Unsuccessful and Successful Complex Problem Solvers – A Log File Analysis of Complex Problem-solving Strategies across Multiple Tasks

Björn Nicolay, Florian Krieger, Jörg-Tobias Kuhn, Arthur C. Graesser, Dirk Ifenthaler, Ryan S. Baker, Samuel Greiff

Making use of digital technologies to evaluate and advance students' skills and competencies has become a hallmark of education in large parts of the world (Curran et al., 2019; Scherer et al., 2019; Selwyn, 2012). By considering such computer-based technologies, educational stakeholders seek to assess and foster relevant competencies in their students in order to equip them with a fundamental knowledge base for their further educational trajectory and beyond (Buitrago Flórez et al., 2017; Martinez, 2000; Shute & Rahimi, 2017). One particular set of competencies that bear crucial relevance for today's students and are therefore comprehensively addressed using digital technologies are 21st century skills, which incorporate a range of different skills (Amar & David, 2016; Eguchi, 2016; Lapek, 2017; Mayrath et al., 2012; Sanabria & Arámburo-Lizárraga, 2017). One of the arguably most studied 21st century skills are Complex Problem Solving (CPS) skills (Ananiadou & Claro, 2009; Geisinger, 2016; Greiff et al., 2014). Notably, CPS and intelligence are strongly intertwined (Goode & Beckmann, 2010; Stadler et al, 2015). As such, several existing studies have emphasized that the ability to successfully solve (complex) problems represents a key aspect of overall intelligence (Beckmann & Guthke, 1995; Lotz et al., 2017; Resnick & Glaser, 1976). Additionally, CPS has been found to be a significant antecedent of successful performance on educational and work-related settings, beyond traditional proxies of intelligence such as reasoning ability (Eseryel et al., 2011;

Greiff et al., 2013; Kretzschmar et al., 2016; Mainert et al., 2015; OECD, 2009; 2014; Rohde &

Thompson, 2007; Wüstenberg et al., 2012). Thus, CPS is a crucial skill to consider in the

assessment and fostering of students' competencies in educational settings (e.g., Deary et al.,

2007).

Moreover, studies of students' CPS skills take place at the intersection of intelligence

and education, two strongly intertwined research fields of equal relevance for effectively

preparing today's students for future challenges (Mayer, 2000; Novalinda et al., 2020).

Therefore, the need to improve scientific knowledge about the underlying key elements of CPS

success becomes paramount. Extending evidence-based understanding of what differentiates

success from failure in CPS is particularly important when seeking to identify factors to include

in computer-based learning simulations aimed at enhancing students' CPS skills (e.g., Molnár &

Csapó, 2018). On a broader level, thoroughly investigating the mechanisms that drive CPS

performance can help us identify beneficial factors for computer-based educational training

programs targeting students' broader development in science, technology, engineering, and

mathematics (STEM) fields as well as related domains sharing underlying characteristics with

CPS (Murphy et al., 2020; Xie et al., 2015).

Since the early 21st century, computer-based microworlds have been a promising way to

capture and assess CPS (Baker & O'Neil Jr., 2002; Wirth & Klieme, 2003). One such approach

that has been increasingly applied recently is to scrutinize data stored in computer-generated log

files while problem solvers work on particular problems (Lin et al., 2016; Liu et al., 2018; Ren et

al., 2019; Teig et al., 2020; Tóth et al., 2017). The unique advantage of log file analysis

compared to other methods such as verbal protocols is the element of objectivity, since each

action by a participant is monitored and stored automatically without the user made explicitly

aware of these processes (e.g., Adams et al., 2015). Several studies have already utilized log files

to uncover information about strategies used or time spent on individual or multiple CPS tasks

(Greiff et al., 2018; Greiff et al., 2016; Greiff et al., 2015; Molnár & Csapó, 2018; Mustafić et

al., 2019; Sonnleitner et al., 2012; Stadler et al., 2019). However, despite revealing some aspects

of strategy use in CPS, to the best of our knowledge, research has yet to simultaneously

scrutinize multiple strategies over multiple tasks in terms of their relationship with CPS

performance. Hence, the current study takes a pioneering role in performing such an analysis.

Among CPS strategies assessed in previous research, one strategy known as *vary-one-thing-at-a-time* (VOTAT; also referred to as *Control of Variables*, CVS; e.g., Kuhn & Dean Jr,

2005), which describes systematically varying only a single variable in order to detect its unique

effect(s) on the remaining variables (Schwichow et al., 2016; Tschirgi, 1980), has been found to

be particularly beneficial for solving complex problems. However, we still lack a detailed

understanding of whether (and, if so, which) additional strategies might also play a (beneficial)

role in CPS. Existing studies have also yielded limited evidence on how the joint application of

different strategies is associated with CPS performance across multiple CPS tasks. Answering

these questions is of particular importance for establishing future training programs to help

students become better complex problem solvers, which is essential to prepare the current

generation of students for life in the 21st century (e.g., Greiff & Neubert, 2014). Hence, this

study examined log files from a large-scale educational dataset of ninth graders solving multiple

complex problem tasks in order to enrich our understanding of how students approach such

complex problems and uncover potential differences between successful and unsuccessful

problem solvers in strategy use and number of interactions with the computer system across multiple CPS tasks with varying characteristics.

1.1 Complex Problem Solving: Definition, Importance for Education, Assessment

There are a number of open questions regarding how to prepare students for their future educational trajectory and beyond and support them during their educational journey through training programs targeting skills that have been shown to be related to educational performance. CPS is one such valuable skill. In general, a *problem* occurs when a desired goal state differs from the actual current state (Mayer & Wittrock, 2006). A *complex problem*, in turn, can be characterized as a situation encompassing a number of features, including multiple interrelated variables ('complexity') with hidden connections ('intransparency') that may autonomously change over time ('eigendynamics'; Stadler et al., 2019) and requiring the user to pursue several goals simultaneously ('polytely'; Buchner, 1995; Dörner, 1980).

The relevance of CPS skills for successful performance on both education and working life has been highlighted in existing research. The dynamic nature of a complex problem is of particular relevance here, as it mirrors the dynamic, continuously changing environment of the 21st century (Trilling & Fadel, 2009; Witherington & Boom, 2019), with non-routine tasks becoming increasingly significant in various life domains (Marcolin et al., 2019; Reijnders & de Vries, 2018). Indeed, CPS skills have shown to be predictive of educational and job-related success (Lotz et al., 2016; Mainert et al., 2019; Schweizer et al., 2013; Sonnleitner et al., 2013; Wüstenberg et al., 2012). Moreover, CPS assessment was included in the 2012 cycle of the Program for International Student Assessment (PISA; termed 'creative problem solving' in this context; OECD, 2014). As PISA aims to test competencies that are particularly relevant for

students' developmental trajectory (OECD, 2014), including CPS skills in this large-scale

assessment program serves as another indicator of the widely recognized educational importance

of these skills. Similarly, previous research has highlighted processes relevant for CPS in

providing educational scaffolding to students via computer-based programs (Gobert et al., 2012).

Thus, CPS skills are a prototypical example of competencies for which digital technologies can

be leveraged to advance education, as they are both trained and evaluated in computer-based

environments (Kretzschmar & Süß, 2015).

A broad range of tools exist to assess individuals' CPS skills (Dörner et al., 1983; Funke,

2003; Greiff et al., 2013; Greiff et al., 2012; Jonassen, 2011; Sonnleitner et al., 2012). Today, the

most common such tools are computer-based microworlds encompassing a number of complex

problem tasks, each with a different set of input variables that can be actively manipulated by the

participant, affecting several output variables (for an overview of CPS assessment tools, see

Stadler et al., 2015).

One commonly used CPS assessment approach in educational and research settings,

which is also employed in this study, is called MicroDYN (Greiff et al., 2012). MicroDYN is

part of the minimally complex systems approach family that applies linear structural equations as

its underlying framework (Funke, 2001; Greiff et al., 2013). This means that assessments based

on the MicroDYN approach consist of multiple items, each with a clearly defined underlying

structure of relations between variables. Like other CPS assessment frameworks, MicroDYN

aims at identifying performance indicators for respective phase of the overall CPS process

(Beckmann et al., 2017; Fischer et al., 2012).

Similar to any type of problem-solving in general, the process of solving a complex problem encompasses two distinct phases (Beckmann et al., 2017; Fiore et al., 2002; Funke, 1993; Purzer et al., 2018; Shute et al., 2016). In the first phase, knowledge acquisition, participants are asked to familiarize themselves with the problem space in order to discover the underlying variable relationships. Due to the unique initial intransparency of complex problems, the solver is required to actively explore the problem space during this phase in order to uncover the variable relationships. Subsequently, participants are asked to apply their previously acquired knowledge in the knowledge application phase. In order to succeed in this phase, they are required to achieve target goal values on certain variables by efficiently manipulating other variables within a given number of steps (e.g., Fischer et al., 2012).

Given the different requirements of the two phases, being able to draw inferences about solvers' performance on each respective phase using fine-grained data sources such as log files represents a promising approach to learn more about the underlying indicators of successful vs. unsuccessful CPS performance. Such log files, which comprise data about a participant's problem-solving behavior for a given complex problem, are created unobtrusively and stored automatically during the problem-solving process. Hence, their analysis represents one way of exploring the relations between variables, such as how behavioral patterns when working on the tasks and overall performance relate to certain background variables (Angeli et al., 2017; Baker & Siemens, 2015). Despite its potential, however, this approach has yet to be fully utilized in the field of CPS (see, e.g., Ifenthaler et al., 2018).

In particular, the information contained in log files encompasses all possible interactions by the participant in the computer-based environment, for instance, which input variables they

manipulated and when, as well as time spent on each round (i.e., each coherent step evaluating the potential impact of certain variables on other variables) and each task, and their score in each respective phase. Therefore, in essence, a log file represents a particular participant's entire interaction pattern while working on a given CPS task (e.g., Xu et al., 2018). Previous studies have advanced knowledge of the processes underlying CPS by analyzing log file data, such as how individual differences in overarching CPS skills relate to the application of specific strategies (Lotz et al., 2017) or time on task (Greiff et al., 2016). However, previous research has been limited to one or two strategies only (e.g., Greiff et al., 2018), making it impossible to draw inferences with regard to other potential strategies or interactions between various strategies. Furthermore, even recent studies have focused rather narrowly on a few salient factors, such as time between actions (i.e., planning; Eichmann et al., 2019) or performance on a single task (e.g., Xu et al., 2018). This illustrates the presence of some significant research gaps despite the large body of research on CPS strategies.

In this study, we aim to extend the CPS literature by analyzing the application rates of all possible strategies (see Section 1.2) in multiple CPS tasks of varying difficulty. In our investigations, we will focus on successful CPS performance during the knowledge acquisition phase, since this phase allows for a greater variety of strategy application given its less stringent limitations in terms of time and/or rounds (Fischer et al., 2012). In the analyses, we aim to obtain a more comprehensive picture of both the actions and behavioral patterns underlying CPS, with the ultimate goal of developing empirically grounded approaches to facilitate computer-based CPS in educational settings.

1.2 Success and Failure in Complex Problem Solving: What We Do and Do Not Know about Strategy Application

The skills to solve a complex problem successfully are manifested in different sets of planned interventions and variable manipulations (i.e., strategies; Vollmeyer et al., 1996). Generally, using a systematic approach, as opposed to a series of random or unplanned operations, leads to increased CPS success rates (Lotz et al., 2017).

Previous research on CPS strategies has uncovered a variety of strategies that are potentially beneficial for solving complex problems. The most widely researched strategy is termed VOTAT (or CVS; Kuhn & Dean Jr., 2005). In using this strategy, the participant varies only one input variable at a time in order to observe its unique effect on one or multiple output variables (van der Graaf et al., 2015). The systematic application of VOTAT has been found to be a precursor of successful CPS performance in several previous studies (Eichmann et al., 2020; Greiff et al., 2015; Herde et al., 2016). In addition, various strategies adjacent to VOTAT exist that have not been studied as thoroughly in the past. For instance, the 'hold-one-thing-at-a-time' (HOTAT) strategy is applied when a participant systematically holds one input variable while actively manipulating the remaining input variables (Tschirgi, 1980). While this strategy might be beneficial for detecting certain interaction effects of two or more input variables on a given output variable, it has generally been treated as not a truly scientific approach, since the participant does not analyze the effects of a single variable in isolation, limiting their ability to draw inferences about the impact of this single variable (Tschirgi, 1980). Furthermore, the 'vary-no-thing-at-a-time' (NOTAT) strategy refers to non-interfering observation (i.e., an idle round) in order to detect possible autonomous changes in the output variables (i.e., eigendynamics).

Importantly, NOTAT has been found to lead to higher success rates in solving complex problems, but only in cases where it is used to supplement the VOTAT strategy and only when eigendynamics are present in a given task (Lotz et al., 2017; Schoppek & Fischer, 2017). Lastly, when trying to solve a particular complex problem, participants can apply the 'change all' (CA) strategy. This strategy refers to manipulating all input variables simultaneously, and has been shown in previous studies to be a precursor of failure in both general problem-solving (Tschirgi, 1980) and CPS (Vollmeyer et al., 1996). To illustrate each strategy, Figure 1 shows their application in a sample task with three input variables that can be manipulated.

Figure 1

Overview of all possible input variable manipulations (i.e., strategies that can be applied) for a given MicroDYN task with three input variables, like the one used in the present study
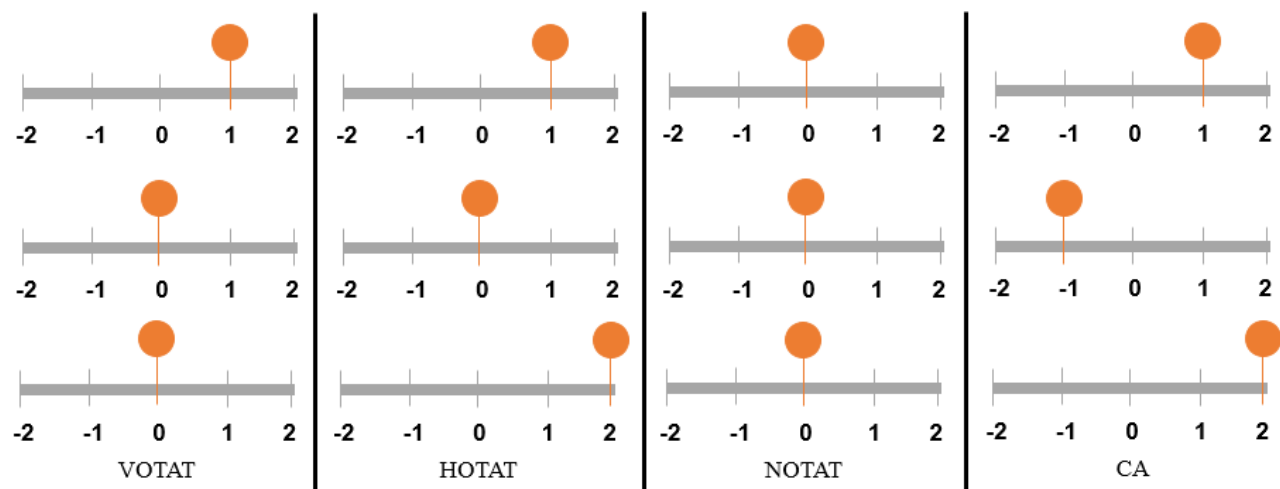
Taken together, the current state of evidence with regard to the underlying actions leading to successfully solving a complex problem suggests that (1) particular, specific strategies are useful when solving a CPS task (e.g., VOTAT). However, it remains unknown how often related

strategies, such as HOTAT or CA, are applied, and if they can also contribute to CPS success under certain conditions. Furthermore, (2) some strategies are only beneficial under certain conditions (e.g., NOTAT when eigendynamics are present), and (3) there are indications that the application of any single strategy is insufficient for successful CPS (see, e.g., Wu & Molnár, 2021).

1.3 The Present Study

Our existing knowledge about strategies as well as the current gaps in the literature had the following implications for the present study. First, while the application of a particular strategy such as VOTAT improves the success rates for complex problem-solving tasks, complementary strategies such as NOTAT, HOTAT, or CA have received little attention in research thus far. Consequently, it remains to be investigated which strategy combinations significantly alter the chances of CPS success and, more generally, to what extent the combination of different strategies acts as a unique precursor of overall CPS performance. Second, it has been shown that NOTAT application is beneficial as a complement to VOTAT in CPS tasks with eigendynamics (Lotz et al., 2017). However, this finding has yet to be replicated with a large comprehensive sample. In sum, these implications led to the following research questions (RQs) and joint expectations, which we approached by coding and analyzing log file data from a large-scale student sample recording each individual action by participants across multiple CPS tasks with varying characteristics:

RQ 1. What strategies do students use when trying to solve a complex problem, including the already well-researched strategy of VOTAT? Given the results of previous research, we

expected students to use VOTAT more than any other strategy, followed by CA, NOTAT, and HOTAT.

RQ 2. Is the use of a particular strategy sufficient to successfully solve complex problems (i.e., to what extent do the rates of use of individual strategies and combinations of strategies play a role in CPS performance)? In particular, we expected that the use of VOTAT would lead to higher chances of success than any other single strategy. In addition, we expected that using multiple strategies together would significantly improve the chances of successful performance compared to using VOTAT alone.

RQ 3. How does the presence vs. absence of eigendynamics in a given CPS task influence CPS performance? In addition, how does the presence vs. absence of eigendynamics influence the usefulness of combining strategies? Based on previous research, we expected that fewer students would be able to successfully solve CPS tasks with eigendynamics. In addition, we investigated whether the presence or absence of eigendynamics modulates the effect of specific strategy combinations on successful performance.

2. Materials and Methods

2.1 Sample Characteristics

This study examined a sample of ninth graders from a panel assessment study conducted in a Finnish municipality, which aimed to help students develop learning-specific competencies. Cohort selection was performed by the municipal school board, with the full grade cohort of students attending different schools in the municipality eligible for participation. The panel can be seen as representative of the Finnish population with respect to several socioeconomic and demographic characteristics (Vainikainen, 2014), thus allowing for generalizable findings. The

initial sample size was $N = 1,508$; 48 percent were female and 50 percent were male (2% missing information). The mean age was 15.8 years ($SD = 0.43$). Informed consent was required from all students before participation. After processing the files according to our inclusion and exclusion criteria (see Section 2.5.1 for further information), the final sample size used for statistical analysis was $N = 1,276$.

2.2 Materials

The present study used the well-established MicroDYN framework (Greiff et al., 2012) to assess the students' CPS performance. MicroDYN inherently captures the two phases, knowledge acquisition and knowledge application, in a variety of different made-up contexts (i.e., tasks) using arbitrary input and output variable names. This eliminates any influence of prior knowledge on a particular topic in order to avoid corresponding bias. In any MicroDYN task, the participant initially has the opportunity to freely manipulate one or more input variables, displayed on the left side of Figure 2 ('Rexol', 'Menol', 'Sarol'), in order to analyze their effects on the output variables, 'Headache', 'Diastolic Blood Pressure', and 'Antibodies' in the depicted task 'Medical Aid'. The three cursors for the respective input variables can be moved independently of one another, and clicking on 'Apply' will result in observable value changes in the output variables. In this phase, the problem solver is asked to draw a model of all existing relationships between input and output variables in a diagram (see bottom of Figure 2). Note that clicking on 'Apply' automatically returned all input variable values to their initial 'neutral' state (as shown in Figure 2), irrespectively of how many variables had been varied and to what extent before clicking on 'Apply'. Afterwards, in the knowledge application phase, the participant receives a specific target goal (displayed as individual ranges of values for each output variable,

e.g., 66-70 for 'Antibodies'), which they must achieve in no more than four steps (i.e., by

clicking on 'Apply' a maximum of four times before task termination is enforced by the

system).

Figure 2

Problem space (top) featuring three input variables (left; e.g., 'Sarol') and three output

variables (right; e.g., 'Headache'), and visual model of variable relationships (bottom) from the

MicroDYN task "Medical Aid" during the knowledge acquisition phase
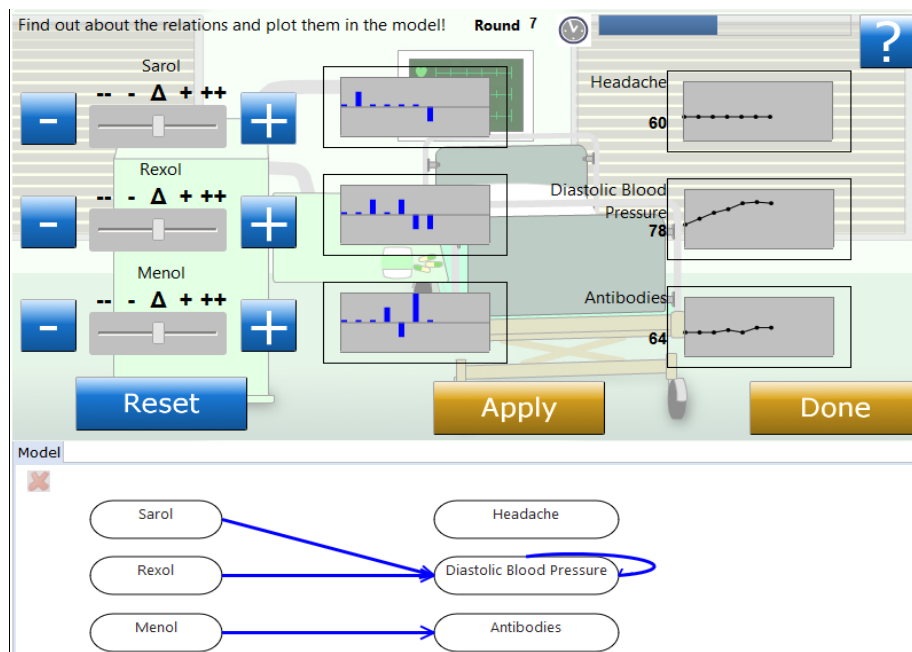
2.3 Procedure

The students were asked to complete a computer-based CPS assessment that took place in

their school's computer lab. Overall, the test consisted of nine consecutive MicroDYN tasks, six

of which were included in the statistical analysis, each taking about five minutes to complete.

We only included the six tasks with three input variables in the analyses (see Figure 2), as the

remaining tasks contained only two input variables, making it impossible to distinguish VOTAT from HOTAT. The students had to complete the tasks in a fixed, predefined order of appearance.

First, an example task with instructions involving text and a short video explanation about the step-by-step process of solving a MicroDYN task appeared, followed by five tasks that did not contain dynamic effects (i.e., eigendynamics). Afterwards, another set of instructions including text and a short video was displayed, this time introducing eigendynamics. Of the four subsequent tasks, three contained eigendynamics. For each task, in the knowledge acquisition phase, participants were allowed to perform an unlimited number of operations within up to 180 seconds in order to discover relationships between the input and output variables, which they then illustrated by drawing arrows on a causal model diagram (see bottom of Figure 2). Afterwards, in the knowledge application phase, participants had to use their acquired knowledge to achieve certain target goals (i.e., different value ranges for all output variables) in a maximum of four steps (after clicking on "Apply" for the fourth time, they received a message that the task was terminated, irrespective of whether they had successfully achieved the goal state or not).

2.4 Variables and Scoring

All actions performed by each participant for each task were stored in XML files. We used a Python script written for this purpose (Van Rossum & Drake, 2009) to code each action and score whether the task had been solved correctly or incorrectly (please find the script under this link: REF). Participants were scored on their performance in the knowledge acquisition phase, meaning that each action performed during this phase was used in the analysis. In addition, a dichotomous overall performance rating was applied (success or failure), with a

successful answer characterized by drawing a correct model indicating the relationships between input and output variables. In turn, any inclusion of incorrect relations and/or omission of correct relations was considered a failure. Regarding the individual strategies used, participants were able to apply one of four different strategies (see Figure 1) in each respective round: 'vary-one-thing-at-a-time' (VOTAT), 'hold-one-thing-at-a-time' (HOTAT), 'vary-no-thing-at-a-time' (NOTAT), or changing all input variables simultaneously (i.e., CA). Each of these strategies was scored on the task level as the absolute frequency with which it was applied for each of the six tasks (e.g., if Participant A used HOTAT five times in the first task, they received an absolute frequency score of five for HOTAT for this task). Furthermore, we coded how many 'strategic' actions each participant performed (i.e., how often the participant clicked on apply) and coded this as the number of rounds.

2.5 Statistical Analysis

2.5.1 Filtering

The initial dataset was filtered to exclude erroneous data outside the scope of the present study. Firstly, we removed all data from participants who did not appear to attempt to explore (i.e., completed fewer than three rounds) and who completed more than 60 rounds of a given task. Note that this filter only removed *items* that met the criteria, but we did not perform listwise deletion. In total, 2528 items were filtered out. In detail, $N = 202$ participants did not meet the inclusion criteria in all six items and were excluded, while $N = 97$ participants (7.6%) remained with one item in the dataset, $N = 90$ participants (7.1%) remained with two items, $N = 78$ participants (6.1%) remained with three items, $N = 103$ participants (8.1%) remained with four items, $N = 134$ participants (10.5%) remained with five items, and the majority of $N = 774$

participants (60.6%) remained with six items. In addition, participants who performed at least

one item twice were excluded, as these were not due to item resumption (picking up where one

left off), but rather to test reloading (starting over; $N = 30$, second filter). A sample of N = 1,276

students was finally included in the following statistical analyses.

2.5.2 Statistical Analyses

To determine how often strategies were used by students across all six CPS tasks (RQ1),

we calculated the mean absolute rates of use for each strategy in each item. In addition, we

performed a Poisson regression within the Generalized Linear Mixed Models (GLMMs)

framework with random intercepts for both items and participants and strategy type as fixed

effect. Since strategy type was categorical, n-1 (i.e., three) dummy coded variables representing

the presence or absence of each category were built by the statistical program and used as

separate strategy predictors in the model. The reference category was set to VOTAT and, thus,

can be considered as a baseline reflected in the general fixed effect (intercept 0). The model is

depicted in equation (1).

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i} + \beta_1\, NOTAT + \beta_2\, HOTAT + \beta_3\, CA$$

(1)

In this model, η_{pi} denotes the logarithm of the Poisson-distributed rate for person p in

item i , 0 the general fixed effect (intercept) of the model (i.e. VOTAT application), b0p the

random intercept for participants, b0i the random intercept for items, and 1 to 3 the fixed effects

of different strategy type (i.e., the application of a certain strategy, compared to VOTAT

application).

In preparation for RQ2 and RQ3, we first calculated the combination of strategies used by each participant in each item. This approach scored which strategies were applied but did not take into account how often each strategy was used in each item or in which order the strategies were applied to reduce complexity of the models. For example, if a participant used VOTAT three times, CA once, and HOTAT three times in an item, we coded this as 'VOTAT-HOTAT-CA' for that item. If a participant used VOTAT ten times in an item, we coded this as 'VOTAT' for that item. If a participant only used strategies that were not expected to be useful for solving an item, we coded this as "not useful". For example, if a participant used CA ten times and then HOTAT three times, we coded this as 'not useful' for that item. Note that we also coded the use of NOTAT without the use of VOTAT as 'not useful', as we assumed, based on previous studies, that the use of NOTAT without VOTAT would not be useful for solving an item.

This coding procedure resulted in a categorical variable with nine meaningful strategy combinations as factor levels. The absolute frequencies for each item and the marginal distributions are shown in Table 1. Note that we excluded the factor level 'V.H' (i.e., VOTAT+HOTAT) for RQ2 and RQ3 because the corresponding standard error of this estimate was unreasonably large (SE = 20.47 for the interaction term in RQ3; see below for the specific model), making this estimate difficult to interpret. This was probably due to the fact that HOTAT was applied quite rarely compared to the other strategies (see also results of RQ1). We provide the analyses for both RQ2 and RQ3 including the factor level 'V.H' in the supplementary material. Please note that although some numerical differences in the coefficients between the models with and without this factor levels have been observed, no differences with regard to the main implications of this study have been found.

Table 1

Frequencies of strategy combinations for each item and marginal distributions

Item	V	NotUseful	V.CA	V.H	V.H.CA	VN	V.N.CA	V.N.H	V.N.H.CA	**Total**
Aid	377	200	21	20	46	235	3	19	10	**931**
Game	626	259	28	48	71	46	2	1	20	**1101**
Gardening	351	224	60	44	93	164	33	35	45	**1049**
Handball	406	217	24	45	61	228	6	17	10	**1014**
Moped	555	283	52	81	113	47	11	9	22	**1173**
Spaceship	452	189	4	29	58	206	5	16	10	**969**
Total	**2767**	**1372**	**189**	**267**	**442**	**926**	**60**	**97**	**117**	**6237**

Note. V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all; NotUseful includes applications of NOTAT, HOTAT, CA without application of VOTAT.

To address RQ2, we used a GLMM based on the 1-parameter logistic item response model with both fixed and random effects. Please note that we previously tested for the absence of local item dependence of the six items to ensure that the results could be properly interpreted. We used Yen's Q statistic (Yen, 1984) and found that no residual correlations exceeded the .20 cut-off, so local independence was assumed.

We used the logistic regression shown in equation (2) to address RQ2, with random intercepts for both participants and items, and a fixed effect for strategy combinations. Since strategy combinations were categorical, n-1 (i.e., seven) dummy coded variables representing the

presence or absence of each combination were built by the statistical program and used as separate predictors, each representing one strategy combination, in the model. The reference category was set to VOTAT and, thus, can be considered as a baseline reflected in the general fixed effect (intercept). The model is depicted in equation (2).

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i} + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 .$$

with

$X_1 = $ NotUseful

$X_2 = $ V.CA

$X_3 = $ V.H.CA,

$X_4 = $ V.N,

$X_5 = $ V.N.CA

$X_6 = $ V.N.H

$X_7 = $ V.N.H.CA

(2)

In this model, η_{pi} denotes the logit of successfully solving an item for person p in item i, β_0 the general fixed effect (intercept) of the model, b_{0p} the random intercept for participants, b_{0i} the random intercept for items, β_1 to β_7 the fixed effects of strategy combinations. Since we used VOTAT as the reference, estimates β_1 to β_7 can be interpreted as an increase or decrease in the logit of the probability of solving an item relative to VOTAT use only. That is, a positive estimate for an estimate β_1 to β_7 indicates an increase in the logit of the probability of solving an item relative to using VOTAT only, while a negative estimate indicates a decrease in the logit of

the probability of solving an item relative to using VOTAT only. We also calculated the

exponentiated estimates of these estimates, which represent the odds ratios (OR) of the odds for

applying a strategy combination other than VOTAT usage only divided by VOTAT usage only.

Hence, OR can be interpreted as an increase (or decrease) of odds of a certain strategy

combination compared to VOTAT. For instance, if the OR of 'V.N' would be 3.2, it would be

3.2 times more likely to solve an item in the case that VOTAT and NOTAT were applied;

compared to when VOTAT only was applied.

To address RQ3, we extended the model depicted in equation (2) by including the

presence or absence of eigendynamics (ED) in items as a further predictor and added an

interaction terms between ED and all predictors of different strategy combinations (β_1 to β_7),

resulting in seven interactions. The model is shown in equation (3).

$$\eta_{pi} = \beta_0 + b_{0p} + b_{0i} + \beta_1 X_1 + \ldots + \beta_7 X_7 + \beta_8 ED + \beta_9 (X_1 \times ED) + \ldots + \beta_{15} (X_7 \times ED).$$

with

$$X_1 = \text{NotUseful}$$

$$X_2 = \text{V.CA}$$

$$X_3 = \text{V.H.CA},$$

$$X_4 = \text{V.N},$$

$$X_5 = \text{V.N.CA}$$

$$X_6 = \text{V.N.H}$$

$$X_7 = \text{V.N.H.CA}$$

(3)

The variable containing the presence vs. absence of eigendynamics was contrast-coded with (1 = ED, -1 = noED) to allow for an interpretation of the estimates for the general intercept, whereas the estimates for β_1 to β_7 were comparable to the model depicted in equation (1). For the interactions, the estimates (9 to 15) can be interpreted as a moderation of the slope of the respective predictors of strategy combinations (β_1 to β_7) by the presence or absence of eigendynamics (8). The exponentiated estimates of the interactions thus represent the change in OR (of a certain strategy combination compared to VOTAT usage only) when eigendynamics are present vs. when they are not present. Hence, these exponentiated coefficients can be interpreted as ratio of OR. A proper interpretation of these interactions thus depends on the consideration of the intercept of the models and the coefficients of all predictors involved in this interaction (strategy combination and eigendynamics). To facilitate the interpretation, we calculated the estimated marginal means for each strategy combination for both levels of eigendynamics (presence vs. absence) and observed the predicted probabilities.

Analyses were conducted using statistical software *R* (R Core Team, 2020). GLMMs were performed with the *lme4* package (Bates et al., 2015). Estimated marginal means and pairwise comparisons were performed using *emmeans* (Lenth, 2022). Further *R*-packages used for analyses included *easystats* (Lüdecke et al., 2022) *jmv* (Selker et al., 2022), *subscore* (Dai et al., 2022), and *jtools* (Long, 2022). The *R*-script containing all statistical analyses reported in this manuscript can be found on OSF (https://osf.io/gh6qj/?view_only=05875ec7604040a08c3b97653a62b7c9).

3. Results

3.1 Which strategies were used how often (RQ1)?

Across all items, VOTAT (M = 4.9) was applied most often, followed by CA (M = 1.98), NOTAT (M = 1.06), and HOTAT (M = 0.63). Poisson regression revealed that VOTAT was applied more often than all other strategies (for more details see Table 2 below).

Table 2

Results of Poisson regression

Predictors	Log-Mean	IRR	p
(Intercept)	1.34 [1.25, 1.42]	3.81 [3.50, 4.14]	**<0.001**
V	*Reference*		
C	-0.82 [-0.84, -0.80]	0.44 [0.43, 0.45]	**<0.001**
H	-1.96 [-1.99, -1.93]	0.14 [0.14, 0.15]	**<0.001**
N	-1.45[-1.48, -1.42]	0.23 [0.23, 0.24]	**<0.001**

Random Effects		
Variance	Participants	0.35
	Item	0.01
N	Participants	1276
	Item	6
Observations	Total	24948
R^2	Marginal	0.397

Conditional	0.662

Note. V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all. Values in parentheses represent bounds of the 95 confidence interval. IRR = Incidence Rate Ratios.

Post-hoc analyses further revealed that CA was more often applied than HOTAT (ratio 3.13), CA was more often applied than NOTAT (ratio 1.88), and HOTAT was less applied than NOTAT (ratio 0.60; for more details see Table 3).

Furthermore, we identified overdispersion (dispersion ratio = 5.1, p < 0.01) in the Poisson regression, which may affect the estimates and corresponding p-values. To ensure the reliability of our analyses, we conducted an observation-level random effects (OLRE) Poisson regression (Harrison, 2014). This model demonstrated no overdispersion (dispersion ratio = 0.13, p > 0.05) and yielded similar directions of estimates and p-values as the initial Poisson regression with overdispersion (including the post-hoc comparisons), which indicates the robustness of the initial model. The supplementary material in the OSF repository of this article contains the results of the additional OLRE model.

Table 3

Post hoc comparisons for frequencies of strategy applications

Comparison	ratio	*SE*	*Z*	*p*
V / C	2.27	0.02	76.42	< 0.001
V / H	7.11	0.12	116.07	< 0.001
V / N	4.27	0.06	106.52	< 0.001

C / H	3.13	0.06	62.79	< 0.001
C / N	1.88	0.03	41.55	< 0.001
H / N	0.60	0.01	-25.53	< 0.001

Note: V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all

3.2 Which strategy combinations were beneficial (RQ2)?

Results of the GLMM for RQ2 are displayed in Table 4.

Table 4

Results of GLMM for RQ2

Predictors	Log-Odds	Odds Ratios	*p*
(Intercept)	-0.49 [-2.58 , 1.60]	0.61 [0.08, 4.95]	0.645
V	*Reference*		
NotUseful	-4.25 [-4.61, -3.89]	0.01 [0.01, 0.02]	**<0.001**
V.CA	-1.81 [-2.34, -1.28]	0.16 [0.10, 0.28]	**<0.001**
V.H.CA	-3.39 [-3.83, -2.95]	0.03 [0.02, 0.05]	**<0.001**
V.N	2.24 [1.94, 2.54]	9.42 [6.99, 12.70]	**<0.001**
V.N.CA	0.56 [-0.35, 1.47]	1.75 [0.70, 4.37]	0.227
V.N.H	1.91 [1.29, 2.54]	6.76 [3.62, 12.65]	**<0.001**
V.N.H.CA	-0.97 [-1.63, -0.31]	0.38 [0.20, 0.74]	**0.004**

Random Effects

Variance	Participants	1.57
	Item	6.81
N	Participants	1272
	Item	6
Observations		5970
R^2	Marginal	0.31
	Conditional	0.80

Note: V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all; AIC = 4112; BIC = 4179

If non-useful strategies only were employed (i.e., NOTAT, HOTAT or CA only), odds ratios were below 1, indicating a disadvantage of an application of these strategies compared to VOTAT usage only. Also, the combinations VOTAT + CA (OR = 0.16, $p < 0.001$), VOTAT + HOTAT + CA (OR = 0.03, $p < 0.001$), and VOTAT + NOTAT + HOTAT + CA (OR = 0.38, $p = 0.004$) were not beneficial, compared to VOTAT usage only. In contrast, the combinations VOTAT + NOTAT (OR = 9.42, $p < 0.001$) and VOTAT + NOTAT + HOTAT (OR = 6.76, $p < 0.001$) were beneficial compared to VOTAT usage only. In more detail, this means it was 9.42 times more likely to solve an item when VOTAT + NOTAT was applied and 6.76 more likely to solve an item when VOTAT + NOTAT + HOTAT were used, compared to VOTAT usage only. In addition, it was 1.75 times more likely to solve an item when VOTAT + NOTAT + CA (OR = 1.75, $p = 0.227$) was applied, although this estimate was not significant and should thus be interpreted with caution.

3.3 Interaction of strategy combinations and eigendynamics (RQ3)?

Results of the GLMM for RQ3 are displayed in Table 5.

Table 5

Results of GLMM for RQ3

Predictors	Log-Odds	Odds Ratios	*p*
(Intercept)	-1.22 [-1.81, -0.63]	0.30 [0.16, 0.53]	**<0.001**
V	*Reference*		
NotUseful	-2.88 [-3.55, -2.21]	0.06 [0.03, 0.11]	**<0.001**
V.CA	-1.24 [-2.33, -0.15]	0.29 [0.10, 0.86]	**0.026**
V.H.CA	-1.75 [-2.41, -1.10]	0.17 [0.09, 0.33]	**<0.001**
V.N	2.27 [1.91, 2.62]	9.66 [6.77, 13.80]	**<0.001**
V.N.CA	1.09 [0.19, 1.99]	2.97 [1.21, 7.30]	**0.018**
V.N.H	1.88 [1.18, 2.59]	6.57 [3.24, 13.32]	**<0.001**
V.N.H.CA	0.04 [-0.57, 0.65]	1.04 [0.57, 1.91]	0.897
ED+	-3.79 [-4.39, -3.18]	0.02 [0.01, 0.04]	**<0.001**
NotUseful × ED+	1.92 [1.27, 2.58]	6.85 [3.56, 13.17]	**<0.001**
V.CA × ED+	1.06 [-0.03, 2.14]	2.88 [0.97, 8.52]	0.056

V.H.CA × ED+	2.27 [1.64, 2.90]	9.64 [5.13, 18.13]	**<0.001**
V.N × ED+	2.35 [2.01, 2.70]	10.54 [7.47, 14.87]	**<0.001**
V.N.CA × ED+	2.04 [1.16, 2.92]	7.68 [3.19, 18.51]	**<0.001**
V.N.H × ED+	2.54 [1.84, 3.23]	12.62 [6.30, 25.29]	**<0.001**
V.N.H.CA × ED+	4.09 [3.48, 4.69]	59.59 [32.59, 108.95]	**<0.001**

Random Effects

Variance	Participants	1.75
	Item	0.42
N	Participants	1272
	Item	6
Observations		5970
R²	Marginal	0.66
	Conditional	0.80

Note: V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all; ED+ = presence of eigendynamics; AIC = 3741; BIC = 3862

First of all, the model with eigendynamics reduces random item variance (6.81 vs. 0.42). Also, the model without eigendynamics fitted worse, compared to the model with eigendynamics ($\chi^2(8) = 387$, p < .001), which was also indicated by AIC (4112 for the model without eigendynamics vs. 3741 for the model with eigendynamics) and BIC (4179 vs. 3862) values.

Results of this model with eigendynamics again revealed that the strategy combinations VOTAT + NOTAT (OR = 2.27, $p < 0.001$), VOTAT + NOTAT + CA (OR = 1.09, $p = 0.018$), VOTAT + NOTAT + HOTAT (OR = 1.88, $p < 0.001$) were beneficial, compared to VOTAT application only. In addition, it was less likely to solve an item if the not useful strategies (i.e., NOTAT, HOTAT, CA) only were applied (OR = 0.06, $p > 0.001$), or the combination of VOTAT + CA (OR = 0.29, $p > 0.026$) or VOTAT + HOTAT + CA (OR = 0.17, $p > 0.001$). These results are in line with the results of the model for RQ2. With regard to the direction of effects and significances, the only difference between models was that the strategy combination VOTAT+NOTAT+HOTAT+CA (OR = 0.38, $p = 0.004$) was significantly related to a lower chance to solve an item in the model for RQ2, but non-significant in this model for RQ3 (OR = 1.04, $p = 0.897$). With regard to eigendynamics, it was less likely to solve an item, if eigendynamics were present (OR = 0.02; $p < 0.001$).

The interactions between different strategy combinations and the presence vs. absence of eigendynamics were all significant (p < .001 for all), except for VOTAT + CA and ED (OR = 2.88, $p > 0.056$). For the significant interactions, this means that the OR of strategy combinations changes, depending on the presence or absence of eigendynamics. To evaluate this in more detail, we computed estimated marginal means for strategy combinations while holding eigendynamics constant on "present" and "absent". Results of these analyses are displayed in Figure 3. Corresponding pairwise comparisons (with Bonferroni-Holm adjustment) between factor levels of eigendynamics (present vs. absent) for each strategy combinations are displayed in Table 6. These pairwise comparisons revealed that all strategy combinations significantly differed between the presence vs. absence of eigendynamics (all $p < 0.001$), except for VOTAT+NOTAT+HOTAT+CA (Z = -0.80, $p = 0.42$).

Figure 3

Estimated marginal means to inspect the interaction between strategy combinations and eigendynamics
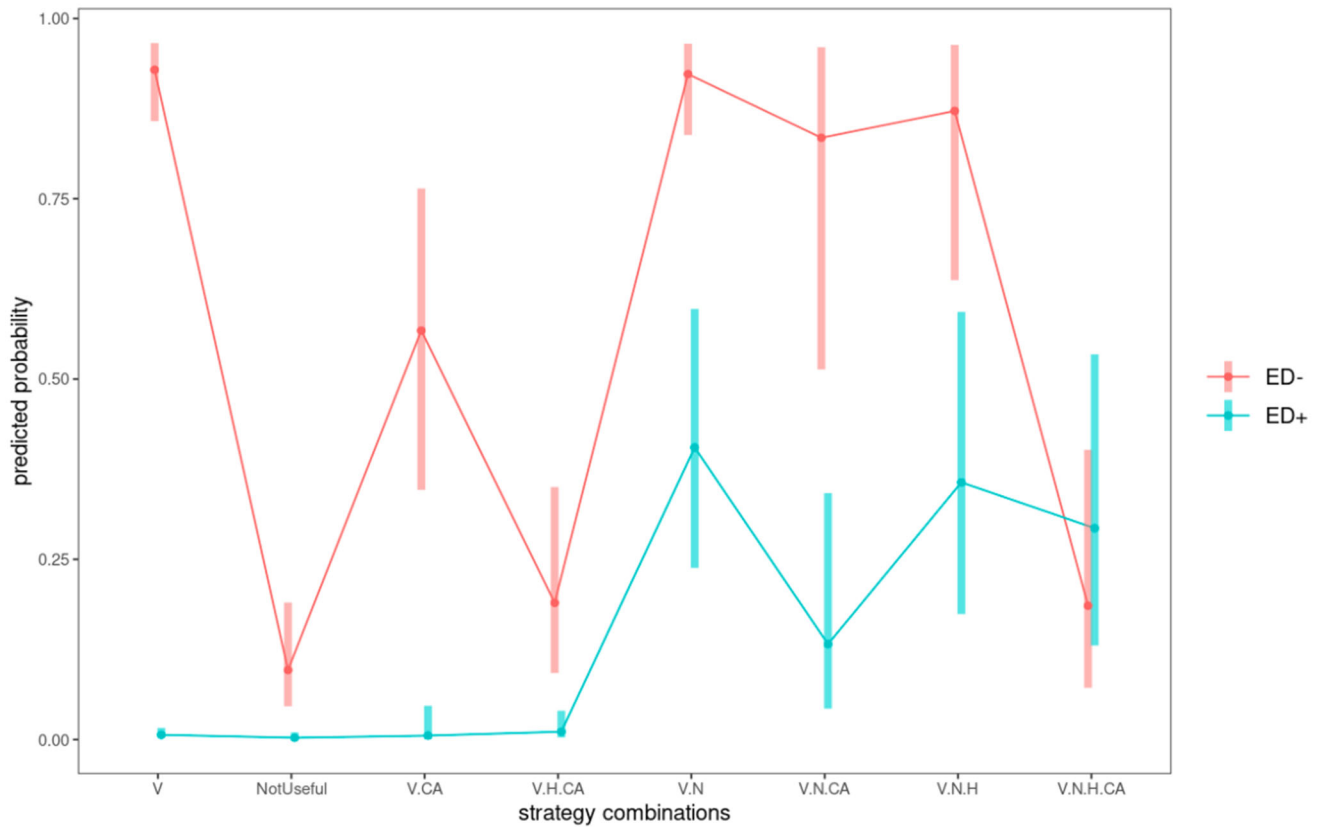
Note. V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all; ED+ = eigendynamics are present; ED- = eigendynamics are not present; error bars indicate 95% confidence interval.

Table 6

Predicted probabilities of strategy combinations when eigendynamics are present vs. absent, and corresponding post hoc comparisons

		predicted probabilties				posthoc comparisons				
strategy combination	ED	pred prob	*SE*	95% CI		estimate	OR	*SE*	*z*	*p*
V	ED-	0.93	0.03	0.86	0.97					
	ED+	0.01	0.00	0.00	0.02	7.58	1952.24	0.62	12.29	**<0.001**
NotUseful	ED-	0.10	0.04	0.05	0.19					
	ED+	0.00	0.00	0.00	0.01	3.73	41.66	0.80	4.67	**<0.001**
V.CA	ED-	0.57	0.11	0.35	0.76					
	ED+	0.01	0.01	0.00	0.05	5.46	235.86	1.20	4.54	**<0.001**
V.H.CA	ED-	0.19	0.07	0.09	0.35					
	ED+	0.01	0.01	0.00	0.04	3.04	20.99	0.78	3.89	**<0.001**
V.N	ED-	0.92	0.03	0.84	0.97					
	ED+	0.41	0.10	0.24	0.60	2.87	17.59	0.58	4.97	**<0.001**
V.N.CA	ED-	0.83	0.11	0.51	0.96					
	ED+	0.13	0.07	0.04	0.34	3.50	33.06	1.00	3.48	**<0.001**
V.N.H	ED-	0.87	0.08	0.64	0.96					
	ED+	0.36	0.11	0.17	0.59	2.51	12.26	0.85	2.97	**0.01**
V.N.H.CA	ED-	0.19	0.08	0.07	0.40					
	ED+	0.29	0.11	0.13	0.53	-0.60	0.55	0.75	-0.80	0.42

Note. V = VOTAT; H = HOTAT; N = NOTAT; CA = Change all; ED+ = eigendynamics are present; ED- = eigendynamics are not present.

4. Discussion

The present article sought to uncover the role of different predictors for CPS performance. As such, we investigated the individual and combined application of all possible strategies on multiple MicroDYN CPS tasks with varying characteristics. Furthermore, the role of eigendynamics was analyzed as additional factor regarding strategy application in determining CPS performance. The results of our study allow for drawing fine-grained inferences concerning which strategy or combination thereof leads to successful vs. unsuccessful CPS performance. Overall, several noteworthy strategy combinations were statistically significant for CPS success. In addition, the results investigating patterns of strategy use across multiple items with vs. without eigendynamics had a number of important implications, which will be discussed below.

4.1 Strategy Usage Across All Tasks

With regard to the general application rates of each strategy across all six tasks, VOTAT was used the most. Given that this strategy is generally associated with successful performance in CPS (e.g., Wüstenberg et al., 2014), it does not come as a surprise that students mostly relied on VOTAT for the purpose of detecting the underlying variable relationships within a particular CPS task. In addition, NOTAT was used fewer times than VOTAT and CA, but more often than HOTAT. Therefore, the results regarding the four possible strategies for solving a complex problem in the MicroDYN assessment approach are in line with those already reported in previous studies (Greiff et al, 2016; Schoppek & Fischer, 2017). Moreover, these outcomes fully

support our initial hypothesis of VOTAT being the most frequently applied strategy, followed by CA, NOTAT, and HOTAT.

Notably, HOTAT had a comparatively low relative frequency compared to VOTAT, NOTAT, and CA. This indicates that students understood that HOTAT application does not bear any unique benefit within the MicroDYN paradigm, particularly since MicroDYN tasks do not feature synergies of two input variables having a unique impact on an output variable.

4.2 Individual and Joint Strategy Contributions to CPS Performance

Generally, our results indicate and confirm that applying VOTAT represents a cornerstone for CPS success. Furthermore, in line with existing research, exclusively applying other strategies was associated with a significantly lower probability of CPS success. Likewise, the drawbacks of using either HOTAT or CA alone have been shown in previous studies (Tschirgi, 1980, see also Lotz et al., 2017). Also, using VOTAT together with NOTAT or HOTAT, or with NOTAT and HOTAT significantly increased the chances of successfully solving a CPS task, compared to only applying VOTAT.

4.3 The Role of Eigendynamics in Strategy Application for CPS Success

Additionally including the separation between CPS tasks with vs. without eigendynamics revealed an even more fine-grained pattern of useful vs. disadvantageous strategy combinations in CPS. In general, students were less likely to solve items containing an eigendynamic correctly, as compared to ones without eigendynamic. Moreover, simultaneously applying VOTAT and NOTAT or VOTAT, NOTAT, and HOTAT or VOTAT, NOTAT, HOTAT, and CA revealed the highest predicted probabilities for solving a given CPS item successfully. Given that VOTAT, HOTAT, and CA application alone are all insufficient means to detect an eigendynamic, this

effect is presumably largely due to the uniquely beneficial nature of NOTAT, which represents the sole strategy for uncovering an eigendynamic without possible confounds. This notion was also mirrored in the significantly lower predicted probabilities for successful CPS performance when students applied strategy combinations without NOTAT in items with eigendynamics.

4.4 The Importance of Strategy Flexibility for CPS success – Indications of the Relevance of Metacognitive Factors

While VOTAT represents a well-known and important strategy for CPS success, the present study has shown that its application alone is insufficient for overarching CPS ability, which spans across a multitude of different scenarios, variables, and problem spaces. Therefore, the simultaneous application of useful supplementary strategies fosters the probability of successful CPS performance. Notably, the results from Figure 3 indicate that VOTAT application alone, whereas technically sufficient for solving CPS items without eigendynamics, did not lead to flawless CPS performance in said items. Similarly, students were largely unable to solve the CPS tasks with eigendynamics despite applying the required strategies, as for instance, VOTAT and NOTAT. Thus, it appears that not only the ability to apply a given strategy, but also the ability to derive meaningful inferences about the variable relationships existing in a given problem environment as a result of one's strategy application, seems to be crucial for CPS success (i.e., being able to integrate the results of a given strategy into and adapt one's mental model of the problem space accordingly; Barrett et al., 2013; Dellaert et al., 2017; Funke, 2012; Halasz, & Moran, 1983). Hence, while students may be able to apply a given strategy in a matter of "trial-and-error" or "simply following instructions", the metacognitive ability to translate their variable manipulations into a comprehensive problem representation

becomes paramount for subsequently not only being able to solve a given, but also other complex problems successfully (Krems, 2014; Osman, 2010; Wüstenberg et al., 2014). We will further elaborate on this aspect in the subsequent paragraph.

4.5 Main Implications and Limitations

In general, our analyses show that how strategies interact to influence CPS success is complex, and that the application of one strategy alone, even VOTAT, does not inherently guarantee that a given CPS task will be solved successfully. However, extensively relying on other strategies, such as HOTAT and CA, appears to decrease the chances of CPS success. In addition, tasks containing eigendynamics are more difficult to solve than the ones with similar overall characteristics not possessing this additional feature. With regard to the interactions of different strategy combinations and their respective role for CPS performance, it thus appears that metacognitive competencies, such as flexible strategy application and adaptation to task demands, seem crucial for achieving CPS success, rather than blindly following a single strategy (Alexander et al., 2017; Bogard et al., 2013; Dahlberg et al., 2019; Kuhn, 2000; Scherer & Tiemann, 2012, 2014; Wüstenberg et al., 2014).

On a broader level, our results carry several implications for the use of digital technologies in the educational realm. For instance, schools are increasingly relying on computer-based learning simulations to equip their students with relevant skills for educational success and beyond (Sánchez-Pérez et al., 2018; Tsarava et al., 2017). Ideally, when implementing such simulations to foster CPS and other domain-general skills (Gobert et al., 2012; Lapek, 2017), as well as skills in STEM domains such as mathematics, chemistry, medicine, or biology (Aurah et al., 2014; Lavi et al., 2019; Su et al., 2016; Wang et al., 2023),

educational institutions should not limit their focus to individual strategies students can apply to achieve success. Instead, students' metacognitive competencies, such as flexibility and adaptability in strategy application based on potentially changing task demands (e.g., Gurbin, 2015), should become a hallmark of computer-based learning simulations for educational purposes. In this regard, three potentially promising metacognitive competencies that have been shown to be associated with successful performance in CPS and beyond are planning, monitoring, and reflecting (De Jong, 2006; McLoughlin & Hollingworth, 2002; Reusser, 1993). Previous studies have already demonstrated the importance of deliberately planning one's next variable manipulation (e.g., Eichmann et al., 2019), monitoring one's progress (e.g., Rudolph et al., 2017), and reflecting on one's CPS performance (e.g., Kauffmann et al., 2008). Likewise, these processes have been discovered to be beneficial in fields adjacent to CPS, such as scientific inquiry (e.g., Sao Pedro et al., 2013), where similar strategies (i.e., VOTAT) are relevant (e.g., Chen & Klahr, 1999). Importantly, these metacognitive facets have also been shown to be important for education more broadly, specifically in the natural science domain (e.g., Avargil et al., 2018), and even general educational success (Cromley & Kunze, 2020; Zohar & Barzilai, 2013).

Importantly, given the importance of metacognition for academic performance (Ohtani & Hisasaka, 2018), existing research has called for its greater facilitation in educational contexts (e.g., Cornoldi, 2010). Thus, specifically addressing metacognitive competencies in CPS training programs would advance not only students' domain-general CPS skills but also their individual toolboxes for achieving educational success across domains.

However, fostering such metacognitive competencies in future CPS training programs should not replace, but rather complement instruction on the application of specific strategies. In this regard, we would like to advocate for the facilitation of two additional strategies apart from VOTAT.

First, in light of the comparatively lower success rates in tasks with eigendynamics, NOTAT should be trained explicitly in such programs (Grežo & Sarmány-Schuller, 2021; Schoppek & Fischer, 2017). The relevance of NOTAT becomes particularly apparent when considering that in real-world scenarios, its application can lead to both beneficial (e.g., a wound healing by itself without the side effects associated with taking a drug) and detrimental outcomes (e.g., keeping the amount of greenhouse gases constant in the face of climate change). Thus, students should be taught the benefits of NOTAT application, ideally by means of computer-based CPS simulations that closely resemble complex real-world problems. This can additionally facilitate students' understanding of why certain variables contain eigendynamics and others do not, for instance by employing assessment approaches such as Tailorshop, which uses 'real-world variables' including 'production', 'demand', and 'worker satisfaction' (e.g., Danner et al., 2011). It might be easier for students to infer that 'demand' fluctuates naturally than to understand why 'Diastolic Blood Pressure' increases autonomously over time, while 'Headache' does not decrease automatically over time in MicroDYN (see Figure 2).

Second, we would like to emphasize the relevance of incorporating an updated version of a strategy resembling HOTAT in MicroDYN in upcoming CPS training programs. While this idea may seem counterintuitive at first, we have seen that, in a CPS task with three input variables, HOTAT essentially involves simultaneously manipulating two input variables in order

to detect potential interaction effects. Therefore, in this case, the equivalent of applying HOTAT would be to apply a strategy called 'vary-multiple-things-at-a-time' (MUTAT). Two (or even more) input variables exerting a joint synergetic effect on one or multiple output variables is a common feature of real-world scenarios. For instance, sowing seed and subsequently adding both water and fertilizer will increase plant growth to a considerably larger extent than just adding either one, and exposure to sunlight may further multiply this effect. Therefore, interaction effects among input variables should become a part of future CPS training programs. Although it would be theoretically possible to incorporate such interaction effects requiring students to apply MUTAT (as HOTAT), they have not yet been implemented in MicroDYN, which may be why the students in our study used HOTAT seldom despite the importance of investigating potential interaction effects between multiple input variables (Funke et al, 2018; Zille et al., 2017).

Thus, while MicroDYN can be considered largely suitable for teaching the underlying strategies and mechanisms of successful CPS under 'clean' conditions (e.g., Schoppek & Fischer, 2017), it should be complemented by additional assessment approaches that a) feature a higher number of input variables, and b) avoid arbitrary labels, such as Tailorshop or related approaches (Danner et al., 2011; Stadler et al., 2015). In addition, we simultaneously advocate for the inclusion of interaction effects between input variables in MicroDYN in order to better highlight such interaction effects. We also recommend extending the set of possible strategies to include MUTAT, which can be implemented in tasks containing as few as two input variables, in order to evaluate its anticipated beneficial effect.

Moreover, we would like to highlight the implications of our results for the relationship between CPS and intelligence in educational settings. The fine-grained investigation of strategy

application as it relates to CPS performance lies at the heart of the present article, and we have

discussed our findings primarily in relation to their implications for future CPS training

programs in educational contexts. However, CPS, intelligence, and education are closely linked

(Greiff et al., 2013; Greiff & Neubert, 2014; Mayer, 2000; Stadler et al., 2015), and it is an

overarching goal of education to shape and foster students' intellectual abilities so that they are

able to successfully deal with the multi-faceted and dynamic demands of 21st-century life

(Martinez, 2000; Ritchie & Tucker-Drob, 2018). As stated by Martinez (2000), intelligence can

be defined as "the knowledge, skills, and strategies necessary to be effective in a world that is

complex and information-rich" (p. 1), thereby building a conceptual bridge to the underlying

facets of CPS. In this vein, Stadler and colleagues (2015) demonstrated in a meta-analysis that

CPS substantially correlates with intelligence. Recently, Grežo and Sarmány-Schuller (2021)

further discovered that the ability to acquire knowledge serves as a partial mediator of the

relationship between intelligence and CPS performance. Taken together, our study's outcomes

have specific implications pertaining to the interplay of CPS and intelligence in education. As

such, our results based on students' overall performance on MicroDYN show that, at the age of

about fifteen, many students have yet to become proficient complex problem solvers. Thus, CPS

training programs in educational settings are needed to foster students' awareness and

application of relevant strategies, including VOTAT, NOTAT, and MUTAT, at particular points

in time in order to successfully navigate complex environments.

    In closing, we would like to address three limitations of our study. First, we only used

one specific kind of microworld to assess CPS. As outlined previously, other microworlds exist

that rely on a higher level of complexity and uncertainty compared to the MicroDYN approach

(Funke, 2003; Stadler, 2015). However, whereas MicroDYN has been criticized for being a pure

VOTAT test (e.g., Dörner & Funke, 2017), this study clearly demonstrates that there is more going on with regard to strategy application even in CPS tasks of comparatively low complexity such as MicroDYN. Still, given MicroDYN's aforementioned shortcomings in adequately representing all possible strategies for uncovering the underlying variable relationships, particularly the underrepresentation of eigendynamics and input variable interaction effects, future studies should take an expanded view of strategy application in CPS by utilizing an updated version of MicroDYN and/or additional or different CPS assessment approaches. Second, this study only analyzed the knowledge acquisition phase, neglecting the knowledge application phase. However, given that the knowledge application phase requires students to solve a given MicroDYN task in no more than four steps, this phase leaves comparatively little room for a rich number of strategy profiles and their respective influences on CPS success. Third, while we thoroughly investigated the absolute application rates of each strategy, we were unable to make inferences about the individual strategy sequences students pursued. Therefore, future research is needed to achieve a more profound understanding of not only how many times students apply each strategy or which strategy combinations were applied, but also at which point during interaction with a complex problem they decide to stick with one or switch to another strategy.

5. Conclusion

Today, an increasing number of training programs in educational institutions are delivered in the form of computer-based learning simulations targeting both domain-specific as well as domain-general skills (Eckhardt et al., 2013; Pedaste & Sarapuu, 2006). Due to their indisputable educational relevance, domain-general skills such as CPS are particularly useful

showcases for how students approach a given task and which strategy application patterns and task characteristics influence successful performance (Gnaldi et al., 2020). In this study, we identified which combinations of CPS strategies are associated with an increased vs. decreased probability of successful CPS performance. In addition, based on our study's outcomes, we argued for the relevance of metacognitive aspects for CPS success. Consequently, these mechanisms should be investigated more thoroughly in upcoming research and should be taken into account in future CPS training paradigms in educational settings rather than merely focusing on strategy usage and repetitive task completion, as was the case in several previous trainings (e.g., Kretzschmar & Süß, 2015). Importantly, training on metacognitive aspects and individual strategies such as VOTAT alike can not only pave the way to successfully solving complex problems, but also increase students' chances of success within and beyond the educational realm. Therefore, computer-based training and assessment programs in schools should incorporate metacognitive aspects, including but not limited to planning, monitoring, and reflecting, alongside showing students how to extend their knowledge by means of fine-grained, yet universally useful strategies such as VOTAT, NOTAT, and MUTAT. By taking into account the relevance of both individual strategies and general metacognitive aspects, computer-based educational CPS training programs can help students become better complex problem solvers and better overall learners.

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In *Assessment and teaching of 21st century skills* (pp. 115-132). Springer, Dordrecht.

Amar, S., & David, N. B. (2016). Realistic intelligence and 21st century skills in adapted learning environment. *American Journal of Educational Research*, *4*(8), 588-596.

Ananiadou, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. OECD Education Working Papers, No. 41, OECD Publishing.

Aurah, C., Cassady, J. & McConnell, T. (2014). Predicting problem solving ability from metacognition and self-efficacy beliefs on a cross validated sample. *British Journal of Education, (2)*1, 49-72.

Avargil, S., Lavi, R., & Dori, Y. J. (2018). Students' metacognition and metacognitive strategies in science education. In *Cognition, Metacognition, and Culture in STEM Education* (pp. 33-64). Springer, Cham.

Baker, E. L., & O'Neil Jr, H. F. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior*, *18*(6), 609-622. https://doi.org/10.1016/S0747-5632(02)00019-5

Baker, R. S., & Siemens, G. (2015). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2 ed., pp. 253–272). Cambridge, UK: Cambridge University Press.

Barrett, J. D., Peterson, D. R., Hester, K. S., Robledo, I. C., Day, E. A., Hougen, D. P., & Mumford, M. D. (2013). Thinking About Applications: Effects on Mental Models and Creative Problem-Solving. *Creativity Research Journal*, *25*, 199–212. https://doi.org/10.1080/10400419.2013.783758

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beckmann, J. F., & Goode, N. (2017). Missing the wood for the wrong trees: On the difficulty of defining the complexity of complex problem solving scenarios. *Journal of Intelligence*, *5*(2), 15. https://doi.org/10.3390/jintelligence5020015

Beckmann, J. F., & Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. In: P.A. Frensch & J. Funke (Eds.), *Complex problem solving: the European perspective* (pp. 177-200). Lawrence Erlbaum.

Beckmann, J. F., Birney, D. P., & Goode, N. (2017). Beyond psychometrics: the difference between difficult problem solving and complex problem solving. *Frontiers in psychology*, *8*, 1739. https://doi.org/10.3389/fpsyg.2017.01739

Bogard, T., Liu, M., & Chiang, Y. H. V. (2013). Thresholds of knowledge development in complex problem solving: A multiple-case study of advanced learners' cognitive processes. *Educational Technology Research and Development*, *61*(3), 465-503. https://doi.org/10.1007/s11423-013-9295-4

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.

    44-75). Lawrence Erlbaum.

Buitrago Flórez, F., Casallas, R., Hernández, M., Reyes, A., Restrepo, S., & Danies, G. (2017).

    Changing a generation's way of thinking: Teaching computational thinking through

    programming. *Review of Educational Research*, *87*(4), 834-860.

    https://doi-org.proxy.bnl.lu/10.3102/0034654317710096

Cornoldi, C. (2010). Metacognition, intelligence, and academic performance. In H. S. Waters &

    W. Schneider (Eds.), *Metacognition, Strategy Use, and Instruction* (pp. 257–277). The

    Guildford Press.

Cromley, J. G., & Kunze, A. J. (2020). Metacognition in education: Translational research.

    Translational Issues in Psychological Science, 6(1), 15–20.

    https://doi.org/10.1037/tps0000218

Curran, V., Gustafson, D. L., Simmons, K., Lannon, H., Wang, C., Garmsiri, M., ... & Wetsch,

    L. (2019). Adult learners' perceptions of self-directed learning and digital technology

    usage in continuing professional education: An update for the digital age. *Journal of*

    Adult and Continuing Education, 25(1), 74-93.

    https://doi.org/10.1177/1477971419827318

Dahlberg, C. L., Lee, S. R., Wiggins, B. L., Leaf, D. S., Lily, L. S., Jordt, H., & Johnson, T. J.

    (2019). A Short, Course-Based Research Module Provides Metacognitive Benefits in the

    Form of More Sophisticated Problem Solving. *Journal of College Science Teaching*,

    48(4), 22-30.

Dai S, Wang X, Svetina D (2022). subscore: Computing Subscores in Classical Test Theory and

Item Response Theory. R package version 3.3,

<https://CRAN.R-project.org/package=subscore>.

Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J.

(2011). Measuring performance on dynamic decision making. *Journal of Individual*

Differences, 32, 225-233. https://doi.org/10.1027/1614-0001/a000055

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational

achievement. *Intelligence*, *35*(1), 13-21. https://doi.org/10.1016/j.intell.2006.02.001

Dellaert, B. G., Arentze, T., Horeni, O., & Timmermans, H. J. (2017). Deriving attribute utilities

from mental representations of complex decisions. *Journal of choice modelling*, *22*, 24-

38. https://doi.org/10.1016/j.jocm.2016.12.001

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation &*

Gaming, *11*(1), 87-106. https://doi.org/10.1177/104687818001100108

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

Frontiers in Psychology, 8, 1–11. https://doi.org/10.3389/fpsyg.2017.01153

Dörner, D., Kreuzug, H.W., Reither, F., & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit*

Komplexität. [Lohhausen: On handling complexity]. Bern: Huber

Eckhardt, M., Urhahne, D., Conrad, O., & Harms, U. (2013). How effective is instructional

support for learning with computer simulations?. *Instructional Science*, *41*(1), 105-124.

https://doi.org/10.1007/s11251-012-9220-y

Eichmann, B., Goldhammer, F., Greiff, S., Pucite, L., & Naumann, J. (2019). The role of

    planning in complex problem solving. *Computers & Education*, *128*, 1-12.

    https://doi.org/10.1016/j.compedu.2018.08.004

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring

    behavioural patterns during complex problem-solving. *Journal of Computer Assisted*

    Learning, 36(6), 933-956. https://doi.org/10.1111/jcal.12451

Eguchi, A. (2016). RoboCupJunior for promoting STEM education, 21st century skills, and

    technological advancement through robotics competition. *Robotics and Autonomous*

    Systems, *75*, 692-699. https://doi.org/10.1016/j.robot.2015.05.013

Eseryel, D., Ge, X., Ifenthaler, D., & Law, V. (2011). Dynamic modeling as cognitive regulation

    scaffold for complex problem solving skill acquisition in an educational massively

    multiplayer online game environment. *Journal of Educational Computing Research,*

    45(3), 265-287. https://doi.org/10.2190/EC.45.3.a

Fiore, S. M., Cuevas, H. M., Scielzo, S., & Salas, E. (2002). Training individuals for distributed

    teams: Problem solving assessment for distributed mission research. *Computers in*

    Human Behavior, *18*(6), 729-744. https://doi.org/10.1016/S0747-5632(02)00027-4

Fischer, A., Greiff, S., & Funke, J. (2012). The Process of Solving Complex Problems. *The*

    Journal of Problem Solving, *4,* 19-42. https://doi.org/10.7771/1932-6246.1118

Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex

    problem solving and experimental results. *Advances in Psychology, 101*, 313-330.

    https://doi.org/10.1016/S0166-4115(08)62663-1

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & reasoning*, *7*(1), 69-89. https://doi.org/10.1080/13546780042000046

Funke, J. (2003). *Problemlösendes Denken [Problem Solving Thinking]*. Stuttgart: Kohlhammer

Funke, J. (2010). Complex problem solving: A case for complex cognition?. *Cognitive processing*, *11*(2), 133-142.

Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: problem solving in the twenty-first century. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and teaching of 21st century skills* (pp. 41-53). Springer, Cham. https://doi.org/10.1007/978-3-319-65368-6_3

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them?. *Applied Measurement in Education*, *29*(4), 245-249. https://doi.org/10.1080/08957347.2016.1209207

Gnaldi, M., Bacci, S., Kunze, T., & Greiff, S. (2020). Students' complex problem solving profiles. *Psychometrika*, *85*(2), 469-501. https://doi.org/10.1007/s11336-020-09709-2

Gobert, J. D., Sao Pedro, M. A., Baker, R. S., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 104-143.

Goode, N., & Beckmann, J. F. (2010). You need to know: There is a causal relationship between structural knowledge and control performance on complex problem solving tasks. *Intelligence*, *38*(3), 345-352. https://doi.org/10.1016/j.intell.2010.01.001

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A

multitrait-multimethod study of assessment instruments for complex problem solving.

Intelligence, 41, 579-596. https://doi.org/10.1016/j.intell.2013.07.012

Greiff, S., Molnár, G., Martin, R., Zimmermann, J., & Csapó, B. (2018). Students' exploration

strategies in computer-simulated complex problem environments: A latent class

approach. *Computers & Education*, *126*, 248-263.

https://doi.org/10.1016/j.compedu.2018.07.013

Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality,

fluid intelligence, and academic achievement. *Learning and Individual Differences*, *36*,

37-48. https://doi.org/10.1016/j.lindif.2014.08.003

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance on

a computer-based assessment of complex problem solving: An analysis of behavioural

data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46.

https://doi.org/10.1016/j.chb.2016.02.095

Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015). Sometimes less is more.

Comparing the validity of complex problem solving measures. *Intelligence, 50,* 100-113.

https://doi.org/10.1016/j.intell.2015.02.007

Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log file analyses as a

window into students' minds?: A showcase study based on the PISA 2012 assessment of

problem solving. *Computers & Education*, *91*, 92–105.

https://doi.org/10.1016/j.compedu.2015.10.018

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin,

    R. (2014). Domain-general problem solving skills and education in the 21st century.

    Educational Research Review, (13), 74-83. https://doi.org/10.1016/j.edurev.2014.10.002

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment

    perspective. *Applied Psychological Measurement*, *36*(3), 189–213.

    https://doi.org/10.1177/0146621612439620

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex

    problem solving in educational contexts—Something beyond g: Concept, assessment,

    measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*,

    364–379. https://doi.org/10.1037/a0031856

Gurbin, T. (2015). Metacognition and technology adoption: exploring influences. *Procedia

    Social and Behavioral Sciences*, *191*, 1576-1582.

    https://doi.org/10.1016/j.sbspro.2015.04.608

Grežo, M., & Sarmány-Schuller, I. (2021). It is Not Enough to be Smart: On Explaining the

    Relation Between Intelligence and Complex Problem Solving. *Technology, Knowledge

    and Learning*, 1-21. https://doi.org/10.1007/s10758-021-09498-2

Halasz, F. G., & Moran, T. P. (1983, December). Mental models and problem solving in using a

    calculator. In *Proceedings of the SIGCHI conference on Human Factors in Computing

    Systems* (pp. 212-216). ACM.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can

    humans process?. *Psychological Science*, *16*(1), 70-76. https://doi.org/10.1111/j.0956-

    7976.2005.00782.x

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count

    data in ecology and evolution. PeerJ, 2, e616.

Ifenthaler, D., Greiff, S., & Gibson, D. C. (2018). Making use of data for assessments:

    harnessing analytics and data science. In J. Voogt, G. Knezek, R. Christensen, & K.-W.

    Lai (Eds.), *International handbook of IT in primary and secondary education* (2 ed., pp.

    649–663). New York, NY: Springer.

JASP Team (2020). JASP (Version 0.14.1) [Computer software].

Jonassen, D. H. (2011). *Learning to solve problems. A handbook for designing problem-solving*

    learning environments. New York: Routledge.

Kauffman, D. F., Ge, X., Xie, K., & Chen, C.-H. (2008). Prompting in Web-Based

    Environments: Supporting Self-Monitoring and Problem Solving Skills in College

    Students. *Journal of Educational Computing Research*, *38*, 115–137.

    https://doi.org/10.2190/EC.38.2.a.

Krems, J. F. (2014). Cognitive flexibility and complex problem solving. In P.A. Frensch & J.

    Funke (Eds.), *Complex problem solving: the European perspective* (pp. 206-223).

    Hillsdale, NJ: Lawrence Erlbaum.

Kretzschmar, A., Neubert, J. C., Wüstenberg, S., & Greiff, S. (2016). Construct validity of

complex problem solving. A comprehensive view on different facets of intelligence and

school grades. *Intelligence, 54,* 55-69. https://doi.org/10.1016/j.intell.2015.11.004

Kretzschmar, A., & Süß, H.-M. (2015). A Study on the Training of Complex Problem Solving

Competence. *Journal of Dynamic Decision Making, 1,* 1-14.

https://doi.org/10.11588/jddm.2015.1.15455

Krkovic, K., Greiff, S., Kupiainen, S., Vainikainen, M.-P., & Hautamäki, J. (2014). Teacher

evaluation of student ability. What roles do teacher gender, student gender, and their

interaction play? *Educational Research*, 56, 243-256.

http://dx.doi.org/10.1080/00131881.2014.898909

Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science*,

9(5), 178–181.

Kuhn, D., & Dean Jr, D. (2005). Is developing scientific thinking all about learning to control

variables?. *Psychological Science*, *16*(11), 866-870. https://doi.org/10.1111/j.1467-

9280.2005.01628.x

Lapek, J. (2017). 21st century skills: the tools students need. *Children's Technology &

Engineering*, *21*(3), 24-26.

Lavi, R., Shwartz, G., & Dori, Y. J. (2019). Metacognition in chemistry education: a literature

review. *Israel Journal of Chemistry*, *59*(6-7), 583-597.

https://doi.org/10.1002/ijch.201800087

Lee, W., & Grimm, K. J. (2018). Generalized linear mixed-effects modeling programs in R for

    binary outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(5), 824-

    828. https://doi.org/10.1080/10705511.2018.1500141

Lenth, R. (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package

    version 1.8.3, <https://CRAN.R-project.org/package=emmeans>.

Lin, L., Mills, L., & Ifenthaler, D. (2016). Collaboration, multi-tasking and problem solving

    performance on shared virtual spaces. *Journal of Computing in Higher Education, 28*(3),

    344–357. https://doi.org/10.1007/s12528-016-9117-x

Liu, H., Liu, Y., & Li, M. (2018). Analysis of process data of PISA 2012 computer-based

    problem solving: application of the modified Multilevel Mixture IRT model. *Frontiers in

    psychology*, *9*, 1372. https://doi.org/10.3389/fpsyg.2018.01372

Long, J.A. (2022). _tools: Analysis and Presentation of Social Scientific Data. R package version

    2.2.0, <https://cran.r-project.org/package=jtools>.

Lotz, C., Scherer, R., Greiff, S., & Sparfeldt, J. R. (2017). Intelligence in action – Effective

    strategic behaviors while solving complex problems. *Intelligence*, *64*, 98–112.

    https://doi.org/10.1016/j.intell.2017.08.002

Lotz, C., Sparfeldt, J. R., & Greiff, S. (2016). Complex problem solving in educational contexts

    Still something beyond a "good g"?. *Intelligence*, *59*, 127-138.

    https://doi.org/10.1016/j.intell.2016.09.001

Lüdecke, Patil, Ben-Shachar, Wiernik, & Makowski (2022). easystats: Framework for Easy

    Statistical Modeling, Visualization, and Reporting. CRAN. Available from

    https://easystats.github.io/easystats/

Mainert, J., Kretzschmar, A., Neubert, J. C., & Greiff, S. (2015). Linking complex problem

    solving and general mental ability to career advancement: Does a transversal skill reveal

    incremental predictive validity?. *International Journal of Lifelong Education*, *34*(4), 393–

    411. https://doi.org/10.1080/02601370.2015.1060024

Mainert, J., Niepel, C., Murphy, K. R., & Greiff, S. (2019). The incremental contribution of

    complex problem-solving skills to the prediction of job level, job complexity, and salary.

    Journal of Business and Psychology, *34*(6), 825-845.

    https://doi.org/10.1007/s10869-018-9561-x

Marcolin, L., Miroudot, S., & Squicciarini, M. (2019). To be (routine) or not to be (routine), that

    is the question: a cross-country task-based answer. *Industrial and Corporate Change*,

    28(3), 477-501. https://doi.org/10.1093/icc/dty020

Martinez, M. E. (2000). *Education as the cultivation of intelligence.* Lawrence Erlbaum

    Associates Publishers.

Mayer, R. E. (2000). Intelligence and education. In R. J. Sternberg (Ed.), *Handbook of*

    intelligence (pp. 519–533). Cambridge University Press.

    https://doi.org/10.1017/CBO9780511807947.024

Mayer, R.E., & Wittrock, M. C. (2006). Problem solving. In P.A. Alexander & P.H. Winne

(Eds.), *Handbook of educational psychology*, (2nd ed., pp. 287–304). Mahwah, NJ:

Lawrence Erlbaum.

Mayrath, M. C., Clarke-Midura, J., Robinson, D. H., & Schraw, G. (Eds.). (2012).

*Technology-Based Assessments for 21st Century Skills: Theoretical and Practical

Implications from Modern Research*. Charlotte, NC, US: Information Age Publishing.

McLoughlin, C., & Hollingworth, R. (2002). *Bridge over troubled water: Creating effective

online support for the metacognitive aspects of problem solving* (pp. 1267-1272).

Association for the Advancement of Computing in Education (AACE).

Meza J., Monguet J.M., Grimón F., Trejo A. (2016). Fostering Collective Intelligence Education.

In G. Vincenti, A. Bucciero, & C. Vaz de Carvalho (Eds.), *E-Learning, E-Education, and

Online Training (eLEOT 2015). Lecture Notes of the Institute for Computer Sciences,

Social Informatics and Telecommunications Engineering, vol. 160*. Springer, Cham.

https://doi.org/10.1007/978-3-319-28883-3_21

Molnár, G., & Csapó, B. (2018). The efficacy and development of students' problem-solving

strategies during compulsory schooling: Logfile analyses. *Frontiers in psychology*, *9*,

302. https://doi.org/10.3389/fpsyg.2018.00302

Murphy S., MacDonald A., & Danaia L. (2020). Sustaining STEM: A Framework for Effective

STEM Education Across the Learning Continuum. In A. MacDonald, L. Danaia, & S.

Murphy (Eds.), *STEM Education Across the Learning Continuum*. Springer, Singapore.

https://doi.org/10.1007/978-981-15-2821-7_2

Mustafić, M., Yu, J., Stadler, M., Vainikainen, M.-P., Bornstein, M. H., Putnick, D. L., & Greiff,

      S. (2019). Complex problem solving: Profiles and developmental paths revealed via

      latent transition analysis. *Developmental Psychology, 55*(10), 2090–2101.

      https://doi.org/10.1037/dev0000764

Novalinda, R., Giatman, M., Syahril, Ambiyar, & Fajra, M. (2020). Problem-Based Learning:

      21st Century Vocational Education. *International Journal Of MultiScience*, *1*(8), 12-19.

OECD (2014). *PISA 2012 results: Creative problem solving: Students' skills in tackling real-life*

      problems (Volume V). Paris, France: OECD Publishing.

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: a meta-analytic review of the

      relationship among metacognition, intelligence, and academic performance.

      *Metacognition & Learning, 13***,** 179–212. https://doi.org/10.1007/s11409-018-9183-8

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

Pedaste, M., & Sarapuu, T. (2006). Developing an effective support system for inquiry learning

      in a web-based environment. *Journal of Computer Assisted Learning*, *22*(1), 47-62.

      https://doi.org/10.1111/j.1365-2729.2006.00159.x

Purzer, S., Moore, T. J., & Dringenberg, E. (2018). Engineering cognition: A process of

      knowledge acquisition and application. In *Cognition, Metacognition, and Culture in*

      STEM education (pp. 167-190). Springer, Cham. https://doi.org/10.1007/978-3-319-

      66659-4_8

Reijnders, L. S., & de Vries, G. J. (2018). Technology, offshoring and the rise of non-routine

    jobs. *Journal of Development Economics*, *135*, 412-432.

    https://doi.org/10.1016/j.jdeveco.2018.08.009

Ren, Y., Luo, F., Ren, P., Bai, D., Li, X., & Liu, H. (2019). Exploring Multiple Goals Balancing

    in Complex Problem Solving Based on Log Data. *Frontiers in psychology*, *10*, 1975.

    https://doi.org/10.3389/fpsyg.2019.01975

Resnick, L. B., & Glaser, R. (1976). Problem solving and intelligence. In L. B. Resnick (Ed.),

    The nature of intelligence (pp. 205–230). Lawrence Erlbaum.

Reusser, K. (1993). Tutoring systems and pedagogical theory: Representational tools for

    understanding, planning, and reflection in problem solving. *Computers as cognitive tools*,

    1, 143-177.

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A

    meta-analysis. *Psychological Science*, *29*(8), 1358-1369.

    https://doi.org/10.1177/0956797618774253

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive

    ability. *Intelligence*, *35*(1), 83-92. https://doi.org/10.1016/j.intell.2006.05.004

Rudolph, J., Niepel, C., Greiff, S., Goldhammer, F., & Kröner, S. (2017). Metacognitive

    confidence judgments and their link to complex problem solving. *Intelligence*, *63*, 1–8.

    https://doi.org/10.1016/j.intell.2017.04.005

Sanabria, J. C., & Arámburo-Lizárraga, J. (2017). Enhancing 21st century skills with AR: Using

    the gradual immersion method to develop collaborative creativity. *Eurasia Journal of*

Mathematics, Science and Technology Education, *13*(2), 487-501.

https://doi.org/10.12973/eurasia.2017.00627a

Sánchez-Pérez, N., Castillo, A., López-López, J. A., Pina, V., Puga, J. L., Campoy, G., ... &

Fuentes, L. J. (2018). Computer-based training in math and working memory improves

cognitive skills and academic achievement in primary school children: Behavioral results.

Frontiers in psychology, *8*, 2327. https://doi.org/10.3389/fpsyg.2017.02327

Sao Pedro, M. A., Baker, R. S., Gobert, J. D., Montalvo, O., & Nakama, A. (2013)

Leveraging machine-learned detectors of systematic inquiry behavior to estimate and

predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction*, *23*(1), 1-

39.

Scherer, R., Siddiq, F., & Tondeur, J. (2019). The technology acceptance model (TAM): A meta-

analytic structural equation modeling approach to explaining teachers' adoption of digital

technology in education. *Computers & Education*, *128*, 13-35.

https://doi.org/10.1016/j.compedu.2018.09.009

Scherer, R., & Tiemann, R. (2012). Factors of problem-solving competency in a virtual

chemistry environment: The role of metacognitive knowledge about strategies.

Computers & Education, *59*, 1199–1214. https://doi.org/10.1016/j.compedu.2012.05.020

Scherer, R., & Tiemann, R. (2014). Evidence on the effects of task interactivity and grade level

on thinking skills involved in complex problem solving. *Thinking Skills and Creativity*,

11, 48–64. https://doi.org/10.1016/j.tsc.2013.10.003

Schoppek, W., & Fischer, A. (2017). Common process demands of two complex dynamic

control tasks: transfer is mediated by comprehensive strategies. *Frontiers in psychology*,

8, 2145. https://doi.org/10.3389/fpsyg.2017.02145

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach:

Complex problem solving predicts school grades beyond working memory capacity.

Learning and Individual Differences, *24*, 42–52. DOI:10.1016/j.lindif.2012.12.011

Schwichow, M., Croker, S., Zimmerman, C., Höffler, T., & Härtig, H. (2016). Teaching the

control-of-variables strategy: A meta-analysis. *Developmental Review*, *39*, 37–63.

https://doi.org/10.1016/j.dr.2015.12.001

Selker, R., Love, J., Dropmann, D., & Moreno, V. (2022). jmv: The 'jamovi' Analyses. R package

version 2.3.4, <https://CRAN.R-project.org/package=jmv>.

Selwyn, N. (2012). *Education in a digital world: Global perspectives on technology and

education*. New York, NY, US: Routledge.

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in

elementary and secondary education. *Journal of Computer Assisted Learning*, *33*(1), 1-

19. https://doi.org/10.1111/jcal.12172

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving

skills via stealth assessment in an engaging video game. *Computers in Human Behavior*,

63, 106-117. https://doi.org/10.1016/j.chb.2016.05.047

Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., … Latour, T. (2012).

The Genetics Lab: Acceptance and psychometric characteristics of a computer-based

microworld assessing complex problem solving. *Psychological Test and Assessment Modeling*, *54*(1), 54–72.

Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence*, *41*(5), 289-305. https://doi.org/10.1016/j.intell.2013.05.002

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence*, *53*, 92–101. https://doi.org/10.1016/j.intell.2015.09.005

Stadler, M., Fischer, F., & Greiff, S. (2019). Taking a closer look: An exploratory analysis of successful and unsuccessful strategy use in complex problems. *Frontiers in Psychology*, *10*, 777. https://doi.org/10.3389/fpsyg.2019.00777

Stadler, M., Niepel, C., & Greiff, S. (2016). Easily too difficult: Estimating item difficulty in computer simulated microworlds. *Computers in Human Behavior*, *65*, 100–106. https://doi.org/10.1016/j.chb.2016.08.025

Stadler, M., Niepel, C., & Greiff, S. (2019). Differentiating between static and complex problems: A theoretical framework and its empirical validation. *Intelligence*, *72*, 1-12. https://doi.org/10.1016/j.intell.2018.11.003

Su, H. F. H., Ricci, F. A., & Mnatsakanian, M. (2016). Mathematical teaching strategies: Pathways to critical thinking and metacognition. *International Journal of Research in Education and Science*, *2*(1), 190-200.

Teig, N., Scherer, R., & Kjærnsli, M. (2020). Identifying patterns of students' performance on simulated inquiry tasks using PISA 2015 log-file data. *Journal of Research in Science Teaching*. https://doi.org/10.1002/tea.21657

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills. In B. Csapó and J. Funke (Eds.), *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning*, OECD Publishing, Paris, https://doi.org/10.1787/9789264273955-14-en

Trilling, B., & Fadel, C. (2009). *21st Century Skills, Learning for Life in Our Times*. San Francisco, CA: John Wiley & Sons.

Tsarava, K., Moeller, K., Pinkwart, N., Butz, M., Trautwein, U., & Ninaus, M. (2017, October). Training computational thinking: Game-based unplugged and plugged-in activities in primary school. In *European Conference on Games Based Learning* (pp. 687-695). Academic Conferences International Limited.

Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, *51*, 1-10.

Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: Dynamic assessment of the control of variables strategy. *Instructional Science*, *43*, 381–400. https://doi.org/10.1007/s11251-015-9344-y

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vainikainen, M. P. (2014). *Finnish primary school pupils' performance on learning to learn assessments: A longitudinal perspective on educational equity. Report 360*. Helsinki, Finland: Unigrafia

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science*, *20*, 75–100. https://doi.org/10.1207/s15516709cog2001_3

Wang, C. Y., Chen, S., & Huang, M. Y. (2023). Exploring medical students' metacognitive and regulatory dimensions of diagnostic problem solving. *Medical Education Online*, *28*(1), 2210804. https://doi.org/10.1080/10872981.2023.2210804

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

Witherington, D. C., & Boom, J. (2019). Conceptualizing the dynamics of development in the 21st century: Process, (inter) action, and complexity. *Human Development*, *63*(3-4), 147-152. https://doi.org/10.1159/000504097

Wu, H., & Molnár, G. (2021). Logfile analyses of successful and unsuccessful strategy use in complex problem-solving: a cross-national comparison study. *European Journal of Psychology of Education*, 1-24. https://doi.org/10.1007/s10212-020-00516-y

Wüstenberg, S., Greiff, S. & Funke, J. (2012). Complex problem solving. More than reasoning? *Intelligence, 40*, 1-14. https://doi.org/10.1016/j.intell.2011.11.003

Wüstenberg, S., Stadler, M., Hautamäki, J., & Greiff, S. (2014). The Role of Strategy

    Knowledge for the Application of Strategies in Complex Problem Solving Tasks.

    Technology, Knowledge and Learning, *19*, 127–146.

    https://doi.org/10.1007/s10758-014-9222-8

Xie, Y., Fang, M., & Shauman, K. (2015). STEM education. *Annual Review of Sociology*, *41*,

    331-357. https://doi.org/10.1146/annurev-soc-071312-145659

Xu, H., Fang, G., Chen, Y., Liu, J., & Ying, Z. (2018). Latent class analysis of recurrent events

    in problem-solving items. *Applied psychological measurement*, *42*(6), 478-498.

    https://doi.org/10.1177/0146621617748325

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of

    the Three-Parameter Logistic Model. Applied Psychological Measurement, 8(2),

    125–145.

Zille, H., Ishibuchi, H., Mostaghim, S., & Nojima, Y. (2017). A framework for large-scale

    multiobjective optimization based on problem transformation. *IEEE Transactions on*

    Evolutionary Computation, *22*(2), 260-275. https://doi.org/10.1109/TEVC.2017.2704782

Zohar, A., & Barzilai, S. (2013). A review of research on metacognition in science education:

    Current and future directions. *Studies in Science education*, *49*(2), 121-169.

    https://doi.org/10.1080/03057267.2013.847261