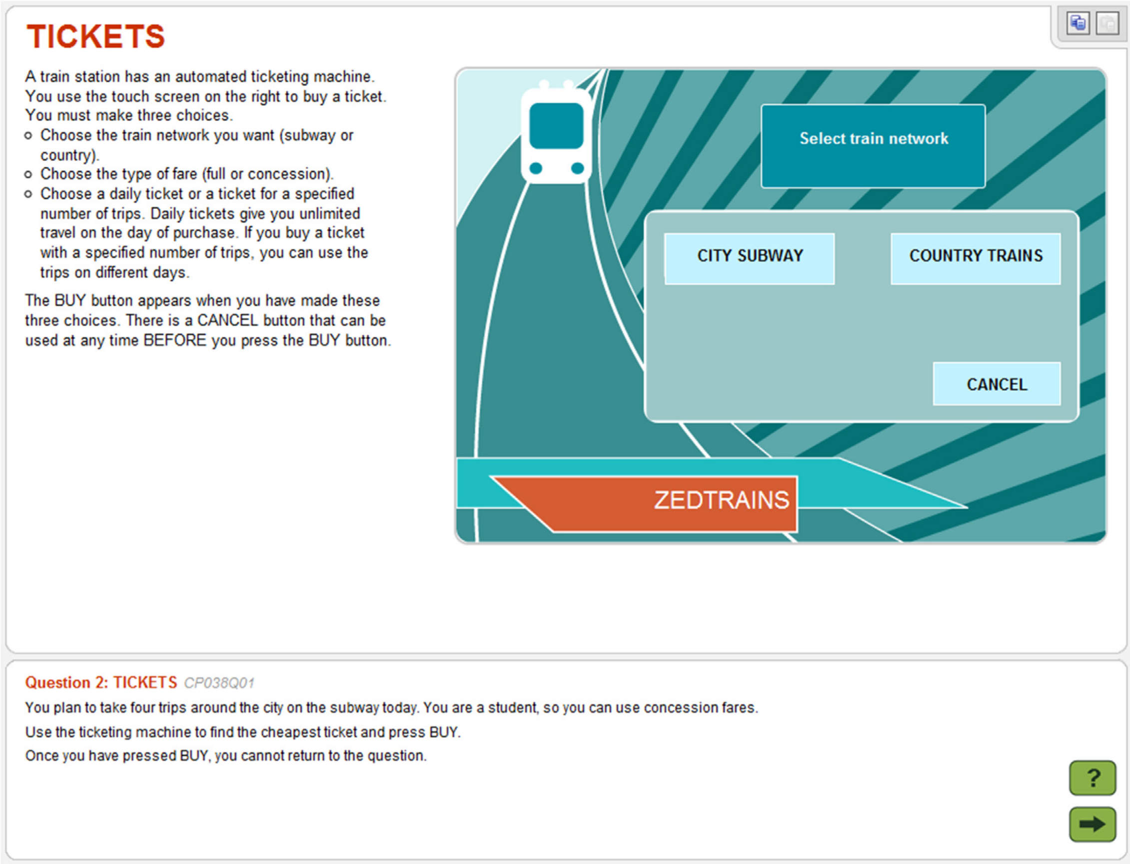

Fig. 1. CPS-task from PISA 2012.

outset) (Frensch & Funke, 1995). Hence, the problem solver has to interact with the problem to overcome these barriers and solve the problem. In the scientific literature many similar definitions of CPS exist, which all stress the importance of dynamics, opacity, and interactivity. To account for the interactivity of complex problems, CPS is mostly assessed via computer-based tasks, as it was the case in the Programme for International Student Assessment (PISA) 2012 (OECD, 2013). PISA measures curricular and cross-curricular competencies of fifteen-year-olds in the participating countries every three years. An example of a problem solving task, which is taken from the PISA 2012 problem solving assessment, is shown in Fig. 1. It displays a simulated ticket machine and an instruction to buy a specific ticket. This scenario can be regarded as a problem: there is a given state (not having a ticket), a goal state (having a ticket) and barriers between them (the user interface of the ticket machine). The problem is complex, in that not all information is present at the outset (e.g. it is not shown how different fare types are selected) and it therefore requires a number of interactions to gain the information needed to solve the task. Because CPS is considered a cross-curricular domain, its assessment can take many different forms and the tasks used in PISA 2012 are rather heterogeneous in nature (Buchner, 1995; Leutner, Funke, Klieme, & Wirth, 2005).

Several cognitive processes are involved in the solving of complex problems. The OECD (2013) established a framework of problem solving that is based upon the work of several cognitive psychologists (Baxter & Glaser, 1997; Blech & Funke, 2005, 2010; Bransford, Brown, & Cockling, 1999; Funke & Frensch, 2007; Greiff & Funke, 2008; Klieme, 2004; Mayer & Wittrock, 2006; Osman, 2010; Reeff, Zabal, & Blech. C., 2006; Vosniadou & Ortony, 1989; Wirth & Klieme, 2003). According to this framework, four main processes are involved in problem solving: “exploring and understanding”, “representing and formulating”, “planning and executing”, and “monitoring and reflecting”. The process of “exploring and understanding” involves interaction with the problem environment to gain information. Furthermore, the given and the found information should be understood by creating mental models of the pieces of information. The process of “representing and formulating” covers the creation of a mental representation of the problem situation as a whole. This includes the selection and integration of information, and the formulation of hypotheses about the problem. “Planning and executing” involve the setting of goals and sub-goals and the selection and execution of steps to achieve these goals. The last process, “monitoring and reflecting”, requires monitoring the progress towards the goals and, if necessary, for the individual to adjust their behavior as well as reflecting on assumptions and solutions. However, not all of the processes have to be present in each particular problem solving process. Furthermore, according to Lesh and Zawojewski (2007), the human brain is capable of parallel information processing, so it is also assumed that in CPS several processes can happen simultaneously. Unterrainer and Owen (2006) argue that even in non-complex problems people tend to switch between the different processes of problem solving.

Notably, planning occurs in three out of the four processes in the PISA framework (OECD, 2013): “exploring and understanding”, “representing and formulating” and “planning and executing” all refer to planning as it is defined by Unterrainer and Owen (2006). They argue that planning requires an individual to form a representation of the current state and goal state of a problem, and also to conceive of a sequence of actions, which transform the current state into the goal state. In the PISA framework, understanding the problem and creating a representation of the problem could therefore be understood as forming a descriptive representation of the possible states of the problem and actions to switch between them, while what is explicitly called “planning” in the framework should target the process of selecting the specific actions required to reach the desired goal (cf. Novick & Bassok, 2005). Formulating hypotheses about the problem refers to both aforementioned parts of planning, since the problem solver is required to formulate hypotheses about both the structure of the problem and steps to solve it. This shows that planning is an important part of CPS, which serves several purposes. In this article we follow this understanding of planning and investigate empirically the relevance of planning in CPS. Although planning is part of many theories about CPS (as in the PISA framework), empirical studies addressing this process and its relevance for success in CPS are scarce. In the following, we summarize relevant results from this domain.

1.2. Literature review

Like any cognitive process, planning cannot be observed directly. Thus, the occurrence of planning has to be inferred from behavior. To the advantage of the researcher, CPS is an interactive process, which makes it empirically accessible for detailed investigation (Greiff, Niepel, Scherer, & Martin, 2016). Although several studies investigated behavior while solving complex and non-complex problems, there are no studies addressing planning in CPS. Therefore, we examine both research about planning in related domains and studies addressing behavior in CPS that is related to planning: Albert and Steinberg (2011) found that first move latency is positively related to performance in non-complex problem solving tasks. They argue that individuals with higher first move latency complete their initial planning phase and therefore show a higher performance rate, in comparison to people with lower first move latency. Furthermore, when instructed to plan ahead the entire process of solving the task before starting to interact with it, the performance in problem solving was improved (Unterrainer & Owen, 2006). Both of these studies demonstrate the importance of prior planning in problem solving. However, both studies also used the “Tower of London” task for their investigations, which is a non-complex problem solving task according to the definition by Frensch and Funke (1995), so it remains unclear whether these findings hold for complex problems. As already mentioned, in complex problems the situation can change dynamically and not all of the information required to find a solution is presented at the outset, so in this case it is not possible to plan ahead completely before having interacted with the problem. Along this line of thinking, Greiff et al. (2016) found a negative correlation between performance in CPS and the frequency of interactions with the task, indicating that fewer and therefore presumably better planned actions also lead to higher performance in complex problems. Although this result indicates that it is beneficial to invest time in cognitive activities such as planning, instead of permanently interacting with a complex problem task, it does not show how this time should be distributed over the process of CPS. Another process measure, which might be related to time spent on planning, is the overall time spent on a task. Goldhammer et al. (2014) argue that, according to dual processing theory, the interpretation of time spent on a task depends on the nature of the task and the cognitive processes required to solve the task. In tasks that require non-routine behavior (like CPS tasks), time spent on a task can be regarded as an indicator of effort, so it can be assumed that in these scenarios students actually spend their time working on the task. In contrast, in tasks which only require routine behavior, time spent on a task can be seen as an indicator of having trouble solving the task. Supporting this theory, Goldhammer et al. (2014) showed that in problem solving tasks a higher amount of time spent on a task was associated with higher performance. Furthermore, they found that the effect of time spent depends on task difficulty and being strong especially in hard tasks. These results also indicate that investing time in cognitive activities (such as planning) might be beneficial in CPS. Also in line with this notion, Naumann and Goldhammer (2017) found that the investment of time in digital reading, a domain which can also be conceived as solving information problems (Brand-Gruwel, Wopereis, & Walraven, 2009; Rouet & Le Bigot, 2007) was positively predictive of successful task completion especially in hard tasks. Moreover, these authors showed that positive effects of time were accelerated with task's navigation demands, thus with the problem space becoming larger, and more opaque.

However, the relationship between time spent on a task and performance in CPS does not seem to be purely linear. Greiff et al. (2016) found a quadric relation between the time spent on a task and performance, where times that were too long or too short were associated with poor performance. Likewise, Naumann and Goldhammer (2017) found a negatively sloped quadratic trend for the time on task relation to performance in digital reading. However, where tasks had high navigation demands, this quadratic trend was more linear, possibly indicating that in these tasks positive effects of cognitively demanding activities such as planning levelled off not that early. In other studies, task-dependent effects of process measures were also found (e.g. Goldhammer, Naumann, & Greiff, 2015). For example, Naumann, Goldhammer, Rölke, and Stelter (2014) found that in technology-based problem solving the effect of the number of performed interactions on performance was moderated by the minimum number of interactions the task required. Hence, in a broad domain such as CPS it can be assumed that the effects of process measures on performance depend on certain task characteristics such as difficulty or the number of required interactions.

1.3. The current study

Although planning is an important part of many theories about CPS (e.g. Dörner, 1989; Leutner et al., 2005; Mayer & Wittrock, 2006) there is no research about the impact of planning on performance in CPS (to the best of our knowledge). Furthermore, there is no research that we are aware of that addresses the effects of different aspects of planning. In addition to the recent findings that general indicators of planning have a positive impact on performance in CPS, we attempt to inspect several aspects of planning by using different process measures: we investigate whether the length of planning intervals, the time when planning takes place, and the variation of planning time over the whole CPS process influence the performance in CPS, and therefore include three different aspects of planning during CPS tasks. For each of those three aspects we developed a process measure that addresses the timespan between two successive interactions. Our definition of planning refers to understanding, representing, formulating, and planning from the PISA framework (OECD, 2013). While exploring and executing explicitly refer to visible action, this leaves the processes that represent planning, as well as monitoring and reflecting to happen in the time between. Therefore, process indicators based on the time between interactions should capture planning behavior while also correlating with monitoring and reflecting and task-irrelevant behavior.

We will further introduce the process indicators we used for planning in the following paragraphs. Because research in this area is scarce, we concentrated on the following explorative research questions:

Research question 1: Is the length of planning intervals related to performance in CPS?

According to several studies, planning activities have a positive impact on problem solving. For example, Albert and Steinberg (2011) showed in their study that first-move latency as a predictor of planning behavior is positively related to task outcome in problem solving. We extend this measure and not only take into account the time before the first move, but also include planning time that appears at any point in the problem solving process. Due to the opacity of complex problems and the need for exploration, we argue that planning might occur at a later point in time, after the collection of necessary information, and not just and exclusively at the beginning.

We use the duration of the longest interval, in which no interaction with the problem occurs, as an indicator of the length of planning intervals (duration indicator). In the example task (Fig. 1), this would be the longest time without a button being pressed. Obviously, this measure does not cover all the planning that happens during the process but it gives information about the longest instance when planning activities might take place. In doing so, we are able to cover main planning activities no matter if they appear at the beginning of the process or at a later stage. However, it is unclear whether planning or any other activities take place during this time. Nevertheless, the results of Goldhammer et al. (2014) indicate that in problem solving tasks it is beneficial to invest more time in the task, so we assume that time spent on a task is mostly spent by engaging in task-relevant cognitive activities such as planning.

We further investigate the relationship between the duration indicator and performance by testing whether the effect varies between tasks. The results of several studies show that the effects of process measures in problem solving are dependent on task characteristics, so this might also be true for planning indicators (e.g. Goldhammer et al., 2014; Goldhammer et al., 2015; Naumann & Goldhammer, 2017; Naumann et al., 2014). The tasks used in this study vary in many respects, such as in level of difficulty, complexity, or in the domain of knowledge, and cover a wide range of everyday problems. Therefore different tasks might require different styles of planning. For example, tasks that require more interactions to get to a solution might also require the individual to plan for longer.

Research question 2: Is the time when planning takes place related to performance in CPS?

The second aspect of planning behavior that we investigate is the time when planning takes place in CPS. As mentioned before, Albert and Steinberg (2011) showed that planning at the beginning of a non-complex problem solving process is beneficial, while there is no research addressing the issue of when planning in CPS is beneficial. In the PISA framework (OECD, 2013) it is stated that planning can happen at any time, but it leaves open when or to what extent planning is beneficial for the task outcome. In complex problems it might not be optimal to make a plan at the beginning of the process but rather after some exploration; or there might not be an optimal time for planning at all.

To investigate this question, we use the delay of the longest interval, in which no interactions take place, as a measure for the time when (most of) the planning might take place. We use the delay of this interval from the beginning of the process as our indicator for research question 2 (delay indicator). In the example task (Fig. 1) this would be the time until the longest pause (when no buttons are pressed) occurs in the process.

In a manner similar to what we assumed in research question 1, the effect of the delay indicator could also depend on task characteristics. Thus, we investigate when planning should optimally occur and whether the effect varies between tasks. The heterogeneity of CPS tasks leads to many different dimensions, in which the tasks may vary. For example, a task that has a very rich environment that can take many different states and requires many interactions to find the solution might also require more exploration before meaningful planning can take place. Conversely, in less rich environments planning could occur earlier.

Research question 3: Is the variation of planning time related to performance in CPS?

Another aspect of planning that might influence the performance of the individual in CPS is the variation of planning time. Variation of planning time over several phases of similar length means that the interactions are performed by the problem solver at an overall constant speed. There are neither especially long nor especially short intervals between the interactions, so we assume that such variation of planning time indicates similar efforts in planning before every interaction. If, in contrast, people systematically allocate the time they take for planning their actions when necessary, this should cause larger variance in planning time across the course of completing a task. Because exploration is a central feature in CPS, we assume that a higher variation in planning time is beneficial: both phases of high exploration and phases of planning are necessary to solve a task.

We use the variance in the length of the intervals between two successive interactions as a measure of variation of planning time (variance indicator). With this measure, we take into account the whole process of CPS. Therefore, it is an extension of the two indicators introduced in the previous research questions, since these indicators only consider the longest planning interval.

The predictive power of the variance indicator, once again, might depend on task characteristics: for example, in a task that requires many similar interactions (like pressing the same button several times) it might be beneficial to perform only a few longer planning intervals when planning a sequence of actions, while in a task that requires different interactions, it might be beneficial to plan in more homogenous intervals when planning one action at a time. For this reason we investigate whether the effect of the variance indicator varies across tasks.

Research question 4: Are there interaction effects between the three process indicators mentioned above?

Few studies consider several behavioral indicators in CPS at once. We investigate interaction effects between the three indicators mentioned above, thus endorsing a comprehensive view on different aspects of planning. As a result, we wish to better understand the relationship between planning behavior and performance, and investigate whether the expected effects of our three indicators depend on each other. For example, the expected effect of the duration indicator could depend on the delay indicator, since a late planning phase might not be as crucial as an early planning phase. On the other hand, the effect of the duration indicator could also depend on the variance indicator: the duration of the longest phase might only play a role if there is any variation between the phases of planning.

2. Material and methods

2.1. Sample

For the analyses reported here, data from the fifth cycle of the PISA study (PISA, 2012) were used. In the present study, only data from the CPS tasks of German students (N = 1350) were used. After clearing invalid data, which might have occurred due to technical issues during data collection, N = 1346 students were left in the sample, of which 48.7% were female.

2.2. Instruments

In PISA 2012, problem solving was assessed via 42 computer-based tasks, which were organized into 16 units of two or three tasks each. An example task is shown in Fig. 1. Tasks in the same unit shared common stimulus material with only minor deviations. Students completed one or two (out of four) different problem solving clusters, which consisted of four units (ten to eleven tasks) each. The order of tasks within units and the order of units within clusters was always the same. Students were not able to return to a former task after finishing it.

According to OECD (2013), only very basic ICT skills were required to work on the computer-based problem solving tasks. To cover a wide range of difficulty, the tasks assessing problem solving varied across the following characteristics: amount, representation, and disclosure of information, internal complexity, distance to the goal, degree of abstraction, familiarity of the context, and reasoning skills required. For a detailed explanation of these task characteristics see OECD (2013). Notably, the tasks required interactive behavior like clicking on virtual buttons and sliders, dragging and dropping, operating simulated machines, exploring simulated environments, and manipulating variables. The response formats used included simple and complex multiple-choice tasks that were answered by clicking radio buttons, tasks that required shapes to be selected and dragged into position, tasks in which selections had to be made from pull-down menus, tasks that required parts of diagrams to be drawn or highlighted, and text boxes (OECD, 2013).

2.3. Scoring

The correctness of students' answers was derived from the computer-generated log data. For overall CPS performance, the response coding from OECD (2015) was used: the responses were coded either as correct (1) or incorrect (0), where partially solved tasks were coded as incorrect. We did not use the original scoring rules because they included partial credit based on students' behavior during the task. However, we were only interested in the effect of students' behavior on task outcome. Out of the total of 42 tasks, only 28 are considered to be interactive and therefore meet the definition of complex problems (Frensch & Funke, 1995). Only these tasks were used for data analyses. Two more tasks had to be excluded because the response included free text input, which was not recorded in the log data, so the correctness of the response could not be inferred. Another two tasks were excluded due to the respective log files including undefined events (clicks on elements with ambiguous IDs). For another three tasks, there was no valid log data in the German sample due to invalid and missing data. Therefore, after selecting data from only German students and clearing invalid data, a total of 21 tasks were left, out of the 28 tasks that could be considered complex (for task examples see Fig. 1 or OECD, 2013).

In accordance with our research questions, the three process indicators mentioned above were extracted from the log data: for research question 1 we used the duration of the longest interval between two successive interactions (duration indicator); for research

Table 1
Descriptive statistics of the three mean-centered indicators.

Indicators	SD	min	max	Skewness
Duration indicator	0.79	−2.14	2.69	0.16
Delay indicator	1.21	−0.59	4.48	1.97
Variance indicator	0.08	−0.04	1.31	4.71

question 2 we used the delay of the longest interval between two successive interactions from the beginning of the process (delay indicator); and for research question 3 we used the variance of times between each two successive interactions (variance indicator). The indicators were extracted for each combination of person and task for which data was available depending on the booklet design.

2.4. Procedure

The computer-based assessment of problem solving was part of the PISA 2012 study, which covered the domains of mathematics, reading, science, problem solving, and financial literacy. The computer-based assessment included the domains of problem solving, mathematics and reading. Students were administered two computer-based task clusters of which none, one, or both were problem solving clusters. An example of a computer-based problem-solving task is shown in Fig. 1. The computer-based assessment took place after the completion of the paper-based PISA tasks. Students were given 20 min per cluster.

2.5. Data preparation

The three indicators were inspected for outliers. An outlier was defined as a data point three standard deviations above/below the mean of the corresponding task, as suggested by Goldhammer et al. (2015). Overall, the sample comprised 6.63% outliers, which were replaced by the value at three standard deviations above/below the respective average as suggested by Goldhammer et al. (2014). Because the distributions of the three indicators showed pronounced skewness (duration indicator: 1.97, delay indicator: 4.99, variance indicator: 7.95), we normalized the distributions, using the logarithms of the three indicators. Since the ranges of the delay indicator and the variance indicator included zero, +1 was added to enable the logarithmic transformation. After this, all three indicators were centered around their respective grand mean. Table 1 shows the descriptive statistics of the three indicators after the described data preparation.

2.6. Data analyses

We estimated generalized linear mixed models (GLMM) to investigate our research questions using the package lme4 in the R-environment (Bates, Mächler, Bolker, & Walker, 2015; R Core Team, 2016). The GLMM framework allows modeling the probability of a correct response as a function of fixed and random effects (Baayen, Davidson, & Bates, 2008). In this way, we could model the overall (fixed) effects and the task-dependent (random) effects together, and at the same time, take into account the multilevel structure of our data (students nested in schools). For research questions 1 to 3, we estimated two models for each of our three indicators respectively: one model only including the overall (fixed) effect and one model containing both the overall and the task-dependent random effects. For the duration indicator, both models are shown in equation (1) and equation (2):

$$\eta_{pi} = \beta_0 + b_{0i} + b_{0p} + b_{0s} + \beta_1 M_{pi} \tag{1}$$

$$\eta_{pi} = \beta_0 + b_{0i} + b_{0p} + b_{0s} + (\beta_1 + b_{1i}) M_{pi} \tag{2}$$

In this model η_{pi} denotes the logit of the probability of a successful solution for person p completing task i, β the fixed effects and b the random effects. M represents the duration indicator. In the model, persons are nested within schools s.

The models for the other two indicators look the same, with the duration indicator being replaced by the delay, and variance indicator, respectively. Subsequently, we compared the two models for each indicator to test the task-dependent effects for significance. We used the likelihood ratio test and the Akaike Information Criterion (AIC) as criteria for model fit (Baayen et al., 2008; De Boeck et al., 2011). From the respective models, that fitted the data best, we took the estimates for the overall effects of our three process indicators as well as their p-values. For research question 4, we did not inspect random effects of interactions between indicators, since these effects could hardly be interpreted. Therefore, we estimated a model including only overall (fixed) effects of our three indicators and interaction effects between them. This model is shown in equation (3):

$$\eta_{pi} = \beta_0 + b_{0i} + b_{0p} + b_{0s} + \beta_1 M_{pi} + \beta_2 D_{pi} + \beta_3 V_{pi} + \beta_4 M_{pi} D_{pi} + \beta_5 M_{pi} V_{pi} + \beta_6 D_{pi} V_{pi} + \beta_7 M_{pi} D_{pi} V_{pi} \tag{3}$$

In this model, M, D and V represent the duration, delay and variance indicator respectively.

3. Results

The model comparisons are shown in Table 2. In the table, each comparison starts with ModelX.1 representing models including

Table 2
Model comparison; K is the number of parameters in the respective model, ΔK is the difference in the number of parameters.

Indicator	Model	K	AIC	BIC	logLik	deviance	X^2	ΔK	p	
Duration	Model1.1	5	9733.5	9769.0	−4861.8	9723.5	46.61	2	< 0.001	***
	Model1.2	7	9690.9	9740.6	−4838.4	9676.9				
Delay	Model2.1	5	9727.2	9762.6	−4858.6	9717.2	16.20	2	< 0.001	***
	Model2.2	7	9715.0	9764.6	−4850.5	9701.0				
Variance	Model3.1	5	9734.8	9770.3	−4862.4	9724.8	34.26	2	< 0.001	***
	Model3.2	7	9704.5	9754.2	−4845.3	9690.5				

Note. *p < 0.05, **p < 0.01, ***p < 0.001.

only overall (fixed) effects and proceeds with ModelX.2 representing models including both overall (fixed) and task-dependent random effects. The model comparisons for research questions 1–3 show that the models including both overall (fixed) and task-dependent random effects fit the data better than models including only overall effects, as shown in Table 2. In all model comparisons the more complex model has the better fit according to the likelihood ratio test and AIC. Therefore we used the estimates of those models to investigate the overall and task-dependent effects of our three indicators.

The model parameters of Model1.2, Model2.2 and Model3.2 are shown in Table 3. For the task-dependent random effects, the variance and standard deviation as well as the correlation between the random effect of the indicator and the task intercept are presented. For the fixed effects, estimates, standard errors, z-values, and p-values are shown.

The task-dependent random effects are depicted in Fig. 2. The effects of the three indicators for every single task are shown as a function of task easiness.

Research question 1: For the duration of the longest interval no significant fixed effect on the probability of success was found. However, a significant random effect was found for this indicator. The correlation between task easiness and the random effect of the indicator was $r = -0.04$. The random effect shows that the duration of the longest interval varies between tasks but is not related to task difficulty.

Research question 2: For the delay indicator, a significant fixed effect and a significant random effect was found. The fixed effect of -0.13 indicates that in general it is beneficial to plan early during the process. The correlation between task easiness and the random effect of the indicator was $r = -0.51$. The random effect shows that the effect of the delay indicator was near zero for difficult tasks and negative for easy tasks. This indicates that in easy tasks students who had their longest interval early during the process showed higher performance than students who had their longest interval later. For students working on difficult tasks this effect was less strong.

Table 3
Model parameters of Model1.2, Model2.2, and Model 3.2

Model1.2	Random effects:	Name	Variance	r		
Duration	Students:school	(Intercept)	0.47			
	School	(Intercept)	0.57			
	Task	(Intercept)	3.24			
		Duration	0.17	-.04		
	Fixed effects:	Estimate	SE	z-value	p-value	
	(Intercept)	0.22	0.40	0.54	0.589	
	Duration	−0.05	0.10	−0.47	0.639	
Model2.2	Random effects:	Name	Variance	r		
Delay	Students:school	(Intercept)	0.46			
	School	(Intercept)	0.56			
	Task	(Intercept)	3.35			
		Delay	0.02	-.51		
	Fixed effects:	Estimate	SE	z-value	p-value	
	(Intercept)	0.24	0.41	0.60	0.546	
	Delay	−0.13	0.05	−2.80	0.005	**
Model3.2	Random effects:	Name	Variance	r		
Variance	Students:school	(Intercept)	0.46			
	School	(Intercept)	0.56			
	Task	(Intercept)	3.42			
		Variance	16.78	-.39		
	Fixed effects:	Estimate	SE	z-value	p-value	
	(Intercept)	0.17	0.41	0.41	0.684	
	Variance	−1.85	1.31	−1.42	0.155	

Note. *p < 0.05, **p < 0.01, ***p < 0.001.

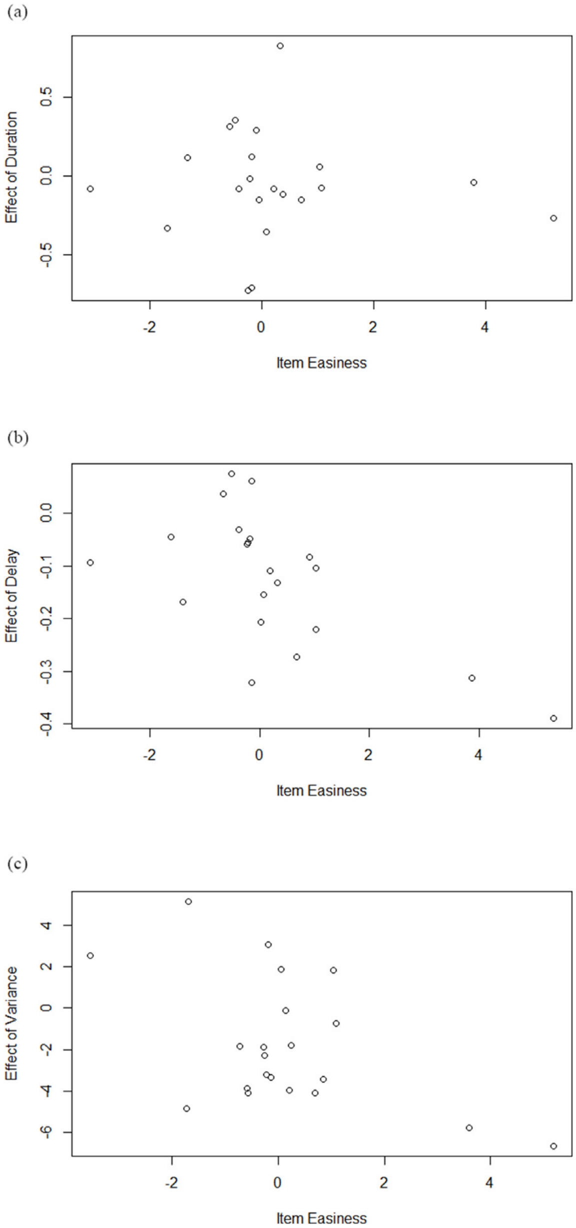

Fig. 2. Random effects of (a) the duration indicator, (b) the delay indicator, and (c) the variance indicator (calculated as fixed effect + task-specific effect) depending on task easiness (calculated as intercept + task-specific effect).

Research question 3: For the variance indicator no significant fixed effect but a significant random effect was found. The correlation between task easiness and the effect of the indicator was $r = -0.39$. The random effect indicates that in difficult tasks a high variation of time is beneficial while in easy tasks a low variation is beneficial.

Research question 4: The model including fixed effects of all three indicators and interactions between them is shown in Table 4. Significant effects were found for the interactions of duration-delay, delay-variance, and duration-delay-variance. Also the significant effect of the delay indicator was found again. The significant interaction effects are depicted in Fig. 3. The interaction between duration and delay shows that the duration indicator is positively related to success when the delay is high but has a very small negative effect when the delay is low (Fig. 3 (a)). The interaction between delay and variance shows that variance is positively related to success when the delay is low and negatively related to success when the delay is high (Fig. 3 (b)). The interaction effect of all three indicators shows that the effect of duration is not only depending on the delay but also on the combination of variance and delay (Fig. 3(c and d)). The positive effect of duration when the delay is high is largest when the variance is also high (Fig. 3 (c)). On the other hand, when the delay is low and the variance is high the effect of duration is negative (Fig. 3 (d)).

Table 4
Model parameters of Model 4.

Random effects:	Name	Variance	SD
Students:school	(Intercept)	0.46	0.68
School	(Intercept)	0.55	0.74
Task	(Intercept)	3.36	1.83

Fixed effects	Estimate	SE	*z*-value	*p*-value	
(Intercept)	0.14	0.41	0.34	0.735	
Duration	0.11	0.06	1.70	0.090	
Delay	−0.30	0.10	−3.12	0.002	**
Variance	−2.27	1.46	−1.56	0.120	
Duration*Delay	0.18	0.07	2.48	0.013	*
Duration*Variance	0.12	0.87	0.14	0.887	
Delay*Variance	−5.46	2.34	−2.33	0.020	*
Duration*Delay*Var.	2.73	1.30	2.09	0.036	*

Note. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

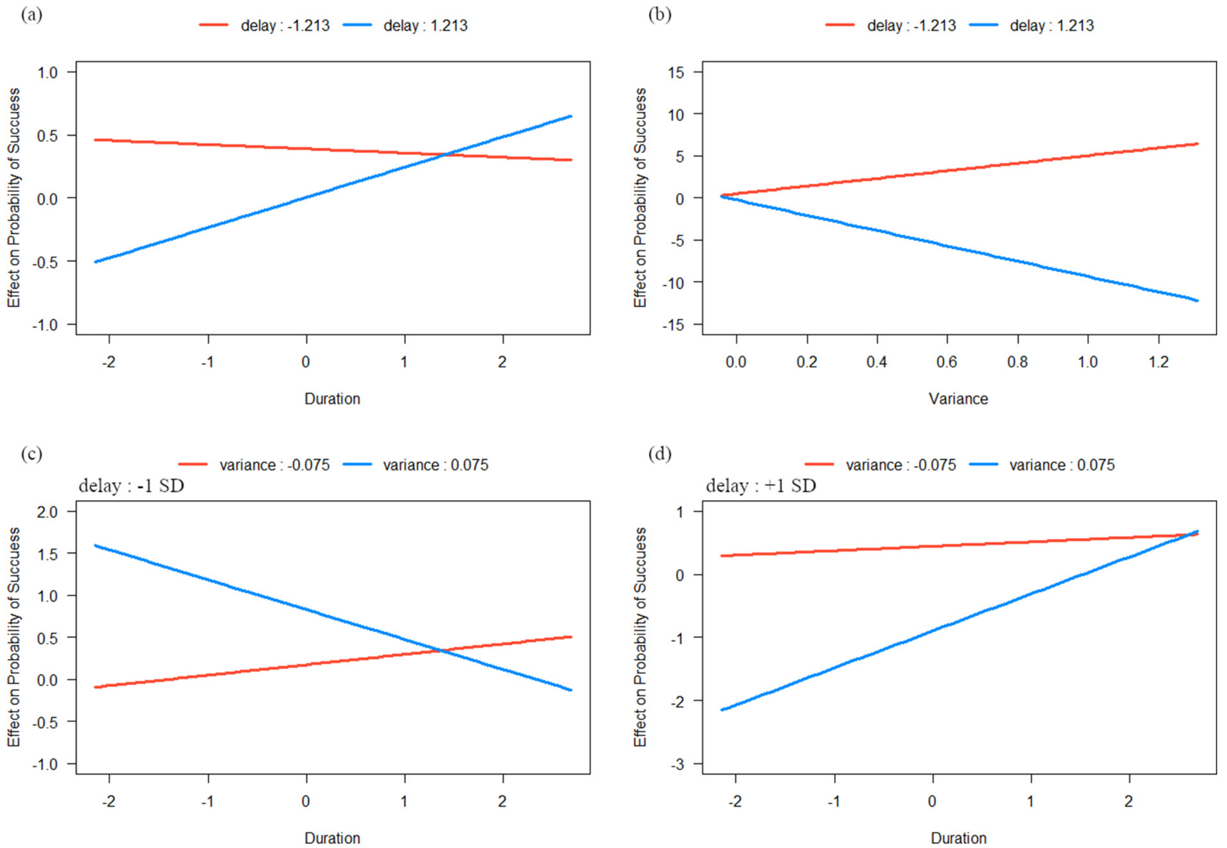

Fig. 3. Interaction effects between the centered variables (a) duration-delay, (b) delay-variance, (c) duration-variance with delay at +1SD, and (d) duration-variance with delay at -1SD.

4. Discussion

The aim of this study was to investigate the effect of planning behavior on performance in CPS. The results show that only one of our three planning indicators has a general effect on CPS performance, but that the effects of all three indicators differ between tasks. In other words, behavior that might be appropriate and highly beneficial in one task might actually be detrimental to performance in another task. Moreover, we found significant interaction effects of the three indicators.

In research question 1 we investigated whether the duration of planning intervals is related to performance. We used the duration of the longest time span between two successive interactions as an indicator for this behavior. The results show that the answer to this question depends on the task: in some tasks it seems to be disadvantageous to perform long planning intervals, while in other tasks relatively long planning phases are beneficial. However, since the correlation between the random effects and task difficulty is rather low, task difficulty does not seem to be crucial for this effect. The PISA framework offers other task characteristics that might lead to such task specific effects such as amount of information, internal complexity, and reasoning skills required (OECD, 2013). For example, a task that has a high amount of information and a high internal complexity might require much planning and therefore benefit from a higher duration of the longest time span between interactions. On the other hand, in tasks that do not require much planning a high duration of the time span between interactions could - in line with the findings of Naumann et al. (2014) - indicate disorientation or mental overload.

In research question 2 we investigated whether the time when planning takes place is related to performance. For this question we used the delay of the longest interval between two interactions as a behavioral indicator. We found a significant negative effect of this indicator suggesting that in general planning should take place at an early stage of the process. Furthermore, the effect of this indicator is depending on task difficulty. In easy tasks it is beneficial to perform the longest planning phase early in the process, while in difficult tasks the effect of this indicator is very small. This supports the assumption that in easy tasks the whole process can and should be planned at an early stage of the task, while in difficult tasks planning ahead the entire process is not possible and, therefore, early planning is not beneficial. Hence, the mechanisms of planning in easy CPS tasks seem to be similar to those in non-complex problems. In these tasks a long initial planning phase was also beneficial (Albert & Steinberg, 2011; Unterrainer & Owen, 2006). In addition, the decrease of the effect in difficult tasks indicates that planning in these tasks can take place later in the process, since difficult tasks might require an initial phase of exploration before planning is useful. Therefore, initial planning seems to be particularly important in easier CPS tasks and non-complex problems.

In research question 3 we investigated whether the variation of planning time is related to performance in CPS. We wanted to know whether it is advantageous to concentrate all planning within a few phases or to spread it equally throughout the process. We used the variance of the times between the interactions as a behavioral indicator for this question. We did not find a general effect of this indicator but, again, we found a task-dependent effect that correlates with task difficulty. The effect is positive in difficult tasks and negative in easy tasks. Therefore, in easy tasks it seems to be important to distribute time equally throughout the process while in difficult tasks it is beneficial to engage in phases of higher and lower activity. This finding supports the assumption that planning is an important part of CPS that should appear at several stages of the process in difficult tasks, as stated in the framework of the OECD (2013).

In research question 4 we investigated whether there are any interaction effects between the three behavioral indicators. We found interaction effects of duration-delay, delay-variance, and duration-delay-variance showing a high interrelatedness of the effects of the three indicators. The interactions show that the effect of the duration of the longest interval depends on the delay: When the delay is high, the duration has a positive effect. When the delay is low the duration does not have an effect. This shows that the absence of early planning might be compensated by investing more time later in the process. The effect of variance also depends on the delay: When the delay is low, variance is positively related to success. When the delay is high, variance is negatively related to success. This indicates that an early planning phase that is longer than the later intervals is most beneficial. If students do not perform such an early planning phase they benefit from an equal time distribution among the whole process, so the absence of an early planning phase can partly be compensated by continued planning activities. Furthermore, the interaction between all three indicators shows that the effect of duration when delay is low also depends on variance. When students concentrate their time in an early planning phase this phase should be rather short. The reason for this negative effect of duration could be that students who take particularly long at the beginning of the task are students with low reading skills that therefore develop a poor understanding of the task. On the other hand, when performing at a steady speed the early planning time should be longer. In this case, students might benefit from an overall lower speed in task processing. On the other hand, when the longest planning period happens late and students work on a steady speed, the duration of planning does not play a role. However, when students concentrate their planning time in a late phase, this phase should be rather long. This could again compensate for the absence of planning at the beginning or during the prior process by investing time at a later stage. All in all, the optimal combination of planning behaviors would be a not too long planning interval at the beginning of the process that is followed by even shorter intervals.

5. Conclusion

To sum up our findings, we argue that the requirements for planning behavior in CPS are highly dependent on the task. The mechanisms in easy tasks seem to be similar to those of non-complex problems: a planning phase at the beginning of the process leads to higher performance. However, with increasing difficulty this effect changes. In difficult tasks, early planning has no benefit since these tasks might require exploration before meaningful planning is possible. In addition, in difficult tasks it is advisable to carefully allocate time at critical points, resulting in a high variation in the intervals between interactions. We also found interaction effects of our three indicators, so we assume that the effects of the three aspects of planning are highly intertwined. The optimal planning behavior identified through the interaction effects is a short planning interval at the beginning of the process that is followed by even shorter intervals.

On a more general level, our results show that the effects of different aspects of planning behavior in CPS tasks are very much interdependent. It seems that a generalization of beneficial behavior in CPS is difficult, since only one of the three behavioral indicators in our study had an overall effect on performance, but the effect of the other two indicators were depending on the values of the other indicators. Moreover, all indicators showed item-dependent effects. Therefore, the appropriateness of behaviors in CPS has to be regarded in the context of the particular task and in the context of other behaviors. One reason for the different requirements of tasks could be task characteristics such as those stated by the OECD (2013). All of these factors might affect the optimal behavior in CPS.

However, this explanation does not stem from empirical findings and therefore the identification of task characteristics, which lead to the mentioned differences, should be subject to future research. Also, the operationalization of planning behavior we chose for our study is only one out of many possibilities, and of course has its weaknesses. As we only rely on time between interactions, it is not certain that students actually did spend this time planning. As we discussed earlier, the time between interactions is also related to other cognitive processes. Moreover, students could have performed other activities such as reading the task or engage in task-irrelevant behavior. Future research might address this issue using techniques like eye tracking or think aloud protocols to find out what mental states occur in students' minds while solving complex problems. Another limitation of this study lies in the method of reducing students' behavior to process indicators. Doing this a lot of information available in the log data is not included in the analysis and only a small proportion of the available information is used. A different approach to gain knowledge from log data is sequential pattern analysis (e.g. lag sequential analysis). Chang et al. (2017) used this method to identify advantageous behavioral patterns in a collaborative problem solving task. Since sequential pattern analysis allows identifying not only single types of behavior but behavioral patterns that are related to success it makes use of information about the whole problem solving process. However, the investigation of process indicators allows for inferences about the strength and shape of the relation between a specific behavior and success. Therefore, both methodological approaches seem suitable to identify beneficial behavior while solving problems.

As a consequence, process data gathered through computer-based assessment contributes to a better understanding of the process of CPS. We introduced three behavioral indicators, which reflect planning behavior in different ways and we showed that these indicators are related to students' performance in CPS tasks. We also showed that tasks require different strategies and that the effects of different behaviors depend on each other. In difficult tasks planning early in the process is less important than in easy tasks and the distribution of time into longer and shorter phases is advantageous. In easy tasks, early planning intervals and a steady process are beneficial. In general, having a short planning interval early in the process followed by more short intervals was identified as the most advantageous behavior. These findings can be used as a basis for both developing further behavioral indicators for cognitive processes and deepening our understanding of planning behavior in CPS.

Source of funding

This research was supported by German Federal Ministry of Education and Research [01LSA1504B].

Declarations of interest

None.

References

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the tower of London. *Child Development, 82*(5), 1501–1517. https://doi.org/10.1111/j.1467-8624.2011.01613.x.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1)https://doi.org/10.18637/jss.v067.i01.

Baxter, G. P., & Glaser, R. (1997). An approach to analyzing the cognitive complexity of science performance assessments: CSE technical report 452. Los angeles. Retrieved from National Center for Research on Evaluation (Los Angeles), website https://www.cse.ucla.edu/products/reports/TECH452.pdf.

Blech, C., & Funke, J. (2005). Dynamis review: An overview about applications of the Dynamis approach in cognitive psychology. Retrieved from Deutsches Institut für Erwachsenenbildung website http://www.die-bonn.de/esprid/dokumente/doc-2005/blech05_01.pdf.

Blech, C., & Funke, J. (2010). You cannot have your cake and eat it, too: How induced goal conflicts affect complex problem solving. *The Open Psychology Journal, 3*, 42–53. Retrieved from http://cogprints.org/6867/1/Blech%26Funke_2010_Polytely.pdf.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education, 53*(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004.

Bransford, J., Brown, A. L., & Cockling, R. R. (1999). *How people learn: Brain, mind, experience, and school.* Washington, D.C: National Academy Press.

Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch, & J. Funke (Eds.). *Complex problem solving: The European perspective* (pp. 27–63). Hillsdale, NJ: L. Erlbaum Associates.

Chang, C.-J., Chang, M.-H., Chiu, B.-C., Liu, C.-C., Fan Chiang, S.-H., Wen, C.-T., et al. (2017). An analysis of student collaborative problem solving activities mediated by collaborative simulations. *Computers & Education, 114*, 222–235. https://doi.org/10.1016/j.compedu.2017.07.008.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., et al. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software, 39*(12), 1–28. https://doi.org/10.18637/jss.v039.i12.

Dörner, D. (1989). *Die Logik des mißlingens: Strategisches Denken in komplexen situationen (17. - 19. Tsd).* Reinbek bei Hamburg: Rowohlt.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving, 4*(1), 19–42. https://doi.org/10.7771/1932-6246.1118.

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. 24–43). Hillsdale, NJ: L. Erlbaum Associates.

Funke, J., & Frensch, P. A. (2007). Complex problem solving: The European perspective-10 Years after. In D. H. Jonassen (Ed.). *Learning to solve complex scientific problems* (pp. 25–47). New York: Lawrence Erlbaum.

Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's matrices. *Journal of Intelligence, 3*(1), 21–40. https://doi.org/10.3390/jintelligence3010021.

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. https://doi.org/10.1037/a0034716.

Greiff, S., & Funke, J. (2008). *Indikatoren der Problemlöseleistung: Sinn und Unsinn verschiedener Berechnungsvorschriften: Bericht aus dem MicroDYN Projekt*.

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. https://doi.org/10.1016/j.chb.2016.02.095.
Klieme, E. (2004). Assessment of cross-curricular problem-solving competencies. In J. H. Moskowitz, M. Stephens, J. Moskowitz, & M. Stephens (Eds.). *Comparing learning outcomes: International assessment and education policy* (pp. 81–107). London: Routledge Falmer.
Lesh, R., & Zawojewski, J. (2007). Problem solving and modeling. In F. K. Lester (Ed.). *The handbook of research on mathematics teaching and learning* (pp. 763–804). .
Leutner, D., Funke, J., Klieme, E., & Wirth, J. (2005). Problemlösefähigkeit als fächerübergreifende Kompetenz. In E. Klieme, D. Leutner, & J. Wirth (Eds.). *Problemlösekompetenz von Schülerinnen und Schülern: Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (pp. 11–19). (1st ed.). Wiesbaden: VS Verl. für Sozialwiss.
Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander, & P. H. Winne (Eds.). *Handbook of educational psychology* (pp. 287–304). (2nd ed.). Mahwah, N.J: Erlbaum.
Naumann, J., & Goldhammer, F. (2017). Time-on-task effects in digital reading are non-linear and moderated by persons' skills and tasks' demands. *Learning and Individual Differences, 53*, 1–16. https://doi.org/10.1016/j.lindif.2016.10.002.
Naumann, J., Goldhammer, F., Rölke, H., & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiebasierten Umgebungen: Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen. *Zeitschrift für Padagogische Psychologie, 28*(4), 193–203. https://doi.org/10.1024/1010-0652/a000134.
Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak, & R. G. Morrison (Eds.). *The cambridge handbook of thinking and reasoning* (pp. 321–349). Cambridge: Cambridge University Press.
OECD (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy.* PISAParis: OECD.
OECD (2015). *Students, computers and learning: Making the connection.* Paris: OECD.
Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin, 136*(1), 65–86. https://doi.org/10.1037/a0017815.
R Core Team (2016). *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.
Reeff, J., Zabal, A., & Blech, C. (2006). The assessment of problem-solving competencies: A draft version of a general framework. Retrieved from Deutsches Institut für Erwachsenenbildung website http://www.die-bonn.de/esprid/dokumente/doc-2006/reeff06_01.pdf.
Rouet, J.-F., & Le Bigot, L. (2007). Effects of academic training on metatextual knowledge and hypertext navigation. *Metacognition and Learning, 2*(2–3), 157–168. https://doi.org/10.1007/s11409-007-9011-z.
Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology Paris, 99*(4–6), 308–317. https://doi.org/10.1016/j.jphysparis.2006.03.014.
Vosniadou, S., & Ortony, A. (1989). *Similarity and analogical reasoning.* Cambridge, New York: Cambridge University Press.
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.