UNIVERSITÄT DES SAARLANDES

FAKULTÄT FÜR EMPIRISCHE HUMANWISSENSCHAFTEN
UND WIRTSCHAFTSWISSENSCHAFT
FACHRICHTUNG BILDUNGSWISSENSCHAFTEN

Shedding Light on the Still Fuzzy Relationship Between Intelligence and Complex Problem Solving: The Role of Process Measures

DISSERTATION
ZUR ERLANGUNG DES AKADEMISCHEN GRADES
EINES DOKTORS DER NATURWISSENSCHAFTEN
DER FAKULTÄT HW DER UNIVERSITÄT DES SAARLANDES

vorgelegt von
M. Sc. Julius Johannes Weise
geb. in München

Saarbrücken, 2024

Dekan:
Univ.-Prof. Dr. Peter Loos

Berichterstatter:
Univ.-Prof. Dr. Jörn R. Sparfeldt, Universität des Saarlandes
Univ.-Prof. Dr. Robin Stark, Universität des Saarlandes

Tag der Disputation: 15.07.2024

Acknowledgements

My utmost gratitude belongs to my supervisor Prof. Dr. Jörn Sparfeldt for his valuable support and his important advice during my time as a doctoral student. Without his tremendous support, guidance, and flexibility regarding my professional career, this dissertation project would hardly have been possible. Thank you, Jörn!

The present dissertation project is based on a dataset whose design is a product of theoretical considerations made by Prof. Dr. Jörn Sparfeldt and Prof. Dr. Samuel Greiff. Furthermore, Prof. Dr. Jörn Sparfeldt and Prof. Dr. Samuel Greiff supervised and supported the collection of the dataset, which was carried out by Dr. Christin Lotz. I am very grateful to Prof. Dr. Jörn Sparfeldt, Prof. Dr. Samuel Greiff, and everyone who was involved in collecting and curating the data for the opportunity to work with this dataset. In addition, I would like to thank Dr. Christin Lotz and her partner Sascha Ludwig for extracting the log-file data from the raw files of the complex problem solving assessment tool, which were used in studies 1 through 3 of this dissertation project. The title of this dissertation was inspired by a study by Prof. Dr. Detlev Leutner (2002) published under the name "The fuzzy relationship of intelligence and problem solving in computer simulations".

I would also like to thank all those who gave me valuable advice and support during my dissertation project, especially Dr. Nicolas Becker and Prof. Dr. Robin Stark. Furthermore, I would like to thank my team and colleagues who have supported me in this dissertation project and with whom I have shared many joyful moments during my time as a Ph.D. student. Thank you, Khanda Adams, Katrin Arendt, Petra Arnold, Annika Blum, Cristina Deeg, Luca Deeg, Stefanie Krämer, Maximilian Krolo, Christin Lotz, Birgit Schneider, Rebecca Schneider, and Sarah Schwabe. In particular, I would like to thank Verena Keimerl, Theresa Krause-Wichmann, Julia Ruby, and Benedikt Schneider for their invaluable friendship and encouragement in challenging times.

Finally, I would like to thank my parents, my sister Klara, my brother Felix, Nattaya, and my close friends for always supporting me and believing in me. Family and friends have been my source of strength to complete this project and have made my life joyful. Thank you all!

Content

Abbreviations

AIC	–	Akaike Information Criterion
BIC	–	Bayesian Information Criterion
BIS	–	Berlin Intelligence Structure test
CFI	–	Comparative Fit Index
CPS	–	Complex Problem Solving
FSE	–	Finite State Automata (Funke, 2001)
g	–	General factor of intelligence
GPA	–	Grade Point Average
HOTAT	–	Strategic CPS exploration behavior to *Hold-One-Thing-At-a-Time*
KPL	–	German: Komplexes Problemlösen
LCA	–	Latent Class Analysis
LSE	–	Linear Structural Equation systems (Funke, 2001)
LTA	–	Latent Transition Analysis
MCS	–	Multiple Complex Systems
MLR	–	Maximum Likelihood Estimation
NOTAT	–	Strategic CPS exploration behavior to *vary NO-Thing-At-a-Time*; also referred to as noninterfering behavior
OECD	–	Organization for Economic Co-operation and Development
PISA	–	Program for International Student Assessment
SCS	–	Single Complex Systems
VOTAT	–	Strategic CPS exploration behavior to *Vary-One-Thing-At-a-time*

Index of Publications

The present thesis is based on four studies either published or submitted in peer-reviewed journals. The published articles are available online through the respective publishers. In the following, I will use the term "we" to refer to the authors of each study.

Study 1

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving – Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502.

Study 2

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2022). Focusing on eigendynamic effects promotes students' performance in complex problem solving: A log-file analysis of strategic behavior. *Computers & Education*, *189*, 104579.

Study 3

Ruby, J., Weise, J. J., Greiff, S., & Sparfeldt, J. R. (submitted). Double-checks or better no repeated steps? The Role of Effective and Efficient Exploration Behaviors for Successful Complex Problem Solving. *Computers & Education.*

Study 4

Weise, J. J., Becker, N., & Sparfeldt, J. R. (2024). Intelligenzmessung in drei Minuten – Evaluation des mini-q und Konstruktion einer Parallelversion [Assessment of intelligence in three minutes – An evaluation of the mini-q and the construction of a parallel form]. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. https://doi.org/10.1024/1010-0652/a000384

Summary

Intelligence and complex problem solving (CPS) are well researched constructs that play an important role in educational science and psychology. The significance of these two constructs is based on their close relationship and significant prediction of success in a variety of important real-life outcomes, such as scholastic achievement, achievement in university, and job performance. Regarding the relationship between intelligence and CPS, the two constructs are closely related at the theoretical level. However, despite extensive research, the empirical relationship between intelligence and CPS has not been conclusively clarified. Therefore, the aim of the present dissertation project was to investigate the factors that influence the strength of the association between the two constructs and to examine observable patterns of behavior in which intelligence manifests itself when solving complex problems. As CPS assessment tools are computer-based and interactive, they allow to extract log-file data of students' problem-solving process from which meaningful patterns of observable behavior can be isolated. Such observable behaviors can be thought of as indicators of cognitive processes and, based on this assumption, can be used to clarify the intelligence-CPS relationship. In the present dissertation project, three studies were conducted to accomplish the aforementioned goal. In a fourth study, a very time-efficient screening instrument to assess an indicator of general intelligence was evaluated and a parallel form was developed, thus expanding the possibilities of intelligence assessment in (large-scale) studies when time is scarce.

In the first study, we investigated the Elshout-Raaheim hypothesis (Leutner, 2002; see Elshout, 1987; Raaheim, 1988), which predicts a curvilinear moderating effect of prior knowledge on the relationship between intelligence and CPS, such that correlation coefficients follow an inverted U-shaped pattern. Across a series of five CPS tasks, we hypothesized that prior knowledge would increase from task to task. To test this hypothesis, we considered the relative frequency of CPS exploration behavior to *vary-one-thing-at-a-time* (VOTAT) in individual CPS tasks as an indicator of prior (strategic) knowledge. The study revealed two main findings: First, the predicted increase in prior knowledge in the knowledge acquisition phase was supported by our data. Second, the pattern of intelligence-CPS correlation coefficients followed an inverted U-shaped pattern in the knowledge acquisition phase and the knowledge application phase of the CPS assessment tool, as predicted. Our findings of an inverted U-shaped pattern of

correlation coefficients suggest that intelligence plays a more important role in solving complex problems when participants have acquired medium levels of prior knowledge, and a less important role when participants have acquired either rather low or rather high levels of prior knowledge.

In the second study, we examined students' strategic exploration behavior with regard to *eigendynamic effects*. Eigendynamic effects are an important characteristic of CPS and can be described as increases or decreases in outcome variables over time without any actions taken by the participant. Specifically, we investigated for the first time the strategic behavior to *identify eigendynamic effects in an early exploration step*, which is an effective strategic behavior if the system comprises different types of effects (i.e., eigendynamic effects along with other effects). The examined strategic behavior predicted the performance in both CPS phases for the respective tasks and for the set of tasks. Moreover, the mediation models showed that intelligence manifests itself in the strategic behavior to identify eigendynamics early, going hand in hand with higher CPS performance in the knowledge acquisition phase. Regarding the knowledge application phase, the pattern of results suggests that more intelligent students achieved higher CPS performance scores by adequately identifying eigendynamics, but regardless of whether they did so in an early or later exploration step.

In the third study, we investigated the effectiveness (i.e., consistent use of VOTAT) and efficiency (i.e., number of non-necessary exploration steps) of students' exploration behavior in CPS tasks and its relation to intelligence and CPS performance. In a first, exploratory analysis of students' exploration behavior, we found an increasingly consistent use of VOTAT across the sequence of tasks, corresponding to previous studies. Going beyond previous studies, we showed an increasingly efficient strategy use across tasks. Using a person-centered approach, we found four distinct (latent) classes of students. With respect to these classes, the pattern of results highlights the importance of effectiveness (indicated by VOTAT) for successful problem solving and its relationship to intelligence. However, classes of students who explored effectively, but either more or less efficiently, showed comparable intelligence and CPS test scores. Thus, a pattern of increasingly efficient exploration behavior across tasks was found, but the relationship of efficiency to CPS performance and intelligence requires further research.

The fourth study focused on the evaluation and extension of a very time-efficient intelligence screening instrument that can be administered in just three minutes, the *mini-q* (Baudson & Preckel, 2016). Time-efficient and well-evaluated intelligence screenings are important instruments in extensive research projects. Specifically, we examined the test characteristics of the *mini-q* and introduced a new parallel version, the *mini-q B*. The distributions of the sum scores of the two test versions indicated discrimination between individuals of different ability and were substantially correlated with each other. With respect to validity, we found correlations with another intelligence test and the GPA that were partially consistent with our assumptions. In summary, the results are promising and indicate the suitability of the *mini-q* and its parallel version as intelligence screening instruments.

Taken together, the results of this dissertation project helped to clarify the ambiguous relationship between intelligence and CPS and to expand the potential applications of a very time-efficient intelligence screening instrument. With regard to the intelligence-CPS relationship, both moderating effects, which influence the strength of the association, and mediating effects, which illustrate how intelligence manifests itself in observable patterns of behavior when solving complex problems, were demonstrated. In addition, the relevance and great potential of log-file data to provide insight into the cognitive processes of problem solvers were demonstrated, helping to clarify the relationship between intelligence and CPS.

Zusammenfassung

Intelligenz und komplexes Problemlösen (KPL) gelten als gut erforschte Konstrukte mit hoher Relevanz für die Bildungswissenschaften und die Psychologie. Die Bedeutung dieser beiden Konstrukte beruht auf ihrem Zusammenhang mit dem Erfolg in zentralen Lebensbereichen wie Schul- und Studienleistungen und beruflichem Erfolg. Auf theoretischer Ebene sind die beiden Konstrukte eng miteinander verknüpft. Trotz zahlreicher Studien ist der empirische Zusammenhang zwischen Intelligenz und KPL jedoch nicht abschließend geklärt. Ziel dieses Dissertationsprojektes war es daher, Faktoren zu untersuchen, die die Stärke des Zusammenhangs zwischen den beiden Konstrukten beeinflussen, und beobachtbare Verhaltensmuster zu identifizieren, in denen sich Intelligenz beim Lösen komplexer Probleme manifestiert. Da KPL-Testinstrumente in der Regel computergestützt und interaktiv sind, ermöglichen sie die Extraktion von Logfile-Daten des Problemlösungsprozesses. Aus diesen Protokolldaten können beobachtbare Verhaltensmuster isoliert werden, die als Indikatoren für kognitive Prozesse interpretiert werden können und somit zur Klärung der Beziehung zwischen Intelligenz und KPL beitragen. Im Rahmen dieses Dissertationsprojektes wurden drei Studien zur Untersuchung des Zusammenhangs zwischen Intelligenz und KPL durchgeführt. In einer vierten Studie wurde ein sehr zeiteffizientes Screening-Instrument zur Erfassung eines Indikators der allgemeinen Intelligenz evaluiert und eine Parallelform entwickelt, welche die Möglichkeiten der Intelligenzmessung in Studien mit begrenztem Zeitrahmen erweitert.

In der ersten Studie untersuchten wir die Elshout-Raaheim-Hypothese (Leutner, 2002; siehe Elshout, 1987; Raaheim, 1988), die einen moderierenden Effekt des Vorwissens auf den Zusammenhang zwischen Intelligenz und KPL in Form eines umgekehrt U-förmigen Musters der Korrelationskoeffizienten vorhersagt. Dabei nahmen wir an, dass über eine Serie von fünf KPL-Aufgaben das Vorwissen von Aufgabe zu Aufgabe zunehmen würde. Um diese Hypothese zu testen, betrachteten wir als Indikator für das (strategische) Vorwissen die relative Häufigkeit des KPL-Explorationsverhaltens, jeweils nur eine Variable zu variieren und alle anderen konstant bei null zu halten (Engl.: *vary-one-thing-at-a-time*; VOTAT) in einzelnen KPL-Aufgaben. Die Analysen zeigten, dass das strategische Vorwissen in der Phase des Wissenserwerbs wie erwartet zunahm. Darüber hinaus folgt das Muster der Korrelationskoeffizienten zwischen Intelligenz und

KPL in der Phase des Wissenserwerbs und in der Phase der Wissensanwendung des KPL-Messinstruments hypothesenkonform einer umgekehrt U-förmigen Kurve. Unsere Ergebnisse eines umgekehrt U-förmigen Musters der Korrelationskoeffizienten deuten darauf hin, dass die Intelligenz eine wichtigere Rolle bei der Lösung komplexer Probleme spielt, wenn die Teilnehmenden ein mittleres Niveau an Vorwissen erworben haben, und eine weniger wichtige Rolle, wenn die Teilnehmenden entweder ein eher niedriges oder ein eher hohes Niveau an Vorwissen erworben haben.

In der zweiten Studie wurde das strategische Explorationsverhalten im Hinblick auf eigendynamische Effekte untersucht. Eigendynamische Effekte sind ein wichtiges Merkmal von KPL und können als Zunahme oder Abnahme von Ergebnisvariablen in Abhängigkeit von der Zeit beschrieben werden. Zum ersten Mal wurde das strategische Verhalten untersucht, eigendynamische Effekte in einem frühen Explorationsschritt zu identifizieren, ein effektives Vorgehen, wenn das System verschiedene Arten von Effekten enthält (d. h. eigendynamische Effekte zusammen mit anderen Effekten). Das untersuchte strategische Verhalten sagte die Leistung in beiden KPL-Phasen für die jeweiligen Aufgaben und für das Set an Aufgaben voraus. Darüber hinaus zeigten die Mediationsmodelle, dass sich Intelligenz in der strategischen Verhaltensweise, eigendynamische Effekte in einem frühen Explorationsschritt zu identifizieren, manifestierte, was mit einer höheren KPL-Leistung in der Wissenserwerbsphase einherging. In Bezug auf die Phase der Wissensanwendung deutete das Ergebnismuster darauf hin, dass intelligentere Schülerinnen und Schüler höhere KPL-Leistungen erzielten, indem sie Eigendynamik adäquat identifizierten, aber unabhängig davon, ob sie dies zu einem früheren oder einem späteren Zeitpunkt taten.

In der dritten Studie untersuchten wir die Effektivität (d. h. die konsistente Verwendung von VOTAT für jede Eingabevariable) und die Effizienz (d. h. die Anzahl unnötiger Explorationsschritte über die einmalige Verwendung von VOTAT für jede Eingabevariable hinaus) des Explorationsverhaltens bei KPL-Aufgaben und deren Beziehung zur Intelligenz und KPL-Leistung. In einer ersten explorativen Analyse des Explorationsverhaltens fanden wir eine zunehmend konsistente Nutzung von VOTAT über die Aufgaben hinweg. Darüber hinaus und in Erweiterung der aktuellen Forschung fanden wir eine zunehmend effiziente Strategieanwendung. Mit Hilfe eines personenzentrierten Ansatzes konnten vier verschiedene (latente) Klassen von Schülerinnen und Schülern identifiziert werden. In Bezug auf diese Klassen unterstreicht das Ergebnismuster die Bedeutung der Effektivität (gemessen durch VOTAT) für erfolgreiches Problemlösen und ihre Beziehung zur Intelligenz. Es wurden jedoch vergleichbare Intelligenz- und KPL-Leistungen zwischen den Klassen von Schülerinnen und Schülern gefunden, die effektiv, aber mehr oder weniger effizient explorierten. Zusammenfassend wurde ein zunehmend effizienteres Explorationsverhalten festgestellt, dessen Zusammenhang mit Intelligenz und KPL-Leistung weiterer Untersuchungen bedarf.

Die vierte Studie konzentrierte sich auf die Evaluation und Konstruktion einer neuen Parallelversion eines sehr zeiteffizienten Screening-Instruments zur Erfassung eines Indikators der allgemeinen Intelligenz in nur drei Minuten, dem *mini-q* (Baudson & Preckel, 2016). Die Verteilungen der Summenwerte beider Testversionen wiesen auf eine Diskrimination unterschiedlich leistungsstarker Personen hin und korrelierten signifikant miteinander. Hinsichtlich der Validität fanden wir Korrelationen mit einem anderen Intelligenztest und der Abiturnote, die teilweise mit unseren Annahmen übereinstimmten. Die Ergebnisse bezüglich des *mini-q* und dessen Parallelform lassen sich als vielversprechend zusammenfassen und weisen auf ihre Eignung als Intelligenzscreening-Instrumente hin.

Insgesamt haben die Ergebnisse dieses Dissertationsprojektes dazu beigetragen, den Zusammenhang zwischen Intelligenz und KPL weiter aufzuklären und die Anwendungsmöglichkeiten eines Screening-Instruments zur Erfassung der allgemeinen Intelligenz zu erweitern. Hinsichtlich des Zusammenhangs zwischen Intelligenz und KPL konnten sowohl moderierende Effekte, die die Stärke des Zusammenhangs beeinflussen, als auch mediierende Effekte, die verdeutlichen, wie sich Intelligenz in beobachtbaren Verhaltensmustern beim Lösen komplexer Probleme manifestiert, nachgewiesen werden. Weiterhin wurde die Relevanz und das große Potenzial von Logfile-Daten aufgezeigt, die Einblicke in die kognitiven Prozesse von Problemlösenden ermöglichen und damit zur Aufklärung des Zusammenhangs zwischen Intelligenz und KPL beitragen.

1 Introduction

How do people solve problems and how do better problem solvers differ from weaker problem solvers? Under what circumstances does intelligence help people solve problems? And what are the mechanisms that underlie our observation that more intelligent people tend to be better problem solvers? These were the fundamental questions that motivated us to conduct the research carried out within the scope of this dissertation project. The questions' relevance to the field of psychology and educational sciences arises from the demands of today's modern society, in which problem solving is considered a crucial 21st-century skill (e.g., Csapó & Funke, 2017; Greiff, Wüstenberg, & Avvisati, 2015; Osman, 2010) and the demands on learners and the education system are rapidly changing. Schleicher (in Csapó & Funke, 2017) concisely summarized this modern phenomenon: "In the past, education was about teaching people something. Now, it's about making sure that students develop a reliable compass and the navigation skills to find their own way through an increasingly uncertain, volatile and ambiguous world" (p. 3). The quotation above highlights the fact that in everyday and professional life, people are increasingly confronted with unfamiliar problems. In order to solve these problems, certain critical thinking skills are required and people need to be able to apply their knowledge to new situations (Mayer & Wittrock, 2006; Scherer & Beckmann, 2014).

The most prominent construct that relates to the application of critical thinking skills in problem solving is intelligence (Rost, 2013). Intelligence is a broad cognitive capability that includes abilities such as reasoning, planning, abstract thinking, learning from experience, and problem solving (Gottfredson, 1997; Schneider & McGrew, 2018). Moreover, intelligence is considered the most important singular determinant of a person's ability that predicts success in the various roles of our contemporary society (Brody, 1999), such as scholastic achievement (Deary et al., 2007; Jensen, 1998; Neisser et al., 1996; Roth et al., 2015; Valerius & Sparfeldt, 2015), achievement in university (Schmitt et al., 2009; Westrick et al., 2015), job performance (Hunter & Schmidt, 1996; Kramer, 2009; Schmidt & Hunter, 2004; Van Iddekinge et al., 2018), and socioeconomic success (Kuncel et al., 2004; Strenze, 2007). However, intelligence is usually assessed through a battery of tests comprised of tasks like figural matrices (e.g., Raven Advanced

Progressive Matrices Test; Raven, 1998), number series (e.g., Intelligence Structure Test; Beauducel et al., 2010), or verbal analogies (e.g., Berlin Intelligence Structure test; Jäger et al., 1997) that are static and well-defined and, therefore, do not allow for any bidirectional interaction with the task (Süß & Kretzschmar, 2018). As such, these intelligence test tasks do not resemble the complex and dynamically changing problems that people interact with in their everyday lives. They are therefore less useful for observing human problem-solving behavior. Moreover, the vision of *intelligent measurement* by Bunderson et al. (1989) outlines the idea of linking the assessment process to a learning experience by offering individualized advice to test takers that is obtained through the analysis of test taking behavior. Due to their static characteristics and the nature of their tasks (e.g., matrices, number series, analogies), intelligence tests are hardly able to combine the function of an assessment tool and a learning opportunity for real-life problems. For this reason, analyzing the test taking behaviors obtained while working on intelligence tests would be an unsuitable way to provide guidance to teachers and learners on how to foster problem solving.

The described discrepancy between intelligence tests and real-life problems is addressed by the research field of complex problem solving (CPS), which utilizes computer simulations of complex systems to investigate problem solving under controlled conditions in the laboratory (e.g., Dörner et al., 1983; Kröner et al., 2005; Greiff et al., 2012). To this end, CPS tasks can be characterized as dynamic task environments in which the regularities of the underlying system must be revealed through interactive exploration and integration of the derived information (see Buchner's definition of CPS in Frensch & Funke, 1995). More specifically, CPS tasks are comprised of several input variables and output variables that are related to each other. Output variables can change as a function of users' intervention and, potentially, as a function of time, both aspects reflecting the dynamic aspect of CPS. The described changes over time are an important feature of CPS that reflects the dynamic aspects of complex systems; they are commonly referred to as *eigendynamic effects* (Dörner et al., 1983; Funke, 1992; Stadler, Niepel, & Greiff, 2016).

Much like intelligence, CPS is related to success in various important real-life outcomes, such as scholastic achievement (e.g., Lotz et al., 2016; Schweizer et al., 2013; Sonnleitner et al., 2013; Wüstenberg et al., 2012), achievement at university (Stadler, Becker, et al., 2016; Stadler et al., 2018), and socioeconomic success (Danner et al., 2011;

Mainert et al., 2019; see also Coyle & Greiff, 2021). Comparing the two predictors, intelligence and CPS, results from recent studies suggest that CPS predicts scholastic achievement (i.e., grades and competence tests) beyond narrow operationalizations of intelligence (e.g., reasoning: Kretzschmar et al., 2014; Mayer, 2010; Wüstenberg et al., 2012), but mostly not beyond a broad factor of general intelligence (Kretzschmar et al., 2016; Lotz et al., 2016; Sonnleitner et al., 2013). As far as the relationship between intelligence and CPS is concerned, the two constructs are closely related at the theoretical level and extensive research exists concerning their empirical relationship. While early studies found correlations close to zero between intelligence and CPS performance scores (e.g., Brehmer, 1992; Dörner et al., 1983; Putz-Osterloh, 1981), a recent meta-analysis reported a mean average correlation of *M(g)* = .43 (Stadler et al., 2015). However, the factors that influence the relationship between intelligence and CPS have not yet been conclusively clarified. Prominently discussed factors that may possibly influence the intelligence-CPS relationship include the reliability of the utilized CPS assessment tool (Rigas et al., 2002), the different cognitive demands between intelligence test tasks and CPS tasks (Rigas & Brehmer, 1999), and the role of problem-specific prior knowledge (Elshout, 1987; Leutner, 2002; Raaheim, 1988). Therefore, the aim of the first study presented in this dissertation project was to clarify the relationship between intelligence and CPS by investigating the moderating effect of prior knowledge for a psychometrically sound assessment tool of CPS, a so-called multiple complex systems (MCS) assessment tool. In perspective, the further elucidation of the intelligence-CPS relationship provides general insights into the empirical relationship and the relevant influencing factors (e.g., Stadler et al., 2015; Kluwe et al., 1991), the mechanism of intelligence when working on problems (e.g., Elshout, 1987; Raaheim, 1988; Lotz et al., 2017), and the relevance for how to structure effective learning environments (e.g., Chen et al., 2020).

As a computer-based assessment, CPS tasks enable the analysis of computer-generated log-file data that provides insight into behavioral processes related to problem solving. In the last decade, the potential of these log-file analyses has been increasingly recognized and has led to an improved understanding of students' problem-solving behavior and its relation to problem-solving performance and intelligence. For example, the strategic exploration behavior to systematically *vary-one-thing-at-a-time* while all other variables are held constant (also referred to as *VOTAT*; Tschirgi, 1980) when exploring a task was positively related to CPS performance scores (e.g., Greiff, Wüstenberg, & Avvisati, 2015; Schoppek, 2002; Vollmeyer et al., 1996), as well as to intelligence test scores (Lotz et al., 2017). Furthermore, recent studies have shown that the relationship between intelligence and CPS performance is mediated by the strategic CPS exploration behavior VOTAT (Grežo & Sarmány-Schuller, 2022; Lotz et al., 2022; Wüstenberg, Stadler, et al., 2014). In addition, further studies revealed significant relationships between CPS performance and several other CPS behaviors, such as to *vary-no-thing-at-a-time* (*NOTAT*; also referred to as *noninterfering behavior*; e.g., Greiff et al., 2016; Schoppek & Fischer, 2017), time on task (e.g., Scherer et al., 2015), and the number of interactions (e.g., Eichmann, Goldhammer, et al., 2020; Greiff et al., 2016). In sum, these findings clarify how specific CPS behaviors are linked to CPS performance and intelligence, and how more intelligent (compared to less intelligent) participants achieve higher CPS performance scores by applying specific strategic CPS behaviors. However, the CPS behaviors that have been investigated to date are not exhaustive, neither with regard to the behaviors theoretically needed to effectively solve CPS tasks (Beckmann, 1994), nor do they entirely clarify the variance in CPS performance (Greiff et al., 2016). Therefore, it is an important keystone in understanding CPS performance to investigate new and potentially meaningful CPS behaviors. One such CPS behavior is the strategic behavior *to identify eigendynamic effects early*, which was investigated in the second study in the present dissertation project. Identifying eigendynamic effects early is an effective strategic behavior if a system comprises different types of effects (i.e., eigendynamic effects along with other effects; see also Beckmann, 1994). Based on prior studies that focused on effective CPS behaviors, the newly investigated CPS behavior should be positively related to CPS performance and intelligence and may also mediate the intelligence-CPS relationship. Furthermore, this dissertation project aimed to clarify the theoretical status and empirical relationship of a CPS behavior that has yielded conflicting results in previous studies, that is, the efficiency of exploration behavior. To this end, the aim of the third study was to investigate whether students form (latent) classes according to the effectiveness and efficiency of their VOTAT exploration behavior, and whether these classes differ in terms of their CPS performance and intelligence, as expected.

A fourth study was related to intelligence and investigated the psychometric properties of an intelligence screening instrument just taking three minutes, the *mini-q* (Baudson & Preckel, 2016). Moreover, the *mini-q* was complemented by a newly designed parallel version, which was evaluated in the scope of this study. The parallel version of the screening instrument provides an important extension of the test's application options, such as for group testing and test repetitions, and permits the determination of the parallel test reliability of the instrument. In the context of the present dissertation project, the fourth study extends a newly established, highly time-efficient intelligence screening instrument, which can be utilized in future investigations of the intelligence-CPS relationship.

Although the questions raised at the beginning of this chapter are certainly too far-reaching to be exhaustively answered in one dissertation project, the present thesis aims to shed further light on the relationship between problem solving and intelligence by investigating the relationship between specific CPS behaviors, intelligence, and CPS performance. Moreover, we investigated potential moderators (i.e., prior knowledge) and mediators (i.e., CPS behaviors) of the intelligence-CPS relationship. Additionally, in the fourth study, we complemented and evaluated a highly time-efficient intelligence screening instrument. The results of this research can provide deeper insights into the mechanisms of intelligence and problem solving and the interrelationship between the two constructs. Therefore, the articles presented within this dissertation project aim to answer the following research questions:

Research Question 1: Does prior strategic knowledge in the sense of the CPS behavior to vary-one-thing-at-a-time (VOTAT) moderate the intelligence-CPS relationship in a way that correlation coefficients follow an inverted U-shaped curve, as predicted by the Elshout-Raaheim hypothesis?

Research Question 2: Is the specific strategic CPS behavior to identify eigendynamic effects early in the knowledge acquisition phase related to CPS performance and intelligence and, moreover, does the described CPS behavior mediate the intelligence-CPS relationship?

Research Question 3: Can students be grouped in (latent) classes according to the effectiveness and efficiency of their VOTAT exploration behavior across a sequence of CPS tasks? Do the assumed (and empirically found) latent classes differ in regard to their CPS performance and intelligence as expected?

Research Question 4: Can the psychometric properties of the *mini-q* be replicated and are the psychometric properties of a newly developed parallel version comparable to the original version? Furthermore, do the two test versions show an appropriate parallel test reliability?

Each of the Research Questions outlined above is addressed by one of the studies presented in this thesis (see Chapter 3: Empirical Studies). To provide the reader a comprehensive overview of the theoretical concepts underlying the four studies, Chapter 2 (Theoretical Background) introduces important concepts and outlines their theoretical and empirical relationships. Specifically, Section 2.1 gives an overview of general intelligence, followed by an overview of the definitions, demands, and the assessment of CPS in Section 2.2. In addition, Section 2.2 introduces various process measures based on log-file data and discusses their relationship to CPS performance measures. Section 2.3 provides an overview of the theoretical and conceptual overlap of intelligence and CPS and presents empirical findings on the relationship between intelligence and CPS performance measures, as well as intelligence and CPS process measures. In the same section, theoretical considerations and empirical findings regarding the mediating effect of CPS process measures on the intelligence-CPS relationship are reviewed. In Chapter 4 (Discussion), the findings are reviewed in light of how students interact with complex and dynamic systems and how intelligence supports and manifests itself in complex problem solving. In addition, this chapter discusses the limitations of the studies presented and previews how our findings may inform potentially fruitful future research projects.

2 Theoretical Framework

2.1 Intelligence

One of the greatest achievements of modern psychological research can be found in the objective measurement of mental abilities and the associated development of comprehensive theories about the structure of intelligence (McGrew, 2009). More than 100 years of intensive research have made intelligence one of the best-studied characteristics in the field of psychology (Rost, 2013), which has also attracted much attention beyond its original field of research in disciplines such as the social sciences, epidemiology, and brain research (Klauer & Spinath, 2010). Intelligence describes interindividual differences between people and can be regarded as the most important singular determinant of a person's ability to predict success in occupational, academic, and a variety of other everyday life domains (Brody, 1999; Deary et al., 2007; Hunter & Schmidt, 1996; Neisser et al., 1996; Roth et al., 2015; Schmidt & Hunter, 2004). Although there is no uniformly established definition of intelligence, there has been a strong consensus on the concept of intelligence. A survey among scientists in the field of intelligence conducted in 1921 revealed that the most common elements in the proposed definitions of intelligence were (a) higher level abilities (e.g., abstract reasoning, mental representation, problem solving, and decision making), (b) ability to learn, and (c) adaptation to meet the demands of the environment effectively (Sternberg, 1997). A more recent and well-established definition of intelligence that reflects mainstream science on intelligence and that was signed by 52 highly recognized researchers of the field is provided by Gottfredson (1997) as follows:

> "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – "catching on," "making sense" of things, or "figuring out" what to do." (p. 13)

A comparison of the most common elements of the 1921 proposed definitions of intelligence (Sternberg, 1997) and Gottfredson's (1997) mainstream definition of intelligence shows a high degree of overlap, reflecting the strong consensus that has formed around the definition of the concept. Among other common denominators, problem solving and the ability to learn from experience and adapt to the demands of our environment are understood as key elements of intelligence in both descriptions.

Closely related to the question of defining intelligence is research on the structural nature of the construct. The scientific discourse revolves around the question of which and how many elements intelligence is composed of and how these elements are arranged in relation to each other. The starting point of today's research on the structure of intelligence was Spearman's (1904) finding that all cognitive tasks, though they may capture different aspects such as mathematical, verbal, or even perceptual skills, are positively correlated with each other, a pattern referred to as *positive manifold*. To explain the positive manifold, Spearman assumed the existence of a general factor of intelligence, symbolized as *g*, that manifests itself in every cognitive task. It is therefore accountable for the pattern of correlations. To account for the variance in cognitive tasks that could not be explained by *g*, Spearman hypothesized further factors, each of them specific to a certain cognitive task and uncorrelated to *g*, as well as to all other specific factors. Importantly, the different cognitive tasks differ in the proportion in which they are composed of the general factor *g* and a particular specific factor. Thus, Spearman assumed that the performance in any cognitive task can be attributed to a general factor of intelligence, to an uncorrelated specific factor, and to measurement error. However, the notion that there are patterns of correlations among groups of similar cognitive tasks, so-called group factors, that are not accounted for by a general factor of intelligence, led to the rejection of Spearman's specific and uncorrelated factors (Kelley, 1928; Thurstone, 1936, 1938). In contrast, the positive manifold has been replicated in numerous factor-analytic studies and is "arguably both the best established and the most striking phenomenon in the psychological study of intelligence" (Van der Maas et al., 2006, p. 855).

Further influential theoretical models of intelligence include Carroll's (1993) three stratum theory and the Cattell-Horn theory of fluid and crystallized intelligence (Gf-Gc theory; Cattell, 1963; Horn, 1991; see also Horn & Blankson, 2005). Regarding a general factor of intelligence, the assumptions of the two models differ: Whereas Carroll (1993) confirmed the positive manifold in hundreds of factor-analytic studies and, therefore, assumes a *g* factor at the apex of his hierarchical model, the Gf-Gc theory denies the existence of a *g* factor. Despite the differences on a *g* factor, the Three-Stratum theory and the Gf-Gc theory also share many common characteristics, for example, they share significant overlap in content and structure with respect to the broad abilities in the second stratum. The umbrella term Cattell–Horn–Carroll (CHC) theory of intelligence combines and unifies the two models mentioned above and represents a common nomenclature and an overarching theoretical framework in the field of human cognitive abilities (Daniel, 1997; McGrew, 2005, 2009; Schneider & McGrew, 2018). Furthermore, the CHC theory seeks to evolve with the gradual progress of scientific knowledge so that every scientist can contribute to the theory's evolution in a transparent and open process. A simplified representation of the model can be found in Figure 1.

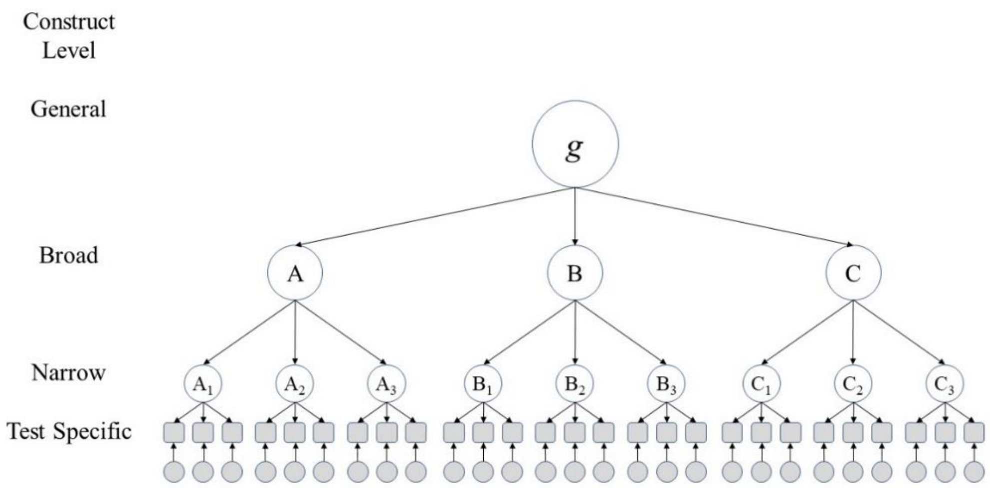

Figure 1: Hierarchical Structure of the CHC theory.

Note. Adapted from "The Cattell–Horn–Carroll theory of cognitive abilities" by Schneider, W. J., & McGrew, K. S. (2018). In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., p. 74). The Guilford Press.

What follows aims to describe the current version of the abilities and factors of CHC theory (Schneider & McGrew, 2018). At the bottom of the hierarchy, there are *specific abilities* that can be measured directly through a specific test or task. Such specific abilities include, for example, correctly continuing a series of numbers or repeating sentences back after hearing them just once. The relationships among the specific abilities are used to derive the narrow and broad abilities as well as the *g* factor, which are higher in the hierarchy and cannot be measured directly. *Narrow abilities* can be described as clusters of specific abilities that are highly correlated to each other. For example, the ability to repeat back letters is highly correlated with the ability to repeat back digits. Both specific abilities can therefore be attributed to the narrow ability of auditory short-term memory storage capacity. *Broad abilities* subsume clusters of narrow abilities, which are more correlated among each other than with narrow abilities in other broad ability clusters. Current broad abilities include, for example, fluid reasoning (Gf), comprehension-knowledge (Gc), domain-specific knowledge (Gkn), reading and writing (Gw), quantitative knowledge (Gq), visual ability (Gv), and auditory ability (Ga; for a complete list of broad abilities, see Schneider & McGrew, 2018). At the top of the hierarchy, there is the *g* factor that represents the general factor of intelligence. The *g* factor indicates the correlation among all of the broad ability factors and refers to the positive manifold, which was first discovered by Spearman (1904) and confirmed many times thereafter.

Although the CHC theory does not settle the dispute between Three-Stratum theory and the extended Gf-Gc theory with regard the *g* factor, leaving it up to the individual scholar to make use of the *g* factor or not (Kaufman et al., 2019), it provides a clear idea of how the *g* factor can be understood. In the present version of the CHC theory, it is assumed that a *g* factor exists, but it remains an open question as to whether it is constituted of one single ability (Schneider & McGrew, 2018). These assumptions imply that one can refer to a person's overall level of intelligence, but it is unclear whether this overall factor of intelligence can be attributed to a single causal force.

In sum, intelligence is one of the best studied characteristics in all of psychology, describing interindividual differences between people. The general factor of intelligence, *g*, reflects a very broad factor of mental capability that is comprised of such abilities as, for example, abstract reasoning, learning, and adapting to new situations. Although problem solving is part of common definitions of intelligence (e.g., Gottfredson, 1997), the *g* factor represents a much broader and more general cognitive ability that is beyond problem solving.

The main research questions of the presented dissertation project revolve around the relationship between intelligence and problem solving (study 1 to study 3). In multiple past studies that investigated different research questions related to this relationship, intelligence was operationalized only through one broad ability (e.g., reasoning; Leutner, 2002; Wüstenberg, Stadler, et al., 2014), thus not reflecting the full variance of the *g* factor. Therefore, a central methodological feature of the studies 1 through 3 of this dissertation project, which examined the relationship between intelligence and CPS, was to extract a strong *g* factor that adequately represented the broad nature of intelligence, relating to the idea of a "good *g*" (Jensen & Weng, 1994). This goal was achieved by utilizing a balanced set of multiple diverse intelligence subtests of high reliability. In a further, fourth study, we investigated the properties of a very time-efficient intelligence screening instrument, the *mini-q* (Baudson & Preckel, 2016), and presented a new parallel version of the instrument that expands its applicability. The *mini-q* and its parallel version provide an indicator of general intelligence that does not reflect the broad nature of *g*. However, the very short administration time of only three minutes allows the integration of this instrument in (large) studies, where time for data collection is often very limited. In the context of the present dissertation project, the *mini-q* and the presented parallel version offer a very time-efficient possibility to obtain a rough indicator of general intelligence in future CPS research projects.

2.2 Complex Problem Solving

2.2.1 Definitions and demands of complex problem solving

Problem solving is essential to master the novel and dynamically changing situations that we encounter in our daily lives. Such problems may be found in, for instance, educational contexts in which students are expected to work with unfamiliar environments in science classes (Csapó & Funke, 2017) or contexts in which people are confronted with processes related to climate change (Osman, 2017). In this vein, problem solving can be understood as a non-routine process, in which a system is transformed from its current state to a goal state (Jonassen, 2000; see also Funke, 2012; Mayer & Wittrock, 2006).

Complex problem solving (CPS) is a specific type of problem solving that can be described by the following five characteristics (Dörner et al., 1983): The underlying system of the corresponding problem is comprised of multiple variables (complexity) that are interrelated to each other (connectedness) and can either change as a result of the problem solver's interventions or over time (dynamics). At the same time, the structure of the system is unknown and opaque, and therefore needs to be actively explored (intransparency). Further, the solution to the problem may consist of several, at times even contradictory goals (referred to as polytely). All of these aspects are summarized in a widely accepted definition of CPS given by Buchner (in Frensch & Funke, 1995):

> "The successful interaction with task environments that are dynamic (i.e., change as a function of user's intervention and/or as a function of time) and in which some, if not all, of the environment's regularities can only be revealed by successful exploration and integration of the information gained in that process." (p. 14)

Buchner's definition emphasizes the character of CPS as a non-routine process that imposes different demands upon the problem solver. To understand these demands, it is useful to distinguish between the problems solver's understanding of the problem and the derived solution, which implies a sequence of operations that transforms the system from its given state to a desired goal state (Mayer, 1992; Novick & Bassok, 2005). In the process of understanding a problem, the problem solver explores the problem and acquires information about the structure of the underlying system and the manipulations that can be applied to it and thereby forms a representation of the problem. CPS research commonly refers to this process as knowledge acquisition (Funke, 2001). To reach the desired goal states, problem solvers manipulate the system by applying the suitable operators based upon the prior assumptions that they have developed during knowledge acquisition (Mayer, 1992; Novick & Bassok, 2005). This process is commonly referred to as knowledge application in CPS research (Funke, 2001). As problem solvers rely on their problem representation when generating a solution, the process of acquiring information about the problem (i.e., knowledge acquisition in CPS) and the derived solution (i.e., knowledge application in CPS) are highly interrelated (e.g., Beckmann & Guthke, 1995; Kersting, 1999; Kröner et al., 2005; Süß & Kretzschmar, 2018).

Importantly, the unique characteristics of the problem, such as context, and the characteristics of the problem solver, such as prior knowledge or cognitive ability, influence how problem solvers construct the representation of the problem (Novick & Bassok, 2005; Mayer & Wittrock, 2006). As concerns context, the embedding of a problem in different story contents, though structurally equivalent, can affect the problem representation and therefore the overall difficulty of the problem (Hayes & Simon, 1977; see also Beckmann & Goode, 2014; Klahr, 2000). Similarly, prior knowledge (Elshout, 1987; Leutner, 2002; Raaheim, 1988) and abstract schemes (Novick, 1988; Novick & Holyoak, 1991; Putz-Osterloh, 1993) influence the problem representation and, in turn, affect the solution that is derived. As relates to cognitive ability, problem solving is a process of "cognitive processing" (Mayer & Wittrock, 2006, p. 287) and therefore a cognitive task that is related to the positive manifold described by Spearman (1904). This assumption is reflected in the results of a recently published meta-analysis, which revealed an average correlation of $M(g) = .43$ between CPS and intelligence (Stadler et al., 2015).

In short, CPS is defined as the interaction with dynamic task environments that need to be actively explored in order to solve the problem. Furthermore, in the process of complex problem solving, a distinction is made between knowledge acquisition and knowledge application, and the performance of the two CPS phases is related to context, prior knowledge, and intelligence.

2.2.2 Assessment of Complex Problem Solving

The historical starting point from which CPS evolved was the alleged criticisms of Dörner and colleagues that intelligence was not related to problem solving in everyday real-life situations (e.g., Dörner & Kreuzig, 1983; Brehmer & Dörner, 1993). Therefore, the aim of CPS was to investigate the interaction and performance with complex everyday problems under controlled conditions in a laboratory (Dörner et al., 1983). In order to realize the characteristic feature of CPS of interacting with dynamically changing task environments, real-world problems were simulated in the laboratory by using computer models of complex scenarios, so-called microworlds. In these classical CPS assessments, problem solvers had to deal with, for example, the responsibilities of a small-town mayor (Lohhausen; Dörner et al., 1983), or those of a business manager of a factory (Tailorshop; Putz-Osterloh, 1981), or had to provide development aid to a small African tribe (Moro; Dörner et al., 1986), or protect a city from fire (Fire Fighting; Brehmer, 1987). Classical CPS assessments featured a high degree of complexity and interconnectivity. For example, the city “Lohhausen” was modeled with over 1000 interconnected variables. Therefore, these microworlds displayed a pronounced resemblance to real-world problems and showed a high level of face validity.

However, these early CPS assessments did not meet the standards of sufficient psychometric measurement, as has been described by various authors (e.g., Funke, 1995; Greiff & Funke, 2010; Kröner et al., 2005; Sternberg, 1995). The single task structure of the classical CPS assessments meant that single or even random errors at the beginning of a task could severely impair the overall test performance, and variation in task difficulty was not possible. On the contrary, problem solvers who performed poorly in such a scenario were confronted with increasing difficulty to reach the desired goal state, a test characteristic that has been described as contra adaptive (Sonnleitner et al., 2012). Moreover, classical CPS assessments were time consuming and prior knowledge about the simulated real-world problems heavily influenced test takers' overall performance. These features resulted in low or even unknown estimates of reliability (see Kröner, 2001; Süß, 1996; Wüstenberg et al., 2012). A further problem was related to the lack of comparability between different problem-solving scenarios that differed with regard to surface features (i.e., semantic context of the problem) as well as to the underlying system structure of the problem (Funke, 2001). Overall, the interpretation of findings based on classical CPS assessments tools was compromised by poor psychometric properties and large differences in the context and underlying system of the simulated problem scenario (i.e., CPS task).

To address these issues of classical CPS assessment tools, Funke (2001) introduced Linear Structural Equation systems (LSE) and Finite State Automata (FSA), which are two formal frameworks that describe a task's underlying structure. Both frameworks enabled modelling different degrees of connectivity and dynamics in simulated complex problems. As LSE are based on linear systems, tasks that utilize this framework deal with quantitative variables that are measured on interval scales (e.g., the quantitative effect of substance A on substance B). In comparison, FSA can be represented by state-transition matrices and are utilized in tasks dealing with qualitive variables measured on nominal scales [e.g., the state of one variable (Yes/No) affects the state of another variable (Yes/No)]. The formal description of the underlying systems made it possible to transfer these systems into different semantic embeddings. Moreover, the frameworks allowed researchers to construct problems without meaningful semantic embedding to reduce the influence of prior knowledge.

The introduction of the LSE framework was the starting point from which a variety of psychometrically sound CPS assessment instruments emerged. Direct implementations of Funke's approach are single complex systems (SCS) that are based on one linear structure equation, such as "MultiFlux" (Kröner, 2001) or "LINAS" (Schoppek, 2002). A further important step was taken by administering several complex systems that are based on LSE chronologically one after the other, an approach referred to as multiple complex systems (MCS; e.g., "MicroDYN", Greiff et al., 2012; "Genetics Lab" Sonnleitner et al., 2012). Most importantly, the MCS approach allows researchers to calculate an aggregated performance score over multiple complex systems. Further, these multiple CPS tasks can vary in task difficulty, circumventing the counter adaptive nature of classical, single complex systems. The semantic embedding of approaches such as MicroDYN or Genetics Lab are entirely fictive, which reduces the influence of prior knowledge. In sum, the MCS approach of CPS showed significant improvements with regard to reliability and validity, overcoming the psychometric shortcomings associated with classical CPS assessment tools (e.g., Greiff et al., 2012; Sonnleitner et al., 2013; Stadler et al., 2018; Wüstenberg et al., 2012). The improved psychometric characteristics make the MCS approach a popular assessment tool in empirical educational research (e.g., Csapó & Molnár, 2017; Kretzschmar et al., 2014) and led to the inclusion of CPS in the 2012 PISA survey (OECD, 2013).

To illustrate the structure and task characteristics of MCS assessment tools, MicroDYN (Greiff et al., 2012) will be presented here as an example. MicroDYN is comprised of several independent and fictive tasks, each of which is comprised of a set of interrelated input and output variables, as displayed in Figure 2.

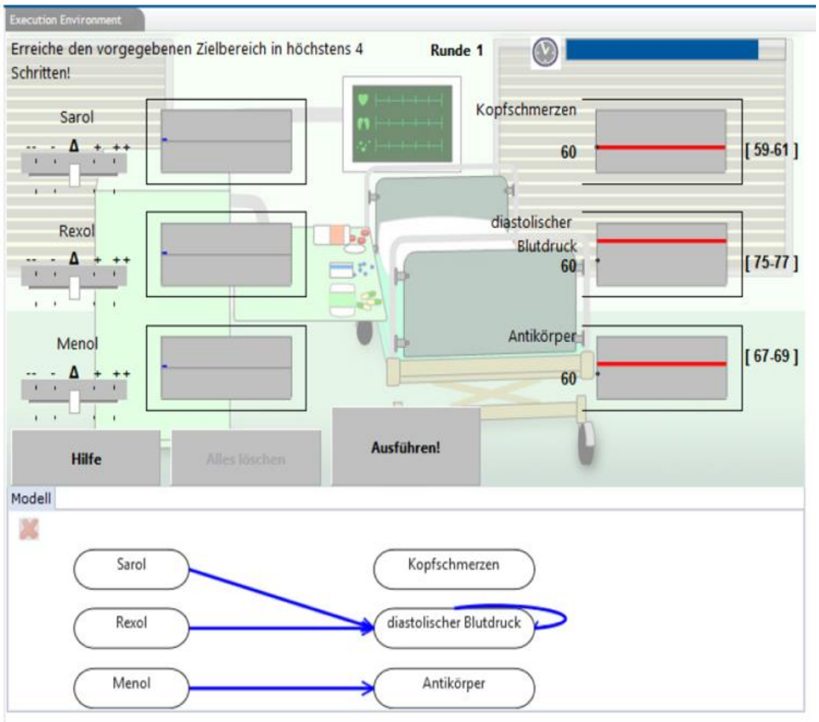

Figure 2: Screenshot of the MicroDYN task *Medical Aid* in the first round of the knowledge application phase (Weise et al., 2022). Red lines in the graphical output of the outcome variables "Kopfschmerzen" ("head ache"), "diastolischer Blutdruck" ("diastolic blood pressure"), and "Antikörper" ("antibodies") symbolize the target values; beneath the execution environment, the correct causal diagram ("Modell") is depicted.

To realize the CPS characteristics of complexity, connectedness, and dynamics, variables are related to each other by direct and eigendynamic effects. Direct effects represent a connection of an input variable to an output variable; eigendynamic effects imply a change of an output variable as a function of time. In the task displayed, the output variable of "Antikörper" is influenced via a direct effect by the input variable "Menol"; "diastolischer Blutdruck" is directly influenced by "Sarol" and "Rexol" and also changes dynamically over time, and "Kopfschmerzen" is not influenced by any variable. In the knowledge acquisition phase, problem solvers freely explore the system and record their conclusions in a causal diagram (see the lower third of Figure 2), which represents the problem representation. In the knowledge application phase, participants are provided with the correct causal diagram and are assigned to reach specific target values of the output variables. For each task, dichotomous performance measures are calculated separately for the two CPS phases. For the knowledge acquisition phase, credit is given if all of the relationships in the system are correctly displayed in the causal diagram, otherwise no credit is given. For the knowledge application phase, credit is given if the target values of all output variables are reached within a limited number of steps, otherwise no credit is given. To evaluate overall CPS performance, aggregated scores across the set of tasks are calculated separately for the two CPS phases.

2.2.3 Process measures of CPS and their relation to CPS performance

The computer-based assessment in combination with the interactive nature of CPS offers a way to analyze behavioral processes and sequences of problem solvers' actions, which provides additional information beyond performance scores. Process measures are obtained from log-files that record participants' single actions and thereby provide insight into how they explore the complex system (e.g., Lotz et al., 2017; Vollmeyer et al., 1996), how much time they spend on working on a given task (e.g., Goldhammer et al., 2014; Scherer et al., 2015), and the number and specific patterns of interactions with the problem (e.g., Eichmann, Greiff, et al., 2020; Stadler et al., 2019).

Although research on process measures has increasingly gained attention in the field of CPS, the potential of such data is not yet being fully exploited. Process measurements of CPS can advance our understanding of complex problem solving and promote improved and more sophisticated use of new methods for assessing problem-solving proficiency. First, they provide information about which behaviors are related to successful problem solving and thereby offer advice to learners on how to become better problem solvers and to teachers about how to better support their students (e.g., Greiff et al., 2016; Greiff, Wüstenberg, & Avvisati, 2015). In future assessment instruments, the evaluation of process measures may open up the possibility to link a learning experience to the assessment process by offering individualized advice to test takers, relating to the idea of intelligent measurement (Bunderson et al., 1989). Second, and from a cognitive perspective, process measures help researchers understand the cognitive mechanisms that underlie test-takers' performance in CPS tasks. Findings obtained through investigating process measures provide potentially deeper insights into participants' cognitive abilities and may eventually shed light on relationships with other cognitive abilities such as intelligence. And third, process measures offer explanations about how group differences in CPS performance (e.g., gender, country, cultural background) can be attributed to different problem-solving behaviors, thus providing greater direction to educational policy makers (e.g., Eichmann, Goldhammer, et al., 2020; Wüstenberg, Greiff, et al., 2014). In the following section, different process measures of (predominantly) MCS assessment tools of CPS and their relation to CPS performance will be introduced and discussed.

2.2.3.1 CPS performance and CPS process measures that capture the exploration process: VOTAT and NOTAT

Important and frequently investigated process measures that provide information about how participants explore the system include the strategic behavior to systematically vary-one-thing-at-a-time (VOTAT; Tschirgi, 1980) and to *vary no-thing-at-a-time* (NOTAT; Lotz et al., 2017; also referred to as noninterfering behavior; Greiff et al., 2016). When using the strategic behavior VOTAT, just one input variable is varied and all of the remaining input variables are held at zero. Therefore, the effect of a variable on the system is isolated and VOTAT constitutes an effective strategic behavior when identifying direct effects in complex systems (Tschirgi, 1980). NOTAT is characterized by holding all of the input variables at zero and observing changes in the system over time, which is an effective strategic behavior used to detect eigendynamic effects (Beckmann, 1994). Both strategic behaviors – VOTAT and NOTAT – are domain-general behaviors and represent the most effective way to conduct unconfounded experiments when exploring an unknown system. As concerns the process measures' relations to CPS performance, effective exploration behavior is both theoretically and empirically linked to performance in the knowledge acquisition and the knowledge application phase. Based upon theoretical considerations, participants can only display the correct relationships in the causal diagram if they have the ability to correctly identify the relationships within the system while exploring. Thus, more effective exploration, such as the use of a strategic exploration behavior (i.e., VOTAT and/or NOTAT), should be associated with higher performance scores in the knowledge acquisition phase (Dörner, 1980; Dörner & Schaub, 1994). With regard to the second CPS phase, participants who explore the system more effectively by using VOTAT and/or NOTAT should achieve higher performance scores in the knowledge application phase due to increased knowledge of the system (Kröner et al., 2005). These theoretical considerations are reflected in the empirical findings on the relationship between strategic exploration behaviors and CPS performance measures.

As concerns empirical findings on VOTAT, evidence from structure equations models indicate that the relative frequency of using VOTAT and performance in the knowledge acquisition phase are significantly related to each other ($\beta = .22$, $SE = .12$, $N = 101$, p-value not reported; Kröner et al., 2005). In a confirmatory factor analysis that used a dichotomous score for VOTAT (full credit was given if the CPS behavior was at least applied once for every input variable), a strong latent relationship between the strategic behavior and performance in the first CPS phase was found ($r = .97$, $p < .05$; Wüstenberg et al., 2012). Data from the PISA 2012 assessment of problem solving revealed that applying VOTAT was positively related to performance in the knowledge acquisition phase of the corresponding item ($r = .67$, $p < .001$) and to overall problem-solving proficiency in a larger set of items ($r = .61$, $p < .001$; Greiff, Wüstenberg, & Avvisati, 2015). Similarly, Vollmeyer et al. (1996) found that a dichotomous score of VOTAT was related to performance in the knowledge acquisition ($r = .76$, $p < .001$), as well as in the knowledge application phase ($r = .32$, $p < .05$). These findings are further supported by the results of person-centered approaches that classified students based on their VOTAT use across multiple items (Greiff et al., 2018; see also Gnaldi et al., 2020): Classes of students who used VOTAT consistently achieved statistically significantly higher performance scores in both CPS phases than classes of students who did not use VOTAT consistently or who used a different CPS behavior.

However, to capture a comprehensive picture of VOTAT's theoretical status, its development across tasks should be considered. In the field of scientific reasoning, VOTAT (also referred to as *control of variables strategy*, CVS; Chen & Klahr, 1999) is regarded as a crucial cognitive skill that develops with practice and exercise (Zimmerman, 2007). Likewise in CPS, the relative frequency of the VOTAT application increased over several rounds when exploring a complex problem (round 1/2/3/4: 19%/47%/47%/56%, linear trend: $F(1, 35) = 23.00$, $p < .001$; Vollmeyer et al., 1996). Further, in a set of five structurally different MCS tasks, a consistent increase in the relative VOTAT frequency was observed (task 1/2/3/4/5: 47%/58%/63%/67%/71%; slope: $\beta = .66$, $p < .05$; Lotz et al., 2017). Likewise, several studies that relied on a person-centered approach to classify students based on their VOTAT use across multiple CPS tasks found a group of so-called "rapid learners", in which students showed increasing levels of VOTAT use over the course of tasks (Greiff et al., 2018; Molnár & Csapó, 2018; Mustafić et al., 2019). Taken together, these findings suggest that working on several different tasks for which VOTAT is an effective strategic behavior can be regarded as a form of practice and exercise for this particular strategic behavior.

To summarize the research on VOTAT, the empirical findings support the theoretical notion that VOTAT is an important strategic exploration behavior that supports and facilitates the correct identification of relationships within a complex system and their recording in a causal diagram (i.e., performance in the knowledge acquisition phase). The substantial relationship between VOTAT and overall performance in both CPS phases and a larger set of problem-solving tasks, suggests that VOTAT can be considered "a good indicator of general strategic knowledge" (Greiff et al., 2016, p. 45). Moreover, working on a set of CPS tasks for which VOTAT is an effective strategic behavior acts as a kind of practice and increases participants' use of this strategic behavior.

As concerns the empirical findings on NOTAT, the aggregated score of the beforementioned strategic behavior was positively related to CPS performance in the knowledge acquisition phase ($\beta = .36$, $p < .001$) and the knowledge application phase ($\beta = .30$, $p < .001$; Greiff et al., 2016; see also Lotz et al., 2017, 2022). Moreover, in an experiment that investigated the transfer of cognitive skills across CPS assessment tools, the use of NOTAT in one assessment tool (i.e., MicroDYN) was substantially correlated with the performance score of controlling a system in a different CPS assessment tool (i.e., Dynamis2; $r = .46$, $p < .001$; Schoppek & Fischer, 2017). An investigation of the strategic behavior's development indicated that the relative frequency of NOTAT increased over four structurally different MCS tasks, for which this behavior was effective (task 6/7/8/9: 10%/10%/10%/13%, slope: $\beta = .30$, $p < .05$; Lotz et al., 2017). In short, the correlation pattern and development of NOTAT suggest that this specific strategic CPS behavior can be interpreted as an indicator of general problem-solving proficiency that increases with practice, similar to VOTAT.

2.2.3.2 CPS performance and CPS process measures related to time on task

In MCS assessment tools of CPS, time on task is commonly defined as the time that participants spend working on the knowledge acquisition phase of a single task (e.g., Greiff et al., 2016; Scherer et al., 2015). Goldhammer et al. (2014) explained the differences in the relationship between performance scores and time on task through the dual process theory framework, which distinguishes between tasks that require routine processing and tasks that require largely non-automated and controlled processing. For routine tasks, the dual process theory assumes the existence of a negative relationship between time on task and task performance. For tasks that require controlled processing, such as CPS, a positive relationship between time on task and performance is assumed. Empirical findings support these considerations, showing a medium-sized relationship between time on task and overall CPS performance scores (Goldhammer et al., 2014: $\beta = .56$, $p = .02$; Scherer et al., 2015: $\rho = .40$, $p < .05$). However, these studies only considered a linear relationship between the two measures. In a different study that incorporated a quadratic term, the relationship between time on task and CPS performance followed an inverted U-shaped curve (knowledge acquisition phase: linear term: $\beta = .39$, $p < .001$, quadratic term: $\beta = .59$, $p < .001$; knowledge application phase: linear term: $\beta = .34$, $p < .001$, quadratic term: $\beta = .38$, $p < .001$; Greiff et al., 2016). This means that medium time on task was associated with the highest CPS performance, and short time on task or long time on task were both associated with poor CPS performance scores. These findings suggest that in tasks requiring controlled processing, there is an optimal time allocation for a specific task. However, the relationship between time on task and task performance is influenced by the person's problem-solving proficiency as well as the overall difficulty of the task (Goldhammer et al., 2014).

So-called planning activities are another process measure that is related to time on task, for which a positive effect on problem solving was demonstrated (Albert & Steinberg, 2011). A recent study investigated planning activities when solving CPS tasks and operationalized this process measure by the longest time interval, in which no interaction with the problem occurs (Eichmann et al., 2019). Whereas the duration of the longest planning interval was not related to CPS performance, the time at which planning activities took place was related to CPS performance (*z-value* = -2.80, $p = .005$), suggesting that early planning phases are related to higher performance scores when solving complex problems. This effect occurred in a pronounced way for easy CPS tasks and decreased for more difficult tasks. The effect of the time when planning took place supports the idea that CPS requires controlled processing (Goldhammer et al., 2014). In summary, the empirical evidence suggests that the relationship between time on task and CPS performance has not been conclusively resolved. However, both a person's ability and the difficulty of the task appear to moderate the relationship between CPS performance and time on task.

2.2.3.3 CPS performance and CPS process measures related to interactions with the system

In contrast to time on task, there is no common definition of the process measure interactions with a system and therefore, studies use different operationalizations of this process measure. Moreover, there are different hypotheses with regard to the relation of interactions with the system to CPS performance. On the one hand, high interaction frequency may indicate a lack of planning and engagement with the task and should therefore be associated with lower CPS performance (Greiff et al., 2016). Following this line of argument, the process measure interactions could be regarded as an indicator of the participants' efficiency. On the other hand, more interactions go along with an intensified exploration process and improved problem representation, meaning that interactions should be related to higher CPS performance scores (Eichmann, Goldhammer, et al., 2020; see also Bell & Kozlowski, 2008; Keith & Frese, 2005).

Empirically, Greiff et al. (2016) examined the relationship between CPS performance and intervention frequency, operationalized as the average number of times the "Apply" button was clicked per time unit in the knowledge acquisition phase when working on MCS tasks. Controlling for time on task, VOTAT, and NOTAT, the intervention frequency was negatively related to CPS performance in both the knowledge acquisition phase ($\beta = -.08$, $p < .05$) and the knowledge application phase ($\beta = -.11$, $p < .05$), indicating that students who showed higher levels of intervention frequency achieved lower CPS performance scores. According to the authors, these results reflect that higher intervention frequencies are associated with a non-targeted exploration process and unplanned behavior. Importantly, the process measure intervention frequency reflects the interactions per time unit, but not the total number of interactions. Different findings were obtained when interactions were operationalized as the total number of click events that occurred during a task in a set of different CPS tasks based upon the LSE and FSA approach (Eichmann, Goldhammer, et al., 2020). CPS performance was positively related to the number of interactions (Model 1/2: $b = 0.71/0.74$, $p < .001/.001$) and the number of additional interactions above the number of necessary interactions (for both models: $b = 0.44$, $p < .001$). Similarly, for problem solving performance in technology-based environments, a higher number of interactions was associated with higher performance scores (Naumann et al., 2014).

In short, these findings on the process measure interactions with the system are not consistent and may depend on its operationalization (frequency per time or the total number of interactions) as well as further characteristics inherent to the task (e.g., task difficulty) and the problem solver (e.g., their cognitive ability). If the process measure interactions should represent an indicator of planning activities and efficiency as outlined by Greiff et al. (2016), then it could be useful to combine the process measure interactions with other process measures that reflect the nature of these actions. In the present dissertation project, we propose to indicate the efficiency of students' exploration behavior by the number of non-necessary exploration steps when applying VOTAT for each input variable (see study 3, Ruby et al., submitted). This operationalization of efficiency is based on the assumption that if a given system consists only of direct effects, then applying VOTAT once for each input variable should provide all the information needed to draw the correct conclusions about the underlying effects of the system. Thus, a more efficient exploration behavior would be indicated by a lower number of non-necessary exploration steps, and a less efficient exploration behavior would be indicated by a higher number of non-necessary exploration steps. In line with the findings of Greiff et al. (2016), we assumed that a more efficient behavior should be related to higher CPS performance scores. However, and in contrast to the study of Greiff et al. (2016), the operationalization of efficiency by the number of non-necessary exploration steps when applying VOTAT provides a more concise insight into how students achieve effective and efficient behavior.

2.2.3.4 CPS performance and other process measures

The large amount of log-file data in computer-based assessment tools of CPS allows for the data to be combined into new process measures that are potentially related to CPS performance. For example, a recent study examined differences in strategic behavior between students who applied the VOTAT strategy but either mastered or failed the knowledge acquisition phase (Stadler et al., 2019). Specifically, strings of behavior distinguishing between working on the input variables of the system (referred to as *S*) and working on the corresponding causal diagram (referred to as *M*) in the knowledge acquisition phase were extracted from the log-file data. The overall pattern of results was that students who achieved higher performance scores in the knowledge acquisition phase tended to work fewer rounds continuously on the scenario (e.g., SSS or SSSS), but spent more time working on the model (e.g., MMM or MMMM). The authors of the study suggest that participants who did not immediately record their findings in the casual diagram after exploring the system increased their cognitive load, thereby making it more difficult to complete the task. In summary, not only the strategic exploration behavior is important for correctly solving CPS tasks (i.e., all students used VOTAT), but also the meta-strategic behavior of when to record conclusions in the causal diagram.

In a further study that comprised two CPS tasks based on the LSE and FSA frameworks, the log-file data were used to distinguish between goal-directed behavior (i.e., all actions necessary to solve the task correctly; for LSE, these were using VOTAT, drawing a correct line, and deleting a wrong line in the causal diagram) and non-targeted exploration behavior (i.e., all actions that are not necessary to solve the task; Eichmann, Greiff, et al., 2020). Moreover, groups of students were identified that exhibited similar behavioral patterns with relation to goal-directed and non-targeted behavior. Students in groups that exhibited long sequences of repeated goal-directed behavior, which the authors interpreted as double-checking, achieved the highest CPS performance scores, followed by students in groups that exhibited short sequences of goal-directed behavior, which can be interpreted as efficient and effective behavior. In contrast, non-targeted exploration was associated with failing to solve the tasks correctly. These findings were in line with previous studies that have indicated that goal-directed behaviors, such as VOTAT, are related to higher CPS performance scores. However, the finding that double-checking behavior was associated with the best CPS performance partially contradicts the findings of Greiff et al. (2016) that higher intervention frequency (interpreted as less efficient) was associated with lower CPS performance scores. Thus, further research is needed to clarify the theoretical status of goal-directed behavior and its relationship to CPS performance.

2.2.3.5 Summary

The empirical findings presented above illustrate the close relationship between performance and process measures of CPS. As a general pattern, process measures obtained in the knowledge acquisition phase are more closely related to performance measures of the same phase compared to performance measures of the knowledge application phase. Findings on VOTAT and NOTAT demonstrate that effective strategic exploration behaviors are related to higher CPS performance scores (e.g., Greiff et al., 2016; Kröner et al., 2005; Vollmeyer et al., 1996). Less clear is the empirical evidence on the relationship between CPS performance and time on task and the efficiency of students' exploration behavior, operationalized as interactions with the system. Whereas some studies have suggested that more time on task (Scherer et al., 2015) and more interventions (Eichmann, Goldhammer, et al., 2020) are related to higher CPS performance scores, other studies have shown that there is an optimal time on task related to higher CPS performance and that a high intervention frequency is related to a lower CPS performance (Greiff et al., 2016). Overall, the relationship between CPS performance and the efficiency of students' exploration behavior requires further research.

With respect to the theoretical requirements of problem solving (see Mayer, 1992; Mayer & Wittrock, 2006; Novick & Bassok, 2005), the research presented here shows that process measures have only captured some of the processes involved in forming a problem representation and deriving a solution to that problem. This theoretical consideration is supported by empirical findings that consider the joint predictive power of multiple process measures. When the process measures VOTAT, NOTAT, time on task, and intervention frequency were considered in concert, about 75% of the variance in performance in the knowledge acquisition phase and about 50% of the variance in performance in the knowledge application phase were explained (Greiff et al., 2016). The substantial amount of unexplained variance in both CPS phases underscores the need for further research on problem solvers' (strategic) behaviors captured in process measures and their relation to successful problem solving.

Therefore, one major goal of the present dissertation project was to investigate the process of how participants record their conclusions in the causal model when exploring the system in the knowledge acquisition phase (see study 2, Weise et al., 2022). This largely unstudied type of process data offers insight into when students identify what type of effects (e.g., direct effects, eigendynamic effects). Based on theoretical considerations, the chronological order in which different types of effects are identified should play an important role in forming an accurate problem representation when a system is composed of different types of effects that potentially confound each other (e.g., direct and eigendynamic effects). More specifically, identifying eigendynamic effects before making inferences about direct effects avoids confounding the two effects and therefore represents an effective strategic behavior (Beckmann, 1994). Similar to effective CPS behaviors such as VOTAT, the strategic behavior to identify eigendynamic effects early should be positively related to CPS performance. As eigendynamic effects are an important part of the dynamic characteristic of CPS tasks (e.g., Dörner et al., 1983; Funke, 1992), investigating the strategic behavior to identify eigendynamics early further clarifies how students interact with complex and dynamic systems and sheds light on the relationship between CPS behaviors and CPS performance (see study 2, Weise et al., 2022).

2.3 Intelligence and CPS

At a conceptual level, intelligence and CPS have much in common: Problem solving is part of widely accepted definitions of intelligence, such as, the mainstream definition proposed by Gottfredson (1997) and an earlier survey of scientists (Sternberg, 1997). In addition, both intelligence and CPS tasks require participants to identify underlying structures (i.e., rules) and integrate information (Carpenter et al., 1990; Funke, 2001). However, because CPS tasks typically involve dynamic aspects, are opaque, and may have multiple (or even conflicting) goals, these tasks require the ability "of self-regulated learning as well as the ability to adapt the problem-solving process to a changing environment by continuously processing feedback information" (Wirth & Klieme, 2003, p. 329). In contrast, intelligence test tasks are typically comprised of static, transparent, and well-defined problems and do not involve any interaction (Süß & Kretzschmar, 2018). In the following two sections, the relationship between intelligence and CPS performance and between intelligence and CPS process measures will be reviewed.

2.3.1 Intelligence and CPS performance measures

As intelligence and CPS are closely related on a theoretical level, the empirical relationship between intelligence and CPS performance measures has been investigated extensively. Early studies using classical CPS assessment tools (e.g., Tailorshop) predominantly found correlations close to zero between intelligence and CPS performance scores (e.g., Brehmer, 1992; Dörner et al., 1983; Putz-Osterloh, 1981). These results led researchers to assume that intelligence and CPS were largely independent constructs (Putz-Osterloh, 1985; Dörner et al., 1986). Similarly, a review of eleven studies revealed that there was no support for the assumption that intelligence predicted performance in classical CPS tasks (Kluwe et al., 1991). However, the authors of the review called attention to psychometric problems related to the objectivity, reliability, and validity of the CPS performance scores measured by classical CPS assessment tools, thus potentially interfering with a meaningful investigation of the relationship between intelligence and CPS.

These findings of a weak to non-existent relationship between intelligence and CPS and the critical considerations involved are summarized in several hypotheses. The different-demands hypothesis assumes that the close to zero correlations between intelligence and CPS performance were caused by more complex cognitive processes of CPS tasks (e.g., imposed by intransparency, dynamics, polytely, and the interactive character) compared to intelligence tests (Rigas & Brehmer, 1999). However, the different-demands hypothesis was challenged by a study that showed a differential pattern of substantial correlations between intelligence and two CPS assessment tools, which were similar in terms of their specific CPS characteristics but posed different cognitive demands on the problem solver ($r = .33$ to $r = .82$, p's $< .05$; Gonzalez et al., 2005).

The low-reliability hypothesis assumes that the low correlations between intelligence and CPS were caused by the low reliability of the CPS performance scores

(Rigas et al., 2002). On the one hand, the authors attributed the low reliability of CPS performance scores to poor psychometric properties of the CPS assessment tools, which needed to be improved. On the other hand, the authors attributed the lower reliability to the fundamental characteristics of CPS tasks, such as complexity, intransparency, and polytely, which should not be changed in order to maintain their alleged ecological validity. Empirical support for the low-reliability hypothesis stems from several studies that utilized psychometrically improved versions of classical CPS assessments. These studies exhibited substantial, mostly medium-sized correlations between intelligence and CPS performance scores (e.g., Süß et al., 1991: $.31 < r < .47$, p's $< .05$).

A further explanation for the so-called fuzzy correlation pattern between intelligence and CPS performance is related to problem solvers' prior knowledge. Specifically, the Elshout-Raaheim hypothesis (Leutner, 2002; see Elshout, 1987; Raaheim, 1988) postulates a moderating effect of prior knowledge on the relationship between problem solving and intelligence: For low levels of prior knowledge, the hypothesis predicts lower correlations between both constructs; for medium levels of prior knowledge, higher correlations are predicted; and for high levels of prior knowledge, the hypothesis predicts again lower correlations between intelligence and CPS performance, resulting in an inverted U-shaped curve of correlation coefficients. Conceptually, the Elshout-Raaheim hypothesis proposes that intelligence is particularly important for solving complex problems when participants have intermediate levels of prior knowledge; when no prior knowledge is present, higher intelligence is not sufficient to solve problems successfully; and when high levels of prior knowledge are present, this knowledge can be used to solve the problem, thereby reducing the influence of intelligence. Empirically, Leutner (2002, second study) examined the Elshout-Raaheim hypothesis, relying on a repeatedly administered classical CPS assessment tool and assumed that prior knowledge increased from administration trial to administration trial. However, the general pattern of the correlation coefficients of Leutner's study was not in line with the Elshout-Raaheim hypothesis. In addition, about half of the correlation coefficients were less than zero, even though the Elshout-Raaheim hypothesis predicts positive correlation coefficients for all levels of prior knowledge.

Since the review on the intelligence-CPS relationship of Kluwe et al. (1991), research in the field of CPS has undergone significant developments. The psychometric properties of classical CPS assessment tools have been significantly improved (e.g., Tailorshop; Süß et al., 1993) and new assessment tools that are based on the LSE and FSE frameworks, so-called single complex systems (SCS, e.g., MultiFlux; Kröner et al., 2005) and multiple complex systems (MCS, e.g., MicroDYN; Greiff et al., 2012) have been developed. Given these advances in the assessment of CPS, a recent and comprehensive meta-analysis summarized the evidence on the relationship between intelligence and CPS over the past several decades (Stadler et al., 2015). This meta-analysis also revisited the different-demands hypothesis, the low-reliability hypothesis, and the Elshout-Raaheim hypothesis. The study considered the operationalization of CPS and the operationalization of intelligence as potential moderators of the intelligence-CPS relationship. All studies included (k = 60), the meta-analysis reported an uncorrected mean relationship of $M(g)$ = .43 (Stadler et al., 2015). As concerns the moderation analyses, results showed that if intelligence was operationalized as general intelligence, the relationship with CPS decreased to $M(g)$ = .36, and if intelligence was operationalized as reasoning, the relationship increased to $M(g)$ = .47. However, the meta-analysis did not consider the combination of psychometrically sound MCS assessment tools and broad operationalizations of general intelligence. Recent studies that used MCS assessment tools and broad operationalizations of general intelligence reported numerically higher correlations of $.65 < r < .85$ than suggested by the meta-analysis (Danner et al., 2011; Kretzschmar et al., 2016; Kröner et al., 2005; Lotz et al., 2016).

Further findings of the meta-analysis revealed that if only psychometrically more convincing MCS assessment tools of CPS were considered, the relationship increased to $M(g)$ = .59, compared to a mean correlation of $M(g)$ = .34 when classical CPS assessments tools were considered (Stadler et al., 2015). This finding was interpreted as support for the low-reliability hypothesis, which assumes that low correlation coefficients of the intelligence-CPS relationship are caused by poor reliability estimates, as can be found in classical assessment tools of CPS. However, when the intelligence-CPS correlations were corrected for plausible average reliability coefficients of different CPS measures, the general pattern of effect sizes did not change, which called the low-reliability hypothesis into question.

A possible explanation for the moderating effect of CPS operationalization on the CPS-intelligence relationship could also be the different-demands hypothesis, assuming that differences in the complexity of cognitive processes between CPS and intelligence lead to relatively low correlations. In this line, classical CPS assessment tools with higher levels of complexity showed a weaker relationship with intelligence, whereas MCS assessments with lower levels of complexity showed a stronger relationship with intelligence (Stadler et al., 2015). However, as the authors of the meta-analysis point out, the reliability of CPS performance measures obtained from different assessment instruments varies, making it difficult to properly examine the different-demands hypothesis.

The authors of the meta-analysis also discussed the Elshout-Raaheim hypothesis by comparing the relationships between intelligence and two different CPS measures, which are both based on the LSE framework: SCS and MCS both use fictive semantic embeddings, so performance should not be affected by domain-specific prior knowledge. Accordingly, Stadler et al. (2015) assumed that the two CPS measures should have an equally strong relationship with intelligence. As the meta-analysis showed different correlational effect sizes for SCS and MCS assessment instruments with intelligence, the authors rejected the Elshout-Raaheim hypothesis. However, this finding regarding the Elshout-Raaheim hypothesis does not reflect the full assumptions of the hypothesis, which predicts an inverted U-shaped pattern of correlation coefficients for different levels of prior knowledge. Specifically, the meta-analysis neglects medium and high levels of prior knowledge and does not account for other differences in assessment instruments that may well account for the reported difference in the relationship to intelligence. In general, it seems difficult to cover the assumptions of the Elshout-Raaheim hypothesis in a moderation analysis of a meta-analytic study.

In summary, a vast body of research shows a high degree of variation in the intelligence-CPS relationship, ranging from zero correlations (e.g., Putz-Osterloh, 1981) to high correlation coefficients (e.g., Kröner et al., 2005; see Stadler et al., 2015, for an overview). Theories, such as the low-reliability hypothesis, the different-demands hypothesis, and the Elshout-Raaheim hypothesis try to shed light on the "fuzzy relationship between intelligence and problem solving" (Leutner, 2002, p. 685). One major goal of the present dissertation project (see study 1, Weise et al., 2020) was to further clarify the moderating effect of prior knowledge on the intelligence-CPS relationship as assumed by the Elshout-Raaheim hypothesis. Specifically, we reinvestigated the Elshout-Raaheim hypothesis for the first time in a psychometrically strong MCS assessment instrument of CPS. The structure of the MCS assessment tools, consisting of multiple short tasks of structural equivalence, allows for the empirical investigation of an increase in prior knowledge over CPS tasks. Overall, identifying moderators of the intelligence-CPS relationship contributes to a deeper understanding of how intelligence might work and the conditions under which it enhances successful problem solving.

2.3.2 Intelligence and CPS process measures

As previously elaborated, process measures play an important role in CPS research. For example, they allow researchers to gain insight into the CPS behaviors that are linked to enhanced problem-solving proficiency (e.g., Greiff et al., 2016) and to better understand how differences in CPS behaviors manifest in differences in CPS performance across groups or countries (e.g., Eichmann, Goldhammer, et al., 2020; Wüstenberg, Greiff, et al., 2014). Unlike the relationship between intelligence and performance measures of CPS, the relationship between intelligence and process measures of CPS has not been extensively researched. In addition, the relationship of each process measure to intelligence requires specific theoretical considerations because process measures differ in their theoretical background.

2.3.2.1 Intelligence and CPS process measures of effectiveness

Effective CPS behaviors in the system exploration phase are behaviors that enable the problem solver to draw the correct conclusions about the system structure. Recognized effective behaviors include VOTAT and NOTAT, which allow for unconfounded experiments, effectively isolating direct and eigendynamic effects. As outlined in the first chapter (*2.1 Intelligence*), general intelligence can be understood as a broad mental capability that is positively related to tasks that require cognitive processes such as reasoning, planning, solving problems, and learning from experience (Gottfredson, 1997; Spearman, 1904). Therefore, higher intelligence should be associated with more effective exploration behavior as, for instance, captured by the process measures VOTAT and NOTAT. This theoretical consideration is supported by multiple empirical findings: The relative frequency of the VOTAT application in a SCS assessment tool of CPS (i.e., MultiFlux) correlated substantially with intelligence indicated by reasoning ($r = .41$, *p*-value not reported; Kröner et al., 2005; $r = .36$, *p*-value not reported; Kröner, 2001). In an MCS assessment tool of CPS (i.e., MicroDYN) intelligence predicted an aggregated, dichotomous VOTAT score over the set of CPS tasks (credit for a task was given if

VOTAT was applied to every input variable; $\beta = .56$, $p < .001$; Wüstenberg, Stadler, et al., 2014). Moreover, in a recent comprehensive study, Lotz et al. (2017) showed that the initial relative VOTAT application rate ($r = .48$, $p < .05$) as well as the increase of the VOTAT application rate ($r = .21$, $p < .05$) were significantly related to a broad factor of intelligence. Similarly, for CPS tasks comprising direct and eigendynamic effects, the initial relative application rate of NOTAT was significantly related intelligence ($r = .27$, $p < .05$; Lotz et al., 2017). Therefore, Lotz et al. (2017) showed that not only application, but also adaptation from less to more effective behavior is related to intelligence. As the CPS tasks used in the MCS assessment tool differed in their problem structure, the VOTAT application rate cannot be directly compared across tasks. However, the adaptation of effective VOTAT behavior is similar to cognitive learning processes, which are likewise associated with intelligence (e.g., Guthke & Stein, 1996; Jensen, 1989; Sternberg, 2013). In summary, intelligence is associated with the use of effective strategic exploration behaviors and with the adaptation from less to more effective exploration behaviors while working on CPS tasks. This finding points to an interpretation that "intelligence manifests itself in specific strategic actions during the exploration phase" (Lotz et al., 2017, p. 111; see also Veenman et al., 2004), such as the effective behaviors VOTAT and NOTAT.

The interpretation of effective behaviors as manifestations of intelligence when working on CPS tasks suggests that these CPS behaviors may constitute underlying factors that explain the intelligence-CPS relationship. This is also reflected in empirical findings showing that the relative use of VOTAT and NOTAT mediated the relationship between a broad factor of general intelligence and CPS performance for both the knowledge acquisition phase and knowledge application phase (Lotz et al., 2022). Similarly, in previous studies, VOTAT mediated the relationship between the three predictors of fluid intelligence, scientific reasoning, and learning orientation and the outcome variable CPS performance (Wüstenberg, Stadler, et al., 2014; see also Grežo & Sarmány-Schuller, 2022). Taken together, the manifestation of intelligence in the form of effective behaviors, such as VOTAT, appears to be a relevant mediator that sheds further light on the relationship between intelligence and CPS.

Importantly, the strategic behaviors VOTAT and NOTAT only partially account for the variance in CPS performance (Greiff et al., 2016). Thus, there are most likely further effective CPS behaviors that represent manifestations of general intelligence when solving complex problems and that may mediate the intelligence-CPS relationship.

Therefore, a major goal of the present dissertation project was to investigate the strategic behavior to identify eigendynamic effects early, its relationship to intelligence, and whether this specific CPS behavior mediates the intelligence-CPS relationship (see study 2). Investigating the strategic behavior to identify eigendynamic effects early may potentially provide important insights into students' behavior when dealing with eigendynamic effects, which is a distinct characteristic of CPS tasks. Moreover, considering a potential mediating effect of the strategic behavior to identify eigendynamic effects early may further contribute to our understanding of how intelligence and CPS are related to one another.

2.3.2.2 Intelligence and CPS process measures of efficiency

Efficient CPS behaviors in the knowledge acquisition phase can be understood as behaviors that do not exceed the necessary number of interactions required to understand the underlying structure of the problem (Eichmann, Greiff, et al., 2020). The number of non-necessary exploration steps when applying VOTAT for each input variable, as investigated in this dissertation project (see study 3), represents a process measure that adequately captures students' efficiency when exploring the system. To date, the relationship between process measures that indicate an efficient exploration behavior and intelligence has not been covered by research in the field of CPS. From a theoretical perspective, general intelligence includes cognitive processes that are related to planning (Gottfredson, 1997) and therefore intelligence should be positively related to students' planned and efficient behavior, indicated by the number of non-necessary exploration steps when applying VOTAT. Moreover, as intelligence is related to adapting more effective behavior over the course of CPS tasks (Lotz et al., 2017), intelligence should also be related to the adaption of a more efficient behavior over the course of CPS tasks.

2.3.2.3 Summary

The relationship between CPS behaviors and intelligence and the potentially mediating effect of these CPS behaviors on the intelligence-CPS relationship is an important aspect in understanding how intelligence manifests itself when solving complex problems (Lotz et al., 2017; Veenman et al., 2004). For effective CPS behaviors (e.g., VOTAT), a significant correlation to intelligence (e.g., Kröner et al., 2005; Lotz et al., 2017) as well as a mediating effect on the intelligence-CPS relationship has been shown (Lotz et al., 2022; Wüstenberg, Stadler, et al., 2014). However, there are further effective CPS behaviors, for which the relation to intelligence and their potential mediating effect remain unclear, such as the behavior to identify eigendynamic effects early when exploring a system, a strategic behavior that was investigated in the scope of this dissertation project (see study 2). Therefore, we hypothesized that the described strategic behavior would be positively related to intelligence and mediate the relationship between intelligence and CPS. Regarding CPS behaviors that indicate efficiency during the knowledge acquisition phase, the relationship to intelligence is still unknown. In this vein, it was a major goal of the present dissertation project to identify latent classes of students according to the effectiveness and efficiency of their VOTAT exploration behavior and to inspect differences in intelligence among them (see study 3). As outlined in the previous sections, intelligence should be positively related to effective and efficient exploration behavior. In perspective, clarifying the relationship between process measures of CPS, performance measures of CPS, and intelligence is an important keystone in understanding the relationship between intelligence and CPS and can provide guidance on how to effectively promote complex problem solving.

2.4 Research aims of the dissertation project

Based on the theoretical framework and the empirical findings outlined in the previous sections, three studies were conducted that formed the core of this dissertation project and investigated the relationship between intelligence and CPS. The primary goals of this dissertation project were to shed light on the relationship between intelligence and CPS by examining the factors that influence the strength of the association between the two constructs, and to examine observable patterns of behavior in which intelligence manifests itself when working on CPS tasks and through which it translates into CPS performance. In addition, in study 4, we evaluated and implemented a very time-efficient intelligence screening instrument that expands the possibilities of intelligence assessment in (large) studies when time is scarce.

In the first study, we investigated the Elshout-Raaheim hypothesis (Leutner, 2002; see Elshout, 1987; Raaheim, 1988), which predicts a curvilinear moderating effect of prior knowledge on the relationship between intelligence and CPS. For the first time, the

Elshout-Raaheim hypothesis was investigated in a psychometrically strong assessment tool of CPS, the MCS-based instrument MicroDYN. The separation of the knowledge acquisition phase and the knowledge application phase in the aforementioned assessment tool provided the opportunity to examine the Elshout-Raaheim hypothesis separately for the two theoretically and empirically distinguishable CPS phases. Moreover, we considered the relative frequency of the strategic exploration behavior VOTAT as an indicator of prior (strategic) knowledge concerning the exploration process of CPS tasks.

In the second study, we investigated for the first time the strategic behavior to identify eigendynamic effects in an early exploration step, which is an effective strategic behavior if the system comprises different types of effects (i.e., eigendynamic effects along with other effects). In addition, we examined whether the newly examined strategic behavior was related to intelligence, CPS performance, and whether it mediated the intelligence-CPS relationship. Exploring strategic behaviors other than VOTAT and NOTAT that contribute to the CPS-intelligence relationship is a keystone to clarify and promote CPS.

In the third study, we investigated the effectiveness and efficiency of students' exploration behavior in CPS tasks and its relation to intelligence and CPS performance by using a person-centered approach. The theoretical status of efficiency of students' exploration behavior and its relationship to intelligence and CPS performance is still unresolved. Therefore, this study offered a new approach to operationalize efficiency and shed further light on its relationship to the two constructs mentioned above.

The fourth study focused on the evaluation and extension of the *mini-q* (Baudson & Preckel, 2016), an intelligence screening instrument that can be administered in just three minutes. Time-efficient and well-evaluated intelligence screenings play an important role in extensive (large-scale) research projects when time is scarce. In the present study, the *mini-q* was complemented by a newly designed parallel version and both test versions were compared and evaluated. The new parallel version of the screening instrument provides an important extension of the test's application possibilities, such as group testing and test repetition, and allows the inspection of the instrument's parallel test reliability.

3 Empirical Studies

3.1 Study 1: Weise, Greiff, & Sparfeldt (2020)

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving – Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502.

In this study, we sought to re-examine the moderating effect of prior knowledge on the relationship between CPS and intelligence, thereby clarifying the fuzzy relationship between intelligence and problem solving.

3.1.1 Theoretical Background and Hypotheses

Although the relationship between CPS and intelligence has been investigated repeatedly (e.g., Dörner & Kreuzig, 1983; Kröner, et al., 2005; Stadler, et al., 2015), the role of prior knowledge and its moderating influence on the relationship between the two constructs remain largely open questions. The Elshout-Raaheim hypothesis (Leutner, 2002; see Elshout, 1987; Raaheim, 1988) fills in this missing link by postulating how prior knowledge moderates the relationship between intelligence and CPS. Specifically, the hypothesis predicts that with increasing prior knowledge, the pattern of correlation coefficients between intelligence and CPS follows an inverted U-shaped curve. Thus, for low levels of prior knowledge, the hypothesis predicts lower correlations between both constructs; for medium levels of prior knowledge, higher correlations are predicted; and for high levels of prior knowledge, the hypothesis predicts lower correlations. Regarding these theoretical considerations, Leutner (2002, second study) examined the Elshout-Raaheim hypothesis relying on a repeatedly administered, classical computer-based CPS scenario, assuming that prior knowledge would increase from trial to trial (three trials per day for three consecutive days, followed by a delayed trial; all together ten trials). The observed pattern of correlation coefficients was in accordance with the hypothesis within each of the three days. However, seven out of ten correlation coefficients were equal to or less than zero, even though zero or positive correlation coefficients were predicted for all levels of prior knowledge. In addition, the overall pattern of correlation coefficients across all ten trials was not consistent with the Elshout-Raaheim hypothesis, so the results were inconclusive.

Recent advances in the assessment of CPS and particularly the emergence of the minimal complex systems approach (MCS; Greiff et al., 2012) allow an advanced reconsideration of the Elshout-Raaheim hypothesis. Specifically, the present study investigated the Elshout-Raaheim hypothesis (Leutner, 2002; see also Elshout, 1987; Raaheim, 1988) using the MCS approach MicroDYN (Greiff et al., 2012) for the two distinct CPS phases knowledge acquisition and knowledge application. In accordance with prior studies, the variation of prior knowledge was operationalized by the assumption of an increase in prior (strategic) knowledge from CPS task to CPS task in both phases. As an indicator of prior strategic knowledge, the application of VOTAT in the knowledge acquisition phase was inspected for each task. In accordance with studies in the field of scientific reasoning and CPS (e.g., Lotz et al., 2017; Schauble, 1996; Vollmeyer, et al., 1996), we expected an increase in prior strategic knowledge across tasks. In summary, this study extended previous research investigating the Elshout-Raaheim hypothesis by (1) relying on a psychometrically strong CPS assessment, (2) administering the CPS assessment in a single session rather than across different days as Leutner (2002) did, (3) examining both CPS phases, and (4) empirically testing the hypothesized increase in prior knowledge:

Hypothesis 1. We expected that item-based correlation coefficients between intelligence and single CPS tasks of the knowledge acquisition phase, ordered in the administered CPS task sequence, would follow an inverted U-shaped curve.

Hypothesis 2. We expected that item-based correlation coefficients between intelligence and single CPS tasks of the knowledge application phase, ordered in the administered sequence, would follow an inverted U-shaped curve.

3.1.2 Methods

Participants and Procedure. The sample consisted of *N* = 495 German high school students (53.33% female; age *M* = 16.40, *SD* = 0.94 years). The students attended two academic-tracked school types (Gymnasium, Gesamtschule). The participation rate reached 87% and students were not graded or rewarded for their participation. Data collection was administered by trained experimenters within two regular school lessons (for more details, see Weise et al., 2020).

Instruments. Intelligence was assessed with the Berlin Intelligence Structure test – Form 4 (BIS-4; Jäger et al., 1997; subtests: Figural analogies, Crossing out letters, Charkow, City map, Number sequences, X greater, Estimation, Story, Fact-opinion, and Verbal-analogies). CPS was assessed by the entirely computer-based program MicroDYN (Greiff et al., 2012), consisting of nine independent CPS tasks, each of them was comprised of a knowledge acquisition phase and a knowledge application phase. In this study, we focused on the first five MicroDYN tasks with only direct and no eigendynamic effects, because for this set of tasks the use of VOTAT is most effective and a steady increase in strategic behavior could be expected.

Concerning the VOTAT strategy as an indicator for prior strategic knowledge, a VOTAT score was retrieved by analyzing each participant's computer-generated log-file of the knowledge acquisition phase (see Lotz et al., 2017, for details). In each exploration step, credit for VOTAT was given (coded as 1), if only one input variable was varied and the other input variables were held at zero; otherwise, no credit was given (coded as 0). For each MicroDYN task, we computed the relative VOTAT frequency.

Analyses. Analyses were conducted with the statistical package Mplus 7.11 (Muthén & Muthén, 1998–2012) and Microsoft Excel. Manifest correlation coefficients between each single CPS task performance score of the two phases and intelligence (*g*) were calculated in Mplus using the "type is complex" option. By doing so, the students' clustering in classrooms was considered and the potential effects of clustering on standard errors and χ^2-statistics could be controlled for. Concerning the intelligence model, the manifest scores of the ten intelligence subtests of the BIS loaded on the corresponding first-order content factors of verbal (VBIS), numerical (NBIS), and figural (FBIS) intelligence, which, in turn, loaded on the second-order *g* factor (*g*BIS), indicating the common higher-order structure of intelligence. The pattern of the correlation coefficients between CPS performance and intelligence was evaluated in three steps, separately for the knowledge acquisition phase and the knowledge application phase. In the first step, we descriptively inspected the correlation coefficients, corresponding standard errors, and the 68% confidence intervals. Non-overlapping confidence intervals were interpreted as indicators for substantial differences in the corresponding correlation coefficients. In a second step, the five correlation coefficients were compared pairwise within each of the two phases (ten pairwise comparisons per phase). This was done in an additional model in Mplus by testing the pairwise differences of path coefficients against zero ($p < .05$). In a third step, the distribution of the correlation coefficients for each CPS phase was analyzed descriptively using Microsoft Excel's graphing tools including the "Trendline"-option to approximate a curve to the correlation coefficients.

3.1.3 Results

Before testing the Elshout-Raaheim hypothesis, we analyzed the relative frequency of the VOTAT use across the five CPS tasks to investigate whether it increased across tasks, indicating increasing levels of prior strategic knowledge. The results showed a steady increase in the relative VOTAT frequencies across all five MicroDYN tasks (task 1/2/3/4/5: 47%/58%/63%/67%/71%; see Lotz et al., 2017, for details).

Concerning Hypothesis (1), the correlation coefficients between intelligence and CPS task performance in the knowledge acquisition phase were analyzed in a first step for the CPS tasks 1, 2, 3, 4, and 5: $r_1 = .41$ ($SE_1 = .07$), $r_2 = .44$ ($SE_2 = .07$), $r_3 = .64$ ($SE_3 = .06$), $r_4 = .44$ ($SE_4 = .07$), $r_5 = .44$ ($SE_5 = .08$). An inspection of the confidence intervals (see Fig. 3) revealed that the confidence interval of the correlation coefficient of task 3 did not overlap with the confidence intervals of the correlation coefficients of the remaining CPS tasks, whereas the confidence intervals of the correlation coefficients of tasks 1, 2, 4, and 5 were overlapping. This pattern of results indicated that the correlation of task 3 was higher than the correlation coefficients of the previous and subsequent CPS tasks.

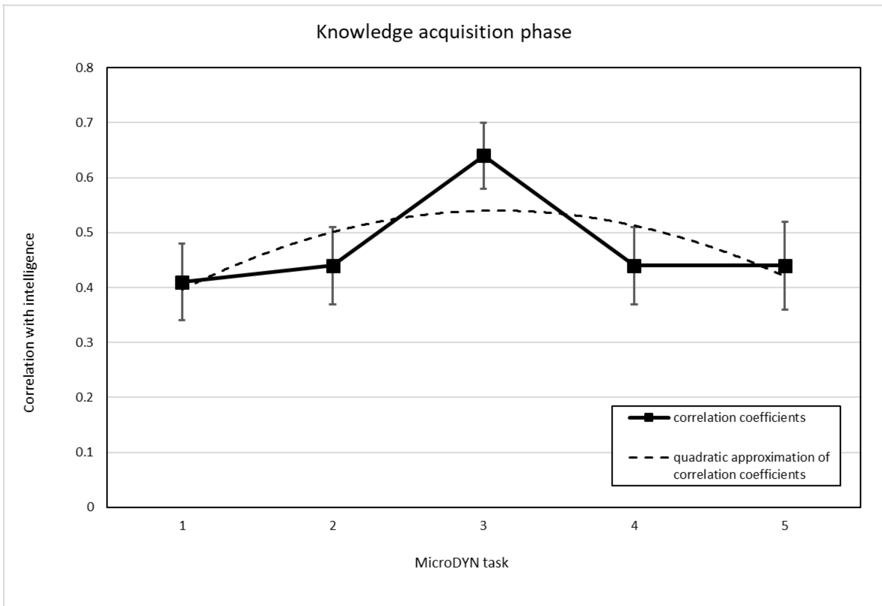

Figure 3. Distribution of correlation coefficients and 68% confidence intervals (±1 standard error) between CPS tasks and *g* for the knowledge acquisition phase augmented by a quadratic equation curve (Weise et al., 2020).

In a second step to analyze the described pattern of correlations, the pairwise comparisons of the correlation coefficients again showed a significant difference between the third (and highest) intelligence-CPS correlation coefficient and the other four correlation coefficients of the knowledge acquisition phase ($p < .01$ for all four comparisons). The correlation coefficients r_1, r_2, r_4, and r_5 did not differ statistically significantly from each other ($p > .55$ for all six comparisons). In a third step, a quadratic approximation of the five mentioned intelligence-CPS correlations using Microsoft Excel showed that the following equation fitted the data best: $y = –.03x^2 + .20x + .23$ (see Fig. 3). The quadratic approximation accounted for 44% of the variance in the correlation coefficients, whereas a linear approximation accounted for only 1% of the variance. In summary, the results indicated an inverted U-shaped pattern of the correlation coefficients in the knowledge acquisition phase and provided support to Hypothesis (1).

Concerning Hypothesis (2), the correlation coefficients of intelligence and CPS performance in the knowledge application phase of CPS tasks 1, 2, 3, 4, and 5 reached $r_1 = .37$ ($SE_1 = .10$), $r_2 = .41$ ($SE_2 = .08$), $r_3 = .53$ ($SE_3 = .07$), $r_4 = .62$ ($SE_4 = .06$), and $r_5 = .56$ ($SE_5 = .08$). Thus, the correlation coefficients increased numerically from task 1 to task 4 and decreased from task 4 to task 5 (see Fig. 4). An inspection of the confidence intervals showed that the confidence intervals of task 1 and task 2 did not overlap with the confidence interval of task 4, and the confidence interval of task 1 did not overlap with the confidence interval of task 5, whereas the remaining confidence intervals were overlapping. This result pattern indicated that the correlations of tasks 1 and 2 were lower than the correlation coefficient of task 4 and that the correlation of task 1 was lower than the correlation of task 5.

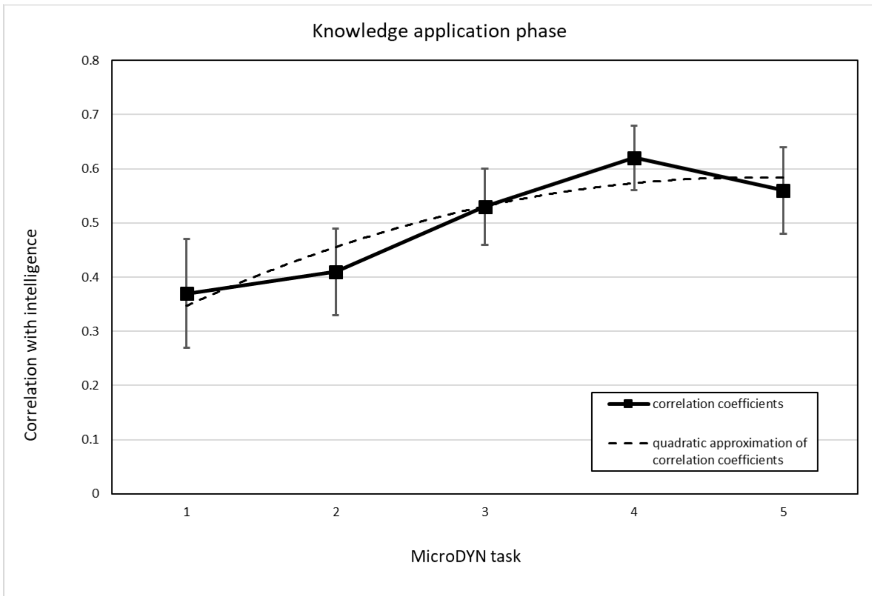

Figure 4. Distribution of correlation coefficients and 68% confidence intervals (±1 standard error) between CPS tasks and *g* for the knowledge application phase augmented by a quadratic equation curve (Weise et al., 2020).

In a second step, the pairwise comparisons of the correlation coefficients showed a statistically significant difference between the first and the fourth intelligence-CPS correlation coefficient ($p = .02$), as well as the second and the fourth correlation coefficient ($p < .01$). None of the remaining eight comparisons of the correlation coefficients revealed statistically significant results ($p > .08$). Concerning the third step, a quadratic approximation of the intelligence-CPS correlations showed the following equation to fit the data best: $y = .02x^2 + .16x + .21$. The quadratic approximation explained 88% of the variance of the correlation coefficients, whereas a linear approximation accounted for 79% of the variance. In summary, the results indicated an inverted U-shaped pattern of the correlation coefficients in the knowledge application phase and provided support to Hypothesis (2).

3.1.4 Discussion

The present study revealed three main findings: First, the predicted increase in prior knowledge in the knowledge acquisition phase, indicated by the relative frequency of the VOTAT strategy, was supported by our data. While our results indicated an increase in prior strategic knowledge across CPS tasks, the development of strategies and test-specific skills was discussed in a recent meta-analysis (Scharfen et al., 2018). Second, and regarding Hypothesis (1), the pattern of correlation coefficients for intelligence and CPS tasks in the knowledge acquisition phase followed an inverted U-shaped curve as predicted. Third, and concerning Hypothesis (2), the pattern of correlation coefficients in the knowledge application phase also followed an inverted U-shaped curve. Taken together, the increase in prior knowledge across CPS tasks and the patterns of correlation coefficients in both CPS phases support the assumption of a moderation of prior strategic knowledge on the relationship between CPS and general intelligence, as assumed by the Elshout-Raaheim hypothesis.

Our findings of an inverted U-shaped pattern of correlation coefficients suggest that for both knowledge acquisition and knowledge application in complex problem scenarios, intelligence plays a more important role when participants have acquired medium levels of prior knowledge and a less important role when participants have acquired either rather low or rather high levels of prior knowledge. Similar to our results, there have been findings of an inverted U-shaped pattern when investigating the moderating effect of prior knowledge on the efficacy of help in learning tasks (Seufert, 2003). If one postulates that increasing prior knowledge reduces the difficulty and complexity of a task, our results also correspond with Jensen's (1980, p. 32) conclusion of higher correlations between learning and intelligence "when the material to be learned is of moderate difficulty and complexity. If a learning task is too complex, everyone … falls back on simpler processes such as trial-and-error".

Furthermore, this study extended the Elshout-Raaheim hypothesis that specified the relation between problem solving and intelligence for (domain-specific) prior knowledge to strategic prior knowledge. In future studies, it would be desirable to test the hypothesis with a larger set of tasks. Taken together, this study was the first to examine the Elshout-Raaheim hypothesis in an MCS approach of CPS, thereby contributing to our understanding of the conditions under which intelligence enhances successful problem solving. In perspective, the identification of moderators further clarifies the specific mechanisms of intelligence in action and their relevance for instructional design and structuring effective learning environments.

3.2 Study 2: Weise, Greiff, & Sparfeldt (2022)

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2022). Focusing on eigendynamic effects promotes students' performance in complex problem solving: A log-file analysis of strategic behavior. *Computers & Education*, *189*, 104579.

In this study, we investigated the effective strategic behavior to identify eigendynamic effects in an early exploration step when exploring complex systems, its relationship to CPS performance and intelligence, and its mediating role in the CPS-intelligence relationship.

3.2.1 Theoretical Background and Hypotheses

In CPS tasks, so-called eigendynamic effects are an important feature reflecting increases or decreases in outcome variables over time without a person's intervention (e.g., Dörner, 1980; Funke, 2001; Stadler, Niepel, & Greiff, 2016). In prior studies, researchers investigated strategic behaviors when dealing with eigendynamic effects while primarily focusing on how students explored the system (i.e., NOTAT; Greiff et al., 2016; Lotz et al., 2017), but not whether the students drew the correct conclusions based on the information they gathered about eigendynamic effects. In this study, we investigated students' exploration behavior, specifically focusing on the strategic behavior to identify eigendynamic effects early, which is an effective strategic behavior if a system comprises different types of effects (i.e., eigendynamic effects along with other effects; see also Beckmann, 1994). Our analysis expands on recent studies by analyzing the log-file data of students' actions in the causal diagram, which provide information about when students have come to which conclusions about the system. In previous studies, strategic behaviors such as VOTAT and NOTAT were related to CPS performance in both the knowledge acquisition and the knowledge application phase (Greiff et al., 2016; Greiff, Wüstenberg, & Avvisati, 2015; Vollmeyer et al., 1996). Moreover, the relationship between intelligence and CPS performance was mediated by strategic exploration behaviors, thereby contributing to the understanding of the relationship between the two constructs (Grežo & Sarmány-Schuller, 2022; Wüstenberg, Stadler, et al., 2014). The log-file based analyses of the strategic behavior to identify eigendynamic effects early offer new insights into how students interact with complex and dynamic task environments. Furthermore, investigating new strategic behaviors is an important keystone to understand how intelligence manifests itself in observable, specific behavior when solving complex problems (e.g., Lotz et al., 2017). Overall, the aims of this study were as follows:

Exploratory analyses. First, we exploratively and descriptively analyzed whether students identified the correct eigendynamic effect in the causal diagram and, if so, in which action.

Hypothesis 1. We expected that students who correctly identified the eigendynamic effect should perform better regarding the correct identification of the direct effects within the same task than students who did not identify the eigendynamic effect. Furthermore, we expected that students who correctly identified the eigendynamic effect earlier should perform better regarding the identification of the direct effects of the respective task than students who identified the eigendynamic effect later.

Hypothesis 2. We expected that the strategic behavior to identify eigendynamic effects early should be related to an increased overall performance in the knowledge acquisition phase, and therefore we expected a similar pattern of results as in Hypothesis (1).

Hypothesis 3. We expected that the strategic behavior to identify eigendynamic effects early should be associated with a higher performance score in the knowledge application phase of the same task. Again, we expected a similar pattern of results as in Hypothesis (1).

Hypothesis 4. We assumed that the strategic behavior to identify eigendynamic effects early should be related to an increased overall performance score in the knowledge application phase and therefore, we expected a similar pattern of results as in the previous hypotheses.

Hypothesis 5. We assumed that students with higher intelligence test scores should show more effective strategic behaviors and higher CPS performance scores compared to students with lower intelligence test scores. Furthermore, we assumed that the strategic behavior to identify eigendynamic effects and the strategic behavior to identify eigendynamic effects in an early exploration step should mediate the intelligence-CPS relationship in the knowledge acquisition phase and the knowledge application phase.

3.2.2 Methods

Participants and Procedure. The present study was based on the same sample as described in study 1. However, due to technical problems, the necessary log-file data of the CPS assessment tool were only recorded for the second half of the schools surveyed, resulting in a sample of $n = 262$ students (61.07% female; age $M = 16.45$, $SD = 0.87$ years). These missing values were assumed to be missing (completely) at random because they were caused by technical problems in entire schools (five out of eleven schools).

Instruments. Intelligence and CPS performance were assessed as described in study 1. With respect to the strategic CPS behavior to identify eigendynamic effects early, we examined tasks six, eight, and nine, which were comprised of eigendynamic and direct effects. Overall CPS performance scores in the knowledge acquisition phase and the knowledge application phase were averaged across all nine tasks.

Analyses. To analyze students' strategic behavior related to eigendynamic effects, for each action in the causal diagram of the knowledge acquisition phase, it was coded whether a path indicating an eigendynamic effect was drawn, deleted, or if no action related to an eigendynamic effect was performed. The specific assignment of credit was done after a complete inspection of the actions. The subsequent analyses [Hypotheses (1) to (5)] were performed in Mplus (Muthén & Muthén, 1998-2012) and the clustered data structure was considered by using cluster robust-standard errors (CR-*SE*s; "type is complex" in Mplus). Hypotheses (1) and (3), which comprised dichotomous outcome variables, were evaluated using probit regression models. Hypotheses (2) and (4) comprised continuous outcome variables, so linear regression models were used. Regarding all regression analyses, we decided to conduct two-sided testing ($p < .05$). We inspected each of the three tasks separately, thereby considering the three tasks in the sense of conceptual replications. Concerning Hypothesis (5), a *g* factor of intelligence was modeled in accordance with the common higher-order structure model and we followed the procedure for mediation analysis with multicategorical independent variables (Hayes & Preacher, 2014; Mplus command "indirect").

3.2.3 Results

The pattern of descriptive results concerning the investigated strategic behavior was likewise for the three MicroDYN tasks six, eight, and nine: Approximately 23% to 33% of students submitted a causal diagram with the correct eigendynamic effect. Of these students, 55% to 60% inserted the correct eigendynamic effect in their first action and the remaining students inserted the eigendynamic effect in their actions two through ten (see Weise et al., 2022, for details).

In a further step, a coding scheme was specified based on the actions related to the eigendynamic effect: If the correct eigendynamic effect was inserted in the first action, we assigned two points (coded as 2); if it was inserted in a later action (i.e., action two through ten), one point was assigned (coded as 1) and if the correct eigendynamic effect was not inserted or if an incorrect eigendynamic effect for a different outcome variable was inserted, zero points were assigned (coded as 0). Students who had inserted the correct effect, then deleted it, and again inserted the effect, were excluded from further analyses. To compare students according to when they had identified the correct eigendynamic effect, a dummy coding scheme that represented the assumed contrasts was adopted: Contrast 1 [0 vs. (1&2)] compared students who did not identify the eigendynamic effect correctly to students who identified the eigendynamic effect correctly; contrast 2 (1 vs. 2) compared students who identified the eigendynamic effect in a later step to students who did so in a first step.

Regarding the probit regression model to inspect Hypothesis (1), students who identified the correct eigendynamic effect showed statistically significantly higher performance scores in terms of the direct effects of the respective task for all three tasks than students who did not identify the correct eigendynamic effect (first contrast: all p's $< .001$). Furthermore, students who identified the eigendynamic effect earlier exhibited statistically significantly higher scores than students who identified the eigendynamic effect later for tasks eight and nine (second contrast: p's $< .001$), but not for task six ($p = .101$). Results concerning Hypothesis (2) showed that the first, as well as the second contrasts, predicted statistically significantly the performance scores of the knowledge acquisition phase for the three tasks (all p's $< .05$). With respect to Hypothesis (3), the first contrast concerning task six yielded a significant effect ($p = .023$), whereas the second contrast was not significant ($p = .341$). For tasks eight and nine, the two contrasts were statistically significant predictors of knowledge application performance in the respective tasks (all p's $< .001$). Regarding Hypothesis (4), for the three tasks, the first and second contrasts were statistically significant predictors of overall performance scores in the knowledge application phase (all p's $< .01$).

Regarding the mediation analyses with the predictor intelligence, the mediator strategic behavior to identify eigendynamics (first and second contrast), and the criterion variables CPS performance in the knowledge acquisition and knowledge application phase, the six mediation models showed at least acceptable fit indices (CFI ≥ .94, RMSEA ≤ .05, for all six models). The results of the six mediation models concerning the knowledge acquisition phase and the knowledge application phase are depicted in Table 1.

Table 1

Results of the mediation models: Estimates (*est.*), standard errors (*SE*), and *p*-values (*p*) (*N* = 262 for all models; Weise et al., 2022).

	Knowledge acquisition phase (mean score)			Knowledge application phase (mean score)		
	est.	*SE*	*p*	*est.*	*SE*	*p*
Task 6						
Direct path	.09	.05	.048	.10	.06	.072
Sum of indirect paths	.08	.03	.005	.06	.03	.023
Indirect path 0 vs. (1 & 2)	.08	.03	.004	.06	.02	.021
Indirect path 1 vs. 2	.01	.004	.039	.01	.004	.128
Task 8						
Direct path	.10	.04	.014	.11	.05	.031
Sum of indirect paths	.08	.03	.020	.05	.03	.038
Indirect path 0 vs. (1 & 2)	.06	.03	.026	.04	.02	.031
Indirect path 1 vs. 2	.02	.01	.033	.01	.01	.151
Task 9						
Direct path	.08	.03	.016	.10	.04	.025
Sum of indirect paths	.09	.04	.014	.06	.03	.055
Indirect path 0 vs. (1 & 2)	.07	.03	.014	.05	.02	.057
Indirect path 1 vs. 2	.02	.01	.037	.01	.01	.164

Note. Due to technical reasons of the statistical software, we needed to use integers as weights for the dummy variables specifying the two Helmert contrasts [i.e., (-2/1/1), (0/-1/1)] in the mediation models.

In summary, the strategic behavior to identify eigendynamics significantly mediated the relationship between intelligence and CPS performance in the knowledge acquisition phase. Thereby, the comparison between students who inserted the correct eigendynamic effect either in their first step or in a later step in contrast to students who did not insert the correct effect (first contrast) and the comparison between students who inserted the correct eigendynamic effect in their first step in contrast to students who did so in a later step (second contrast) contributed significantly to the indirect path. For CPS performance in the knowledge application phase as an outcome variable, significant indirect paths were shown for tasks six and eight. Concerning the two contrasts, only the comparison between students who inserted the correct eigendynamic effect and students who did not (first contrast) acted as a significant mediator. Our expectations concerning the second contrast (i.e., inserting the correct eigendynamic effect in the first step vs. in a later step) were not supported.

3.2.4 Discussion

In this study, the strategic behavior to identify eigendynamic effects early was empirically investigated for the first time. The researched strategic behavior predicted the performance in the respective tasks and overall performance in a set of CPS tasks in both CPS phases. The latter finding suggests that the investigated strategic behavior can be interpreted as an indicator of a general strategic proficiency that is relevant for both CPS phases. In future studies, researchers should investigate how different strategic behaviors such as VOTAT, NOTAT, and the strategic behavior to identify eigendynamic effects early, are interrelated and complement each other. Importantly, NOTAT on the one hand, and the strategic behavior to identify eigendynamics early on the other hand capture different aspects of successful exploration in the knowledge acquisition phase. The mediation models showed that intelligence manifests itself in the strategic behavior to identify eigendynamics early going hand in hand with higher CPS performance in the knowledge acquisition phase. Concerning the knowledge application phase, the pattern of findings suggests that more intelligent students achieved higher CPS performance scores by adequately identifying eigendynamics (first comparison), but regardless of whether they applied the strategic behavior to identify eigendynamic effects early or not (second comparison). Possibly, the ability to adequately identify eigendynamics in the first CPS phase helped more intelligent students to transfer their system knowledge to the second CPS phase: a process at which surprisingly many students fail (Nicolay et al., 2021). An explanation for the non-significant mediation effect of the second contrast could be that the strategic behavior to identify eigendynamic effects early might be too distant to the second CPS phase to mediate the intelligence-CPS relationship in that CPS phase.

To conclude, the results of this study provide important insights into students' behavior when dealing with eigendynamic effects in CPS tasks and contribute to our understanding of the CPS-intelligence relationship. In perspective, the identification of specific and observable behaviors such as the strategic behavior to identify eigendynamics early can support learners and give guidance to teachers to improve students' performance when solving computer simulated and real-world complex problems.

3.3 Study 3: Ruby, Weise, Greiff, & Sparfeldt (submitted)

Ruby, J., Weise, J. J., Greiff, S., & Sparfeldt, J. R. (submitted). Double-checks or better no repeated steps? The Role of Effective and Efficient Exploration Behaviors for Successful Complex Problem Solving. *Computers & Education*.

In this study, we investigated the effectiveness and efficiency of students' exploration behavior in a sequence of five CPS tasks. Using latent class analysis, we identified several theoretically hypothesized classes of students and examined whether they differed in terms of intelligence and CPS performance.

3.3.1 Theoretical Background and Research Questions

Whenever test takers explore an unknown CPS scenario in the knowledge acquisition phase, they may strategically explore the problem space to gather the information needed to successfully solve the problem. For CPS tasks with only direct effects, applying VOTAT for each input variable provides the information needed to gain complete knowledge of the effects of the system. Furthermore, applying VOTAT exactly once for each input variable is not only effective, but also efficient to obtain complete information about the direct effects; any further exploration step only provides redundant information about the system. Correspondingly, a lower number of non-necessary exploration steps indicates efficient exploration behavior. However, empirical results regarding the number of interactions with the system are contradictory: One study showed that a lower number of interactions with the system, which the authors interpreted as efficient exploration behavior, was positively related to CPS performance (Greiff et al., 2016), while other studies found that a higher number of interactions (Naumann et al., 2014) and a so-called double-checking behavior (Eichmann, Greiff, et al., 2020) were positively related to CPS performance.

In addition to the effectiveness and efficiency of students' exploration behavior, we also examined the course of the aforementioned exploration characteristics over a sequence of five CPS tasks. Based on previous studies, it can be assumed that students will maintain or even increase the proportion of their strategic exploration behavior if the behavior has proven useful in previous CPS tasks and, furthermore, that an increased rate of strategic exploration behavior is related to intelligence and CPS performance (e.g., Lotz et al., 2017). Correspondingly, there are studies utilizing person-centered approaches [e.g., latent class analysis (LCA), latent transition analysis (LTA)] to investigate whether groups of students differ in their exploration behavior over the course of CPS tasks. For example, Molnár and Csapó (2018) found six classes of students that can be characterized as follows: Students of class 1 did not use and students of class 2 rarely used an effective exploration behavior, whereas students of class 3 consistently used effective exploration strategies. Students of class 4 partially used effective explorations strategies in the first and easy tasks, but not in the later and more complex ones. Students of class 5 mostly used effective explorations strategies in the first and easy tasks, but not in the later and more complex tasks, and students of class 6 did not use an effective exploration behavior in the beginning, but in the end (see also Greiff et al., 2018; Mustafić et al., 2019). However, the before-mentioned person-centered approaches only considered the effectiveness (i.e., VOTAT) and not the efficiency of the exploration behavior over the course of tasks. Furthermore, the question of how effective and efficient exploration behavior is related to intelligence and CPS performance remains unresolved. Therefore, the overall objectives of this study were as follows:

Research Question 1. First, we analyzed the prevalence of students' effective (i.e., VOTAT for each input variable) and efficient (i.e., the number of non-necessary exploration steps while applying VOTAT for each input variable) VOTAT behaviors across tasks, expecting an increase in effective exploration behavior and a decrease in inefficient exploration behavior over the sequence.

Research Question 2. Next, we examined individual exploration patterns, expecting to identify distinct classes of participants based on the effectiveness and efficiency of their exploration behavior across tasks (see Greiff et al., 2018; Molnár & Csapó, 2018).

Research Question 3. We then examined whether class membership was related to participants' intelligence, hypothesizing higher intelligence scores for participants in classes that showed an effective and more efficient exploration behavior than for students in classes that showed an ineffective or an effective, but less efficient exploration behavior.

Research Question 4. Finally, we examined whether class membership was related to participants' CPS performance scores, expecting the same pattern of results as in Research Question (3).

3.3.2 Methods

Participants and Procedure. Study 3 was based on the same sample as study 1; $N = 495$.

Instruments. Intelligence and CPS performance were assessed as described in study 1. Regarding the effectiveness and efficiency of the participants' exploration behavior, we focused on the first five tasks of the CPS assessment tool MicroDYN, which consisted of direct effects only. The log-file data from the knowledge acquisition phase were used to obtain a combined score for effectiveness and efficiency as follows: If VOTAT was not shown at least once for each input variable, the corresponding strategic behavior was scored as "ineffective". If VOTAT was shown once for each input variable (i.e., effective behavior), we scored the efficiency based on the number of non-necessary exploration steps.

Analyses. Analyses were conducted with the statistical packages IBM SPSS Statistics 26 and Mplus 7.11 (Muthén & Muthén, 1998–2012). Concerning Research Question (1), we adopted the following coding scheme: If students did not apply VOTAT for every input variable of a task, the behavior was considered as non-effective and coded as 8. If students showed VOTAT for every single input variable, their exploration behavior was coded according to the efficiency, which is the number of non-necessary steps: For zero steps more than the number of necessary steps, the exploration behavior was coded as 0; for one step more than the number of necessary steps, the exploration behavior was coded as 1; …; for six steps more than the number of necessary steps, the exploration behavior was coded as 6; and for seven or more steps more than the number of necessary steps, the exploration behavior was coded as 7. If students did not explore the respective CPS task, exploration behavior was coded as missing. By applying this coding scheme, we considered differences in the number of input variables between the five CPS tasks.

To address Research Question (2), we conducted a LCA in Mplus and estimated models with one to six latent classes using robust maximum likelihood estimation (MLR). The observed variables were the categorically ordered indicators of exploration effectiveness and efficiency. For this analysis and all following analyses, only students who worked on all five CPS tasks were included, resulting in $n = 469$ students. The optimal number of classes was determined on the basis of the following analysis strategy

(Collins & Lanza, 2010; Geiser, 2012): To determine absolute fit, we inspected the Pearson χ^2 statistic, the Likelihood-Ratio χ^2 statistic, and χ^2/df. To determine relative fit, we inspected the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample-size adjusted BIC (adjBIC). Simulation studies showed that BIC and adjusted BIC are more informative indices than AIC when determining the number of classes (Nylund et al., 2007). Furthermore, we inspected the interpretability in terms of the meaningfulness of the classes and the average latent class attribution probabilities of the LCA models with different numbers of latent classes. In accordance with prior studies (e.g., Greiff et al., 2018), we additionally inspected the parametric bootstrapped likelihood ratio test, which compares a model with k latent classes and a model with k – 1 classes.

Regarding Research Question (3) and (4), we used Mplus to model the *g* factor of intelligence in accordance with the common higher-order structure and CPS performance scores of the knowledge acquisition phase and the knowledge application phase as latent factors. Students' clustering in classrooms was considered by using the "type is complex" option that controls for potential effects of clustering on standard errors and χ^2-statistics. To compare the students of the latent classes with regard to their intelligence and CPS performance, we inspected linear regression models and used pairwise Scheffé comparisons. Class membership served as predictor, the latent *g* factor and CPS performance in both phases as criterion. The significance level was set to $p < .05$; in order to control for the accumulation of alpha errors, we adjusted according to the Bonferroni-Holm procedure (Holm, 1979).

3.3.3 Results

Regarding Research Question (1), the proportion of students who applied an effective VOTAT behavior without non-necessary exploration steps for all input variables (coded as 0) increased from 1.0% in task one to 22.2% in task five. Correspondingly, the proportion of students who showed VOTAT for each input variable, but used seven or more non-necessary exploration steps (coded as 7) decreased from 31.5% in task one to 8.7% in task five (see Appendix A, Ruby et al., submitted, for further details).

Regarding Research Question (2), relative fit indices indicated that either the three or the four-class solution were most appropriate. The three-class solution was supported by the BIC, whereas the four-class solution was supported by the AIC. Notably, the adjusted BIC was comparably low for both, the three- and the four-class solution. The bootstrapped likelihood test showed a significantly better fit for the four-class solution compared with the three-class solution ($\Delta\chi^2 = 121.60$; $df = 41$; $p < .001$). Taking the relative fit indices, the results of the bootstrapped likelihood test, and the interpretation of the different solutions into account, the four-class solution appeared to be the most appropriate solution. Therefore, we selected the four-class solution as the basis for all further analyses.

The profiles of the four-class solution are visualized in Figure 5 displaying the conditional probabilities of students' exploration behavior in the five CPS tasks. Students of the first class showed for all CPS tasks a high probability not to apply VOTAT consistently and, therefore, they were labeled as *ineffective explorers* (1). Concerning the second class, the main characteristic of students' behavior was a consistent but inefficient use of VOTAT, so they were labeled as *inefficient explorers* (2). The third class and the fourth class both consisted of students who showed a high probability of the most efficient VOTAT-related exploration behavior at the end of the task set, but differed in regard to their exploration behavior when starting to work on the CPS tasks. Students of the third class showed a rather high probability not to use VOTAT consistently for the first and second task and, therefore, they were labeled as *emerging explorers* (3). For the students of the fourth class, the main characteristic was that the less efficient behavior decreased, while the most efficient behavior increased over the set of tasks and showed a rather high value for the last task, resulting in the label of *proficient explorers* (4).

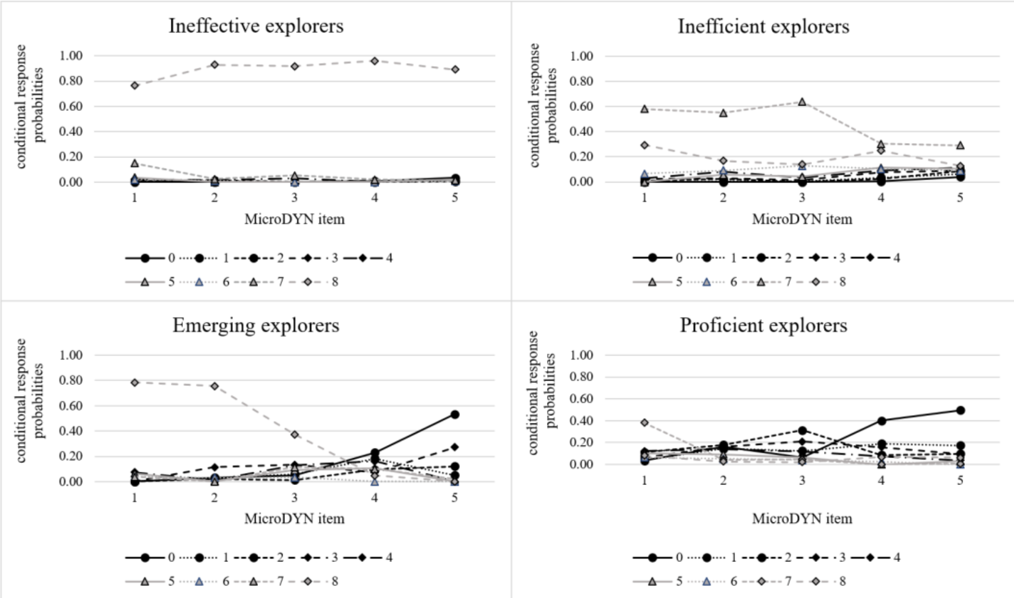

Figure 5: Profiles of the four classes; displayed are the conditional probabilities for the nine ordered categories (i.e., from “0” to “6”, “7”, and “8”) across the five CPS tasks (i.e., MicroDYN items).

Results regarding Research Question (3) and (4) are depicted in Table 2. In accordance with our expectations, students of classes with a mostly effective exploration behavior (i.e., classes 2, 3, and 4) achieved statistically significantly higher intelligence scores than students of the class with an ineffective exploration behavior (i.e., class 1). Contrary to our expectations, students of the latent class 2, who showed an effective but inefficient exploration behavior, reached statistically significantly higher intelligence scores than students who started with an ineffective exploration behavior and changed to an effective and efficient behavior (i.e., class 3) and students who started with an effective but inefficient exploration behavior and changed to an effective and efficient behavior (i.e., class 4).

Concerning CPS performance (Research Question 4), and in line with our expectations, students of classes with a mostly effective exploration behavior (i.e., classes 2, 3, and 4) performed statistically significantly better than students of the class with an ineffective exploration behavior (i.e., class 1). Only partially in line with our expectations, students of class 2 who showed an effective but inefficient exploration behavior reached statistically significantly higher CPS performance scores than students of class 3 (emerging explorers). Furthermore, students of class 2 did not differ significantly regarding their CPS performance from students of class 4 (proficient explorers).

Consistent with our hypothesis, the aforementioned students of class 4 achieved statistically significantly higher CPS scores in both CPS phases than students of class 3.

Table 2

Results of the linear regression models (*n* = 469).

Comparisons of latent classes	*est.*	*SE*	*p*
Knowledge acquisition phase (mean score task 1–5)			
1 vs. 2	.78	.03	< .001*
1 vs. 3	.67	.04	< .001*
1 vs. 4	.73	.05	< .001*
2 vs. 3	-.29	.06	< .001*
2 vs. 4	.01	.08	.91
3 vs. 4	.29	.07	< .001*
Knowledge application phase (mean score task 1–5)			
1 vs. 2	.50	.05	< .001*
1 vs. 3	.41	.05	< .001*
1 vs. 4	.49	.05	< .001*
2 vs. 3	-.24	.05	< .001*
2 vs. 4	-.02	.05	.67
3 vs. 4	.20	.06	< .001*
Intelligence (latent *g* factor)			
1 vs. 2	.44	.05	< .001*
1 vs. 3	.28	.06	< .001*
1 vs. 4	.27	.06	< .001*
2 vs. 3	-.27	.05	< .001*
2 vs. 4	-.16	.06	.004*
3 vs. 4	.08	.07	.254

* significant on a 5%-level after adjusted according to the Bonferroni-Holm procedure.

3.3.4 Discussion

In the present study, we investigated the effective and efficient use of VOTAT in a series of five CPS tasks. The inspection of log-file data showed that exploration behavior became more effective and efficient across the task sequence. The increase in effectiveness is consistent with the findings of previous studies (e.g., Vollmeyer et al., 1996). However, the present study also showed an increase in the efficiency of exploration behavior, going beyond previous studies. With regard to Research Question (2), we found four distinct classes of students meaningfully differentiating (1) ineffective explorers, (2) inefficient explorers, (3) emerging explorers, and (4) proficient explorers.

Some of these findings are similar to those of previous studies using person-centered approaches (e.g., Greiff et al., 2018; Molnár & Csapó, 2018; Mustafić et al., 2019). However, because we examined both effectiveness and efficiency of exploration behavior in the present study, we found two different classes that consistently exhibited effective behavior but differed in their efficiency (i.e., classes 3 and 4).

With regard to intelligence, the findings were partially consistent with our expectations, highlighting that VOTAT application (i.e., effectiveness) is related to intelligence (e.g., Kröner et al., 2005; Lotz et al., 2017). In contrast to our expectations, inefficient VOTAT explorers (2) reached higher intelligence test scores than emerging explorers (3) and proficient explorers (4), not differing among each other. Students in class 2 may have repeatedly checked their hypotheses about the structure of the complex system using VOTAT, either to be more confident about their conclusion or because they expected a sudden change in the system (e.g., new effects appearing after a certain time/number of exploration steps). Moreover, our results showed that this exploration behavior was associated with higher intelligence test scores. This unexpected pattern of results requires further clarification in future studies. Possibly, the reasons for inefficient explorers (2) to reevaluate their hypotheses could be investigated in thinking aloud protocols (Klopp et al., 2020).

With regard to CPS performance scores, the pattern of results was also partially in line with our expectations, showing that classes of students who explored more effectively exhibited higher CPS performance. In contrast to our expectations, class 2 (inefficient explorers) and class 4 (proficient explorers) did not differ from each other. The finding that inefficient exploration behavior, as demonstrated by class 2, can be associated with (comparatively) high CPS performance scores is somewhat similar to the finding of Eichmann, Greiff et al. (2020), who found that double-checking was associated with higher CPS performance. However, the aforementioned study focused on goal-directed exploration behaviors (i.e., any action necessary to successfully complete the task) and not specifically on the number of non-necessary VOTAT actions as considered in our study.

Notable limitations of the present study are, first, that effective exploration was defined only as the use of VOTAT for each input variable. However, students may use other strategic behaviors to effectively explore the system, such as to *hold-one-thing-at-a-time* (*HOTAT*; i.e., a participant holds one input variable while manipulating the other input variables; Tschirgi, 1980; see also Nicolay et al., 2023; Rollett, 2008). Although not every possible way to successfully explore the system was included in this study, the included strategic exploration behavior represents the most important and apparent strategy. Second, the instructions for the CPS assessment tool did not explicitly emphasize efficiency. In addition, the time limit in the knowledge acquisition phase of 180 seconds for each task was rather liberal. Therefore, the effects of efficiency found in this study can be considered as lower bounds, and we expect pronounced effects if the instruction emphasizes efficiency and/or the time to explore the system is reduced.

Taken together, the present study found four meaningful classes of students who differed in the effectiveness and efficiency of their exploration behaviors. With respect to CPS performance and intelligence, the results highlight the importance of VOTAT (i.e., effectiveness) for successful exploration of complex systems. However, the efficiency of VOTAT use was less consistently related to CPS performance and intelligence, and future research needs to clarify the theoretical status of efficient exploration behavior and how it relates to successful problem solving.

3.4 Study 4: Weise, Becker, & Sparfeldt (2024)

Weise, J. J., Becker, N., & Sparfeldt, J. R. (2024). Intelligenzmessung in drei Minuten – Evaluation des *mini-q* und Konstruktion einer Parallelversion [Assessment of intelligence in three minutes – An evaluation of the mini-q and the construction of a parallel form]. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. https://doi.org/10.1024/1010-0652/a000384

Study 4 investigated the psychometric properties of a three-minute intelligence screening tool, the *mini-q* (Baudson & Preckel, 2016), and introduced a parallel version to expand its application options and determine its parallel test reliability. This work contributes to future research on the relationship between intelligence and CPS by providing a time-efficient intelligence screening instrument.

3.4.1 Theoretical Background and Research Questions

There are numerous established instruments for measuring intelligence that exhibit satisfactory to good psychometric properties (Preckel & Brüll, 2008; Sparfeldt et al., 2022). For large-scale research projects, however, comprehensive intelligence diagnostics can be too time-consuming, so time-efficient and well-evaluated intelligence screenings with good psychometric properties are highly desirable. The *mini-q* (Baudson & Preckel, 2016) provides an indicator of general intelligence within a test-taking time of only three minutes and is made available by the test authors for non-commercial use in scientific research. In addition, promising results regarding the psychometric properties of the *mini-q* have been reported in initial studies (Baudson & Preckel, 2016).

The origins of the *mini-q* (Baudson & Preckel, 2016) go back to Baddeley's (1968) reasoning test, which is based on grammatical transformations. Participants are asked to evaluate verbal statements (e.g., "A is not followed by B") in relation to a subsequent letter sequence (e.g., "BA"). The statements regarding the letter sequence differ in regard to six characteristics, resulting in $2^6 = 64$ distinct items to be completed by the test participants within a time limit of three minutes. In developing the *mini-q*, the basic concept of Baddeley's (1968) reasoning test was largely adopted. However, because of the need to conjugate verbs in the passive voice, a direct translation of the verbal statements was not possible. Instead, Baudson and Preckel (2016) used the German semantic contrast pair “*vorziehen*” (English: “*prefer*”) and “*ablehnen*” (English: “*reject*”) to refer to a sequence of symbols consisting of a square, a triangle, and a circle (see Figure 6). The statements always refer to the position of the triangle, which is consistently located in the center and varies in distance to the neighboring symbols. The variations in the verbal statements in Baddeley’s test have been fully adopted, so that the *mini-q* also consists of 64 items.

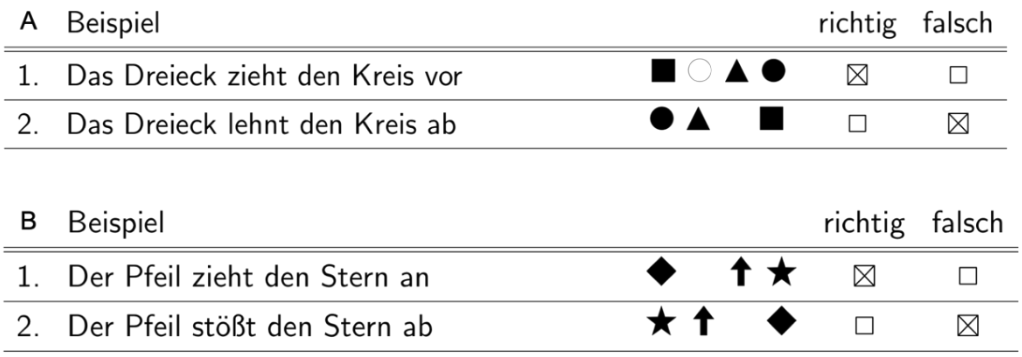

Figure 6: Two correctly solved items each from the instruction of the mini-q (upper example A; Baudson & Preckel, 2016, p. 183) and the parallel version mini-q B (lower example B; Weise et al., 2024).

Baudson and Preckel (2016) used the *mini-q* in seven sub-studies (*N* = 478, mostly university students; mean raw score: *M* = 35.05, *SD* = 9.70, range = 11 – 63). An odd-even reliability of *r* = .98 indicated high reliability. Evidence of convergent validity included a significant correlation with the GPA of the high school graduation certificate (German Abitur; *r* = -.28) and at least medium-sized correlations with other intelligence indicators assessed in subsamples: figural reasoning (CFT 20-R; Weiß, 2008; *r* = .51), reasoning/general intelligence (IST-Screening; Liepmann et al., 2012; *r* = .67), crystallized intelligence (Mehrfachwahl-Wortschatz-Test; Lehrl, 1999; *r* = .32) and processing speed (Ulmer-Speed-Batterie, USB; Schmitz & Wilhelm, 2016, 2019; *r* = .73). In stepwise regressions with the reverse order of the predictors IST-Screening (reasoning) and USB (processing speed) as well as the *mini-q* as criterion, similar proportions of incremental variance were explained for both intelligence facets (15% increment; variance explained by the first predictor in each case approx. 50%). Therefore, Baudson and Preckel (2016, p. 194) argued that the newly developed screening instrument captured “speeded reasoning”.

The *mini-q* has now been used in more than 20 studies (e.g., Leichner et al., 2022; Meuer & Imhoff, 2021; Nalis et al., 2018; Pollet & Schnell, 2017; Schlegel & Mortillaro, 2019). However, these publications did not report reliability measures beyond odd-even correlations and Cronbach's alphas, nor did they report further construct or criterion validity results obtained from independent samples. Therefore, further research is needed. In addition, a parallel version would be a valuable addition, for example for group testing and test repetitions.

Research Question. In the present study, *the mini-q* was to be tested on an independent sample and a newly created parallel version was to be evaluated. For both versions, (**1**) central psychometric properties were to be examined, (**2**) evidence regarding parallel test reliability, convergent validity with another intelligence test, and criterion validity with the GPA of the high school graduation certificate (Abitur) were to be investigated, and (**3**) mean differences between the two test versions and practice effects were to be analyzed.

3.4.2 Methods

Participants and Procedure. The *mini-q* and *mini-q B* (see below) were tested as part of an educational science lecture. A sample of $N = 163$ student teachers (different types of schools and subjects) participated in a group test. For this purpose, students were randomly assigned two paper versions of the test booklet that included both *mini-q* versions in counterbalanced order. To avoid administering the two *mini-q* versions directly one after the other, 22 personality items of no relevance to our research questions of this study were interspersed. In the first lecture session, approximately ten weeks prior to the above-mentioned data collection, intelligence (see below), age, gender, and the GPA of the high school graduation certificate (Abitur) were collected ($n = 136$; age: $M = 23.12$; $SD = 5.09$; 77.2% female, 21.3% male, 1.5% without gender information). The GPA of the high school graduation certificate was available for $n = 135$ students ($M = 2.01$; $SD = 0.62$; range 1.0 – 3.7).

Instruments. In addition to the *mini-q* (Baudson & Preckel, 2016), we created a parallel version (*mini-q B;* Weise et al., 2024) by replacing the graphic elements (circle, square, triangle) with new symbols (arrow, diamond, star) and the verbs „*vorziehen*" (English: prefer) and „*ablehnen*" (English: reject) with „*anziehen*" (English: attract) and „*abstoßen*" (English: repel; see Fig. 6). Both, *mini-q* and *mini-q B*, each consisted of an instruction, six sample items, and 64 items (administration time: three minutes; for more details, see Weise et al., 2024). As a further measure of intelligence, we used the IST-Screening (version A; Liepmann et al., 2012) with the total score as an indicator of general intelligence.

Analyses. To assess the *mini-q* and *mini-q B*, we assigned 1 point for each correctly solved task and 0 points for each incorrectly solved or unfinished (omitted) task, as in Baudson and Preckel (2016), and added them to the total score. In addition to the descriptive parameters, we also examined the average difficulty of the items. For these particular analyses, only the tasks that students had worked on were considered, and no omitted tasks were included in determining the difficulty of the tasks. Reliability coefficients were provided by Cronbach's alpha, odd-even reliability, and parallel test reliability of both *mini-q* versions per test booklet. As validity indicators, we calculated correlations and 90% confidence intervals (CI) with the sum score of the IST-Screening and the GPA of the high school graduation certificate. In addition, we determined reliability and validity across the test booklets (total sample) for both test versions. The sum scores of the first administered *mini-q* version were compared using an independent samples t-test (supplemented by the effect size *d*). Practice effects, that is, within-individual mean differences between *mini-q* and *mini-q B*, were analyzed using a t-test for dependent samples per test booklet.

3.4.3 Results

Regarding the *mini-q*, the mean sum score for the first/second *mini-q* version administered (test booklet 1: *mini-q*, *mini-q B*; test booklet 2: *mini-q B*, *mini-q*; $n = 81/82$) was $M = 32.79/43.54$ ($SD = 7.88/11.39$; Table 3). Skewness and kurtosis did not indicate a significant deviation of the *mini-q* sum score distribution from a normal distribution, and there was no evidence of floor or ceiling effects. The mean number of items that students worked on was $M = 33.75/44.63$ ($SD = 7.70/10.93$). The probabilities of correctly solving an item ranged between .92 – 1.00/.88 – 1.00 ($M = .98/.97$; $SD = .02/.03$). An indicator of reliability obtained using the odd-even method showed a correlation of $r = .96/.99$ (total sample: $r = .98$), Cronbach's alpha was $\alpha = .94/.96$ (.96). The sum score of the first administered *mini-q* version correlated statistically significantly with the sum score of the IST-Screening ($r = .25$, $p = .042$, CI: .05, .43, $n = 69$; total sample: $r = .37$, $p < .001$, CI: .24, .49, $n = 136$), but not with the GPA of the high school graduation certificate (*r* = -.07, *p* = .581, CI: -.26, .13, *n* = 69; total sample: *r* = -.24, *p* = .006, CI: -.37, -.10, *n* = 135).

Table 3

Descriptive results of the *mini-q* and *mini-q B* for test booklet 1 (order: *mini-q*, *mini-q B*) and test booklet 2 (order: *mini-q B*, *mini-q*; Weise et al., 2024).

	Test booklet 1 (*n* = 81)				**Test booklet 2** (*n* = 82)			
	M (*SD*)	range	skewness (*SE*)	kurtosis (*SE*)	*M* (*SD*)	range	skewness (*SE*)	kurtosis (*SE*)
mini-q	32.79 (7.88)	12-55	.32 (.27)	.55 (.53)	43.54 (11.39)	19-64	.13 (.27)	-.66 (.53)
mini-q B	44.00 (9.27)	19-63	-.05 (.27)	.22 (.53)	38.24 (9.50)	19-58	.11 (.27)	-.69 (.53)

Regarding the *mini-q B*, a mean sum score of *M* = 38.24/44.00 (*SD* = 9.50/9.27; Table 3) was obtained for the first/second administered version (test booklet 2/1; *n* = 82/81). Skewness and kurtosis also indicated no significant deviation from a normal distribution, and there was no evidence of floor or ceiling effects. On average, students worked on *M* = 38.94/45.42 (*SD* = 9.18/9.15) items (solution probabilities: .94 – 1.00/.83 – 1.00; *M* = .99/.96; *SD* = .02/.04). Odd-even reliability coefficients reached *r* = .98/.96 (total sample: *r* = .97), Cronbach's alpha was α = .95/.95(.95). The *mini-q B* sum score of the first administered version correlated statistically significantly with the IST-Screening (*r* = .47, *p* < .001, *CI*: .29, .61, *n* = 67; total sample: *r* = .32, *p* < .001, CI: .19, .44, *n* = 136) and the GPA of the high school graduation certificate (*r* = -.25, *p* = .046, CI: -.43, -.05, *n* = 66; total sample: *r* = -.14, *p* = .118, CI: -.27, .01, *n* = 135).

The correlations of the sum scores of *mini-q* and *mini-q B* in test booklets 1 (*r* = .76, CI: .67, .83, *p* < .001) and 2 (*r* = .80, CI: .73, .86, *p* < .001) provided evidence of parallel test reliability. The sum scores of the test version administered second were each statistically significantly higher than those of the version administered first (*t*(80/81) = 16.50/7.07, both *p* < .001, *d* = 1.84/0.78). A comparison of the two test versions which were administered first (test booklet 1: *mini-q*, test booklet 2: *mini-q B*) showed a statistically significant higher mean sum score for the *mini-q B* than for the *mini-q* (test for homogeneity of variance: $F(1,161) = 5.60$, $p = .019$; independent samples t-test: $t(156.45) = -4.00$, $p < .001$, $d = 0.63$).

3.4.4 Discussion

In the present study, we investigated the psychometric properties of the very time-efficient intelligence screening instrument *mini-q* and introduced a parallel version in order to expand its application options and determine its parallel test reliability. The distributions of the *mini-q* and *mini-q B* sum scores showed substantial variances and no substantial deviations from a normal distribution, indicating reasonable test difficulty and discrimination between individuals of different cognitive abilities. Regarding reliability, both test versions showed high values of odd-even reliability ($.96 \leq r_{tt} \leq .99$) and good values of parallel test reliability ($r = .76$, $r = .80$) in the present sample of university students. With regard to convergent validity, the results of the two test versions were only partially in line with our expectations: The correlation between the *mini-q* and the IST-Screening was numerically lower ($r = .25$) than the correlation expected between intelligence tests and reported in previous studies (Baudson & Preckel, 2016: $r = .67$; Schubert et al., 2024: $r = .57$). With regard to the second indicator of convergent validity, the *mini-q* was not significantly related to the GPA ($r = -.07$). For the *mini-q B*, our results showed a more convincing correlation of $r = .47$ with the IST-Screening and of $r = -.25$ with the GPA, similar to what Baudson and Preckel (2016) reported for the correlation with the GPA for the original version ($r = -.28$). The partially unconvincing results regarding the validity were unexpected and cannot be explained by fundamental weaknesses of the sample, which included student teachers for different types of schools and different school subjects and was therefore rather heterogeneous. Considering that the *mini-q* is intended for use in non-commercial research, the sample of university students should be well suited to evaluate the two versions of the screening instrument. In future research, it would be desirable to evaluate the *mini-q* and *mini-q B* in a larger sample and to clarify our findings and the differences with the results reported by the test authors.

Of note, the significantly higher mean score of the *mini-q B* ($d = 0.63$) compared to the *mini-q* should be considered when comparing the two versions. Because the test booklets were randomly assigned and the comparison above considered the versions that were completed first, the reported difference most likely reflects the effect of the modifications between the test versions (e.g., verbs and symbols in the items). With respect to the question of stability and the effects of test repetition, we were able to show an increase in test scores from the first to the second administration (d = 1.84/0.78). Our results exceeded the meta-analytic test-retest effects of intelligence tests by Scharfen et al. (2018; Bias Corrected Standardized Mean Change = 0.37/0.23 for identical/alternate test form), however these results were based on longer test-retest intervals.

A limitation of the present study is that it was not possible to collect further evidence on the extent to which the *mini-q* captures reasoning and processing speed. However, a recently published study by Schubert et al. (2024) reported that both processing capacity (β = .23) and processing speed (β = .34) contributed to the *mini-q* performance in comparable magnitudes. Regarding the skepticism that the *mini-q* may be a test of working memory capacity rather than intelligence, Schubert et al. (2024) found in their mediation analysis that the relationship between intelligence and *mini-q* performance was largely driven by working memory capacity. However, in a multiple regression analysis when controlling for the covariance between a general factor of cognitive abilities and working memory capacity, only g (i.e., intelligence), but not working memory capacity, predicted *mini-q* performance. Overall, the results of the aforementioned study supported Baudson and Preckel's (2016) assumption that the *mini-q* is a test of speeded reasoning.

Taken together, the results of the *mini-q* and the *mini-q B* reported in the present study indicate a good reliability of the two test versions and provide an estimate of the test-retest reliability. In terms of convergent validity, some of the correlations were lower than expected and thus only partially in line with our expectations. Although the *mini-q* cannot replace a comprehensive intelligence diagnostic (see also Baudson & Preckel, 2016), our results indicate that both test versions are suitable for obtaining a rough intelligence score in research projects. Particularly noteworthy about the *mini-q* and its parallel version are the promising psychometric properties, the free availability, and the very time-efficient administration, making both versions valuable intelligence screening tools for research projects.

4 General Discussion

The primary goals of this dissertation project were to shed light on the relationship between intelligence and CPS by investigating the factors that influence the strength of the association between the two constructs, and to examine observable patterns of behavior in which intelligence manifests itself when solving complex problems and through which it translates into CPS performance. The following discussion will therefore focus primarily on study 1 through study 3, which revolve around this topic. The implications and limitations of study 4, which examined and extended the intelligence screening *mini-q,* are discussed in a chapter in the implications section (see chapter 4.1.4. The *mini-q* as a time-efficient intelligence screening).

In study 1, we investigated the Elshout-Raaheim hypothesis (Leutner, 2002; see Elshout, 1987; Raaheim, 1988) that predicts a curvilinear moderation of prior knowledge on the intelligence-CPS relationship. Specifically, the hypothesis assumes higher correlation coefficients between intelligence and CPS at medium levels of prior knowledge and lower correlations at both higher and lower levels of prior knowledge. Results revealed that in a set of five CPS tasks, prior strategic knowledge, indicated by the relative frequency of VOTAT use, increased in the knowledge acquisition phase. With increasing prior knowledge, correlations followed the predicted inverted U-shaped pattern in the knowledge acquisition phase and the knowledge application phase of CPS, thus supporting the Elshout-Raaheim hypothesis.

In study 2, we investigated students' strategic behavior when dealing with eigendynamic effects, which constitute an important feature of CPS tasks. Specifically, we investigated the strategic behavior to identify eigendynamic effects early while exploring the system, which is reasonably an effective strategic behavior (Beckmann, 1994). Results showed that the strategic behavior to identify eigendynamic effects early in the knowledge acquisition phase predicted CPS performance in both the knowledge acquisition and the knowledge application phase of the utilized CPS assessment tool. In mediation models, the investigated strategic behavior mediated the intelligence-CPS relationship in the knowledge acquisition phase, but not in the knowledge application phase.

In study 3, we investigated (latent) classes of students based on their effective and efficient use of the strategic exploration behavior VOTAT, and whether these classes differed in terms of CPS performance and intelligence. The results of the LCA revealed four distinct classes, that can be described as ineffective explorers (1), inefficient explorers (2), emerging explorers (3), and proficient explorers (4). Further analyses showed that students of the classes which explored (mostly) effective achieved higher intelligence test scores and CPS performance scores than the class of students not exploring effectively (i.e., class 4). This finding highlights the role of VOTAT in successful problem solving and its relationship to intelligence. With regard to efficiency, students, who showed an effective but inefficient exploration behavior (i.e., class 2) achieved higher intelligence test scores than students who started with an ineffective or inefficient exploration behavior and changed to an effective and efficient behavior (i.e., classes 3 and 4). A similar pattern was found for CPS, showing that students of class 2 achieved higher performance scores than students of class 3 and did not differ from students of class 4. Therefore, the role of efficiency in successful problem solving and its relationship to intelligence should be further clarified.

In study 4, we investigated the psychometric properties of the *mini-q* (Baudson & Preckel, 2016), a three-minute intelligence screening tool, and introduced a parallel version to expand its application options and determine its parallel test reliability. In perspective, time-efficient and well-evaluated intelligence screenings are important instruments in research projects when time is limited. The distributions of the sum scores of the two test versions indicated discrimination between individuals of different ability and were substantially correlated in terms of parallel test reliability. In terms of validity, we found correlations with another intelligence test and GPA that were partially consistent with our assumptions. In summary, the results are promising and indicate the suitability of the *mini-q* and its parallel version for obtaining a rough indicator of intelligence.

4.1 Implications

4.1.1 Intelligence

The findings of the studies 1 through 3 highlight the importance of strategic behaviors (i.e., VOTAT, the strategic behavior to identify eigendynamic effects early) as a keystone to understand the relationship between intelligence and problem solving. Mediation models showed that for the knowledge acquisition phase, intelligence manifested itself in the strategic behavior to investigate eigendynamic effects early (study 1). In a recent study by Lotz et al. (2022), results showed a manifestation of intelligence in VOTAT and NOTAT going hand in hand with higher performance in both the knowledge acquisition phase and the knowledge application phase in a series of CPS tasks (for mediation of only VOTAT, see also Grežo & Sarmány-Schuller, 2022; Wüstenberg, Stadler, et al., 2014). As the strategic behavior to identify eigendynamic effects early complements VOTAT and NOTAT in terms of the sequence of their effective application, a mediation model examining the three strategic behaviors in concert would most likely explain further variance in the intelligence-CPS relationship. The results of study 2 and Lotz et al. (2022) suggest that, in general, process measures that represent effective strategic behaviors and are related to intelligence and CPS performance are promising candidates for mediating the intelligence-CPS relationship. Moreover, in study 1, VOTAT served as an indicator of prior strategic knowledge that moderated the intelligence-CPS relationship. Possibly, in a set of CPS tasks consisting of direct and eigendynamic effects, a composite score of effective process measures, such as VOTAT, NOTAT, and the strategic behavior to identify eigendynamic effects early, could be a good indicator of prior strategic knowledge and, accordingly, moderate the relationship between intelligence and CPS performance.

Taken together, study 1 and the findings of Lotz et al. (2022) suggest that VOTAT, when considered as an indicator of prior knowledge, is a moderator of the intelligence-CPS relationship and, when considered as an effective strategic behavior, is a mediator of the intelligence-CPS relationship, highlighting its importance as a key strategic behavior for problem solving (Lotz et al., 2017; Nicolay et al., 2023; Wüstenberg et al., 2012). Combining the results of study 1, study 2, and Lotz et al. (2022), it can be argued that there may be further effective process measures, such as NOTAT and the strategic behavior to identify eigendynamic effects early, that could also be moderators and mediators of the intelligence-CPS relationship, similar to VOTAT. Strategic behaviors that simultaneously mediate and moderate the intelligence-CPS relationship would represent the essential observable behaviors that constitute the intelligence-CPS relationship. That is, these strategic behaviors would be the key mechanisms that translate intelligence into successful problem solving. Furthermore, in terms of moderation, the degree to which these strategic behaviors are known and used by problem solvers (i.e., considered as indicators of prior strategic knowledge) would determine the strength of the relationship between intelligence and CPS. Taken together, these considerations imply that if intelligence cannot be translated into successful problem solving through the aforementioned strategic behaviors, the relationship between intelligence and CPS would weaken, that is, intelligence would play a less important role in solving complex problems. In such a case, problem solvers might use other, non-strategic behaviors, such as trial-and-error, so that intelligence would be much less associated with successful problem solving (see also Jensen, 1980, 1989, on the relationship between intelligence and learning). In perspective, a deeper understanding of the intelligence-CPS relationship, as well as insights into how intelligence translates into problem solving, opens avenues for future research to promote problem solving in adaptive learning environments (Bunderson et al., 1989).

However, these assumptions regarding a moderating and mediating effect of multiple relevant strategic behaviors should be tested in future research and there are most likely several limitations: First, the set of strategic behaviors for which the idea was exemplified is most likely not exhaustive. In addition to VOTAT, NOTAT, and the strategic behavior to identify eigendynamic effects early, there might be other relevant strategic behaviors, some of which are already known (e.g., HOTAT; Nicolay et al., 2023; see also Rollett, 2008), some of which are still unknown. Second, the set of tasks that moderates and mediates the relationship between intelligence and CPS could vary according to the types of effects that the CPS tasks are composed of. And third, for the two CPS phases knowledge acquisition and knowledge application, there is probably a different set of strategic behaviors that serve as moderators and mediators (e.g., the strategic behavior to identify eigendynamic effects early did not mediate the relationship between intelligence and performance in the knowledge acquisition phase in study 2, so that the ideas outlined above may not apply to the second CPS phase for this strategic behavior). Nevertheless, the idea of multiple strategic behaviors that simultaneously mediate and moderate the relationship between intelligence and CPS may offer great insight and should be pursued in future research.

4.1.2 Strategic behaviors

With regard to the analyzed strategic behaviors, each of the three studies 1 through 3 demonstrated the usefulness and importance of the respective CPS behavior. However, in study 2 as well as in the study by Lotz et al., (2022), there was still a significant proportion of the intelligence-CPS relationship that could not be explained by the investigated strategic behaviors (see significant direct paths of mediation models). Following the idea outlined in the last section, future studies should investigate further strategic behaviors that mediate the intelligence-CPS relationship and examine these behaviors in concert. Obtaining a more complete set of CPS behaviors has practical relevance beyond the theoretical implications of an improved understanding of the intelligence-CPS relationship. Such a set of strategic behaviors improves the possibilities to offer advice to learners and teachers how to become better problem solvers (Greiff, Wüstenberg, & Avvisati, 2015) and goes hand in hand with the idea of intelligent measurement that proposes an assessment in which performance is continuously assessed and combined with learning opportunities (Bunderson et al., 1989). Only if there is a known set of strategic CPS behaviors translating intelligence into CPS performance that can be considered almost complete for a certain CPS assessment tool, conceptualizations of intelligent measurement can be implemented.

As previously outlined, a set of multiple effective strategic behaviors should be examined to gain a better understanding of the intelligence-CPS relationship and support learners. This idea is in line with theoretical considerations as well as empirical findings, that multiple strategic behaviors need to be combined in order to effectively reveal the relations of the complex system (Nicolay et al., 2023). For example, in a CPS task that is comprised of direct and eigendynamic effects, an effective behavior would be to first identify eigendynamic effects by using NOTAT and then identify direct effects by using VOTAT. This example illustrates that the strategic behavior to identify eigendynamic effects early requires that participants have the means to do so effectively, for example, by using NOTAT. Therefore, future studies should investigate the relationship between the strategic behavior to identify eigendynamics early and further effective CPS behaviors, especially NOTAT, which is theoretically closely linked. Such an investigation could reveal how many of the students apply NOTAT effectively in their sequence of actions, that is, at the beginning of the exploration process. This question highlights that the use of NOTAT and VOTAT is not necessarily always effective (e.g., depending on how and when the strategic behaviors are applied in the exploration phase) and does not necessarily lead to the correct conclusions about the structure of the complex system (Kuhn & Dean, 2005; Rollett, 2008; Stadler et al., 2019). Taken together, clarifying the relationships among strategic CPS behaviors is a keystone to understanding and promoting CPS performance, and sheds further light on how intelligence contributes to successful problem solving.

The yet unknown relationships among strategic CPS behaviors leads to the question whether these strategic behaviors differ in a qualitive way. That is, the strategic behavior to identify eigendynamic effects early could be viewed as a behavior that orchestrates the effective use of NOTAT (i.e., early in the presence of multiple types of effects) and VOTAT (i.e., after identifying eigendynamic effects). In line with this idea, identifying eigendynamics early requires the use of a strategic behavior such as NOTAT (or some other strategic behavior to identify the effect); but conversely, NOTAT can be used without the strategic behavior to identifying eigendynamics early (which would be ineffective for the exploration process). Therefore, the question arises if identifying eigendynamics early is a meta-strategic behavior rather than a strategic behavior, because it includes knowledge on how to use strategic behaviors, such as VOTAT and NOTAT (e.g., Nicolay et al., 2023). Although research on meta-strategic behavior or metacognition in CPS is yet scarce, first results showed that knowledge about problem solving strategies is related to the application of VOTAT (Scherer & Tiemann, 2012; Vollmeyer & Rheinberg, 1999; Wüstenberg, Stadler, et al., 2014; Zohar & Peled, 2008). In a recent study, Stadler et al. (2019) found that among students who applied VOTAT, it was beneficial for CPS performance to plot the findings immediately in the causal diagram after exploring the system's relations. The authors of the study concluded that not only the application of VOTAT, but also the handling of the information gained is an important factor in successful problem solving. In a similar way, Rollett (2008) differentiated between the generated information about the system and the used information about the system and showed that even though the application of strategic behaviors (e.g., NOTAT and VOTAT) led to a sufficient amount of information, the majority of the generated information was not used. Thus, a major problem while exploring the system seems to be the meta-strategic knowledge how to recognize and extract the potentially useful information that students generate through their strategic behavior (see also Nicolay et al., 2023; Stadler et al., 2019). The results described above suggest that there appear to be qualitative differences between the use of strategic behaviors and the meta-strategic knowledge about how to use them. For the research conducted in this dissertation project, this distinction was not relevant because the studies considered CPS behaviors individually and not in concert. However, in future studies that aim to investigate a variety of CPS behaviors in concert to further understand CPS performance and the intelligence-CPS relationship, indicators of meta-strategic behavior should be considered, because they reflect how students actually use effective strategic behaviors and handle the information gained in this process. Such indicators of meta-strategic knowledge are yet scarce, however the investigated behavior to identify eigendynamic effects early (see study 2) and to plot the findings immediately in the causal diagram (Stadler et al., 2019) are starting points for future research. Given that many meta-strategic processes are still unknown, rather abstract, and of a higher order than strategic behaviors such as VOTAT, thinking aloud protocols would be a first approach to discover what further meta-strategic behaviors might look like and how they can be captured in a standardized way (Vollmeyer & Rheinberg, 1999; see also Klopp et al., 2020).

4.1.3 Complex problem solving

With regard to CPS performance, the studies 1 through 3 used the computer-based assessment tool MicroDYN, which is based on the MCS approach (Funke, 2001; Greiff et al., 2012). The MCS approach has been criticized for its alleged lack of ecological validity and low complexity (Dörner & Funke, 2017; Süß & Kretzschmar, 2018). However, skepticism regarding the ecological validity can be refuted by recent findings on the benefit of MCS assessment tools in educational contexts (Lotz et al., 2016, 2017; Sonnleitner et al., 2013; Stadler et al., 2018; see also Greiff, Stadler, et al., 2015). Concerning the complexity of CPS tasks, research has shown that this CPS characteristic is closely related to task difficulty (e.g., Stadler, Niepel, & Greiff, 2016). In the present sample, individual MicroDYN tasks had satisfactory task difficulties between $M = .10 – .79$ ($SD = .30 – .48$) in the knowledge acquisition phase and between $M = .21 – .89$ ($SD = .31 – .49$) in the knowledge application phase. Thus, a wide range of difficulties was covered, and the difficulties suggest that the complexity of the task set was appropriate for the sample. In addition, the interplay between process measures, intelligence, and CPS performance demonstrated in studies 1 and 2 further supports the notion that the MicroDYN tasks consist of complex systems that place appropriate demands on intelligence. Taken together and despite the alleged disadvantages, the studies presented demonstrate the excellent suitability of MCS-based assessment tools, such as the MicroDYN, to investigate how intelligence translates into problem solving and by what factors this relationship is influenced.

In futures studies, more effect types could be included in an MCS-based assessment tool. Such tasks would certainly show an increased face validity and complexity; however, the tasks could easily become too difficult (Stadler, Niepel, & Greiff, 2016). Compromising on the empirical findings on task difficulty and the idea of more complex systems, an MCS-based assessment tool could be developed that, overall, includes more effect types (e.g., direct effects, eigendynamics, interactions, delayed effects, etc.), but single tasks should combine only a limited number of effect types in order to keep difficulty within a reasonable range. The question of whether an increased face validity and complexity, as suggested above, would lead to improved predictions of real-life outcomes (e.g., educational success) remains open. However, such an approach could reduce the effects of training (e.g., Schoppek & Fischer, 2017) if MCS assessment tools continue to gain popularity in educational contexts and address the criticism that has been raised.

4.1.4 The *mini-q* as a time-efficient intelligence screening

In study 4, we investigated the psychometric properties of the *mini-q* (Baudson & Preckel, 2016) and introduced a parallel version, the *mini-q B*. Both test versions can be administered very time-efficiently within three minutes and are available from the test authors for non-commercial use in research projects (Baudson & Preckel, 2016; Weise et al., 2024). The ongoing investigation of the psychometric properties of the *mini-q*, as conducted in study 4 of this dissertation project, contributes to the establishment of a well-evaluated screening instrument that is particularly useful when time is limited. Although the indicators of convergent validity for the *mini-q* (and partially also for the *mini-q B*) reported in study 4 were lower than expected, they suggest that the screening instrument and its parallel version are valuable tools for capturing a rough score of intelligence. This assumption is supported by a recent study by Schubert et al. (2024), who found a correlation between the *mini-q* and a broad measure of intelligence of $r = .57$, similar to the results reported by the test authors (Baudson & Preckel, 2016: $r = .67$).

In the context of the present dissertation project, the *mini-q* and its parallel version offer a number of opportunities for future research, but also come with some limitations. Because the screening instrument captures a rough measure of intelligence, it is less suitable for obtaining a concise estimate of the relationship between intelligence and CPS. As outlined in the present thesis, such an investigation should use a broad measure of intelligence (i.e., "good *g*"; Jensen & Weng, 1994; see also Lotz et al., 2016). Nevertheless, because both CPS performance and CPS process measures are related to intelligence (e.g., Kretzschmar et al., 2016; Kröner et al., 2005; Lotz et al., 2017; Stadler et al., 2015; Wüstenberg, Stadler, et al., 2014), the *mini-q* and its parallel version may play an important role in future research investigating the interplay of intelligence, CPS performance, and CPS process measures when only a rough indicator of intelligence is needed. For example, future research investigating the relationship between the strategic behavior to identify eigendynamic effects early and NOTAT should incorporate a rough indicator of intelligence. Potentially, the relationship between the strategic behavior to identify eigendynamic effects early and NOTAT could vary for different levels of cognitive abilities. For such a moderating effect, the rough indicator of intelligence captured by the *mini-q* might be sufficient and offers great advantages in terms of data collection (e.g., when administered in schools with limited time slots). Furthermore, in research projects that examine the development of exploration behaviors over a series of CPS tasks, a rough indicator of intelligence might be sufficient to estimate whether an increase/decrease in a particular behavior is related to intelligence. Moreover, in future (experimental) research projects investigating how strategic behaviors can be supported in learning environments (e.g., teaching/transferring strategic or meta-strategic behaviors), a rough indicator of intelligence might be sufficient to control for different levels of cognitive abilities in (groups of) students.

Taken together, the examples outlined above illustrate how the *mini-q* and its parallel version can contribute to future research investigating the interplay between intelligence, CPS performance, and CPS process measures. As CPS is often studied in educational contexts (e.g., Csapó & Molnár, 2017; Kretzschmar et al., 2014; OECD, 2013), data collection may take place in schools with limited time slots for data collection. Therefore, the evaluation and expansion of time-efficient screening tools, such as the *mini-q*, are of great value and expand the possibilities for future research.

4.2 Limitations and future research

The following sections address noteworthy limitations of the three core studies 1 through 3 by considering the design and methodological approach of the studies, the instruments used, and the design of the data set.

4.2.1 Study design

Study 1 investigated the moderating effect of prior knowledge, operationalized as the strategic behavior to apply VOTAT, on the intelligence-CPS relationship (Elshout-Raaheim hypothesis; Leutner, 2002; see Elshout, 1987; Raaheim, 1988). In accordance with the hypothesis, we found that prior knowledge increased from task to task over the set of CPS tasks and therefore, the degree of prior knowledge varied between the tasks. However, study 1 did not consider a variation of VOTAT between participants, which should be assumed based on prior studies (Greiff et al., 2016; Lotz et al., 2017; Vollmeyer et al., 1996). Therefore, prior strategic knowledge does not only vary between tasks, but it varies also between participants within one task. Theoretically, an investigation that considers the described variation between participants would have been possible on the basis of the present dataset; however, it was not pursued in order to limit the scope of the study. To investigate the assumed variation of prior strategic knowledge between participants in future studies, two viable options can be outlined: First, the moderating effect could be analyzed separately for each CPS task. Following this idea, the Elshout-Raaheim hypothesis would be transformed into a (logistic) regression design (i.e., dichotomous CPS task performance on intelligence moderated by the frequency of the VOTAT application). A second possible design would be to consider the variance between participants at a task-aggregated level and to test the following question: Is the regression of the aggregated variable CPS performance on intelligence moderated by an aggregated score of the VOTAT application rate? Following the Elshout-Raaheim hypothesis, a curvilinear moderation of prior strategic knowledge on the intelligence-CPS relationship should also be found when considering the VOTAT variation between participants. Moreover, prior knowledge could be operationalized more broadly by multiple strategic behaviors in future studies, for example, VOTAT, NOTAT and to identify eigendynamics early. Such a broader operationalization could show a clearer picture of the moderating effect assumed by the Elshout-Raaheim hypothesis. To realize this operationalization of prior strategic knowledge, the respective CPS assessment tool should include at least five tasks, each of them comprised of direct and eigendynamic effects.

A limitation of study 2 concerns the number and arrangement of the CPS tasks that comprised eigendynamic effects. As the strategic behavior to identify eigendynamics early is an effective behavior, it should increase over a set of tasks that comprise eigendynamic effects (see Lotz et al., 2017). However, the utilized CPS assessment tool only included three tasks with eigendynamics and, moreover, the three tasks were not arranged one after the other (i.e., tasks six, eight, and nine in a set of nine CPS tasks). Therefore, the development of the strategic behavior to identify eigendynamics early could not be adequately investigated in the present study. Findings on the development of VOTAT suggest that further studies investigating a possible increase in the strategic behavior to identify eigendynamic effects early should incorporate at least five consecutive CPS tasks that comprise eigendynamic effects (see Lotz et al., 2017).

With regard to study 3, the person-centered approach (i.e., LCA) as well as the design of the CPS assessment tool limited the possibility to investigate whether the effective and efficient use of VOTAT mediated the intelligence-CPS relationship. The results of several previous studies have shown that the effective use of VOTAT mediates the relationship between intelligence and CPS, meaning that intelligence contributes to successful problem solving through the use of VOTAT. (Grežo & Sarmány-Schuller, 2022; Lotz et al., 2022; Wüstenberg, Stadler, et al., 2014). However, the question of whether the efficiency of the VOTAT application is a further mediator of the intelligence-CPS relationship could not be answered by our study. On a methodological level, the four different classes are neither theoretically nor statistically suitable to serve as a mediating variable to investigate the intelligence-CPS relationship. Furthermore, and with regard to the assessment tool, the instructions did not explicitly emphasize efficiency and the time limit in the knowledge acquisition phase was rather liberal. In order to investigate a potential mediating effect of the effective and efficient VOTAT use, problem solvers should be instructed to explore the problems "thoroughly and efficiently". Future studies should investigate whether a changed instruction changes the CPS working style of participants such that a more efficient VOTAT use is related to higher CPS performance scores for all groups of students. In such a study, a regression analysis could investigate a possible mediating effect of the effective and efficient VOTAT use on the intelligence-

CPS relationship. Nevertheless, the person-centered approach of our study showed that different exploration behaviors (i.e., less and more efficient) are related to CPS performance and intelligence, and provided first insights into the corresponding developments of exploration behaviors. If we had not used a person-centered approach, but had investigated the efficiency of the VOTAT application only in regression analyses, the development of exploration behaviors would not have been visible, and corresponding results might have been non-significant or inconclusive. Thus, the results of the person-centered approach regarding the effectiveness and efficiency of exploration behaviors are a valuable starting point for future research.

A major limitation, common to all three studies 1 through 3 in this dissertation project, is the cross-sectional design, which does not allow for causal inference. The correlations and regression analyses reflect only statistical relationships between variables, not necessarily causal relationships. For example, in the second study, the results suggest that the use of strategic behavior to identify eigendynamic effects early is associated with improved CPS performance and higher intelligence test scores. However, it is also possible that the early identification of eigendynamics is the byproduct of some other, as yet unstudied, behavior that participants who perform better on CPS tasks and intelligence tests engage in. To gather evidence of causality, experimental designs should be implemented. One such design might be to compare students who have been taught the strategic behavior to identify eigendynamics early with students who have been taught another strategic exploration behavior that is irrelevant to eigendynamic effects (e.g., VOTAT; see also Zohar & Peled, 2008). Similarly, to investigate the findings of the first study, an experimental study design could be adopted that implements different levels of prior knowledge through learning materials (e.g., Leutner, 2002, first study). Regarding study 3, which examined the role of effectiveness and efficiency of exploration behavior, the research questions are less suitable for an experimental design (i.e., efficient exploration behavior may be related to CPS performance and intelligence, but it is most likely not the cause of successful problem solving). Taken together, especially our findings on the strategic behavior to identify eigendynamic effects early and the Elshout-Raaheim hypothesis should be further investigated in experimental designs to provide evidence for causal relationships.

4.2.2 CPS assessment tool

In all three studies, CPS was assessed by performance in the knowledge acquisition phase and the knowledge application phase of MicroDYN (Greiff et al., 2012), an MCS assessment tool based on Linear Structural Equation systems (LSE; Funke, 2001). MicroDYN is an instrument that meets high psychometric criteria (e.g., Greiff et al., 2012; Greiff & Neubert, 2014; Schoppek & Fischer, 2017; Wüstenberg et al., 2012) and represents many of the criteria outlined by common definitions of CPS, such as the need for interaction, a dynamically changing system, and opaqueness (see Buchner's definition in Frensch & Funke, 1995; see also Dörner et al., 1983). However, a limitation of all three studies presented within this dissertation project concerns the generalizability of findings for different CPS assessment tools.

Concerning study 2 that investigated the strategic behavior to identify eigendynamics early, the result should be generalizable to other CPS assessment tools based on the LSE approach that can also be comprised of eigendynamic and direct effects, such as Genetics Lab (Sonnleitner et al., 2012) and MultiFlux (Christ et al., 2020). For some CPS assessment tools based on LSE, slight adaptations regarding the operationalization of the investigated strategic behaviors might be necessary. For example, in the LINAS assessment tool (Schoppek, 2002), the eigendynamic effect only unfolds as soon as the outcome variable is moved to values different from zero. After a first input that moves the outcome variables from its zero state, the input variables can be set to zero to observe possible eigendynamics, a strategy referred to as *PULSE* (Schoppek & Fischer, 2017). Thus, in the LINAS assessment tool, the effectiveness of the strategic behavior to identify eigendynamics early and its positive relations to CPS performance and intelligence should hold true, except that an effective behavior to detect eigendynamics would be PULSE instead of NOTAT. However, the effectiveness of the strategic behavior to identify eigendynamics early does not necessarily apply to all assessment tools based on LSE. If complex systems are comprised of eigendynamic effects combined with different types of effects such as delayed effects, interactions, indirect effects, or further possible effects, the strategic behavior to identify eigendynamics early could possibly not be effective any more. This general consideration holds especially true for more complex CPS assessment tools such as Tailorshop (Putz-Osterloh, 1981) and FSYS (Wagener, 2001). These CPS tools exhibit a considerably more complex system structure that is comprised of numerous effect types. In addition, these CPS assessment instruments do not separate the knowledge acquisition phase from the knowledge application phase. Therefore, the results of our study on the strategic behavior to identify eigendynamic effects early are likely to be only partially generalizable to these CPS assessment tools.

Regarding the results' generalizability of study 1 (moderation of the intelligence-CPS relationship) and study 3 (effective and efficient VOTAT behavior), the same applies as outlined for study 1: The results should be transferable to similar LSE-based assessment tools, but not to more complex tools that include different types of effects. Overall, the effectiveness of strategic behaviors depends on the types of effects used in the CPS assessment tools and, therefore, the individual system architecture of each tool must be considered when generalizing findings.

4.2.3 Intelligence

Concerning intelligence, a broad factor of general intelligence was extracted by utilizing a balanced set of ten diverse and reliable intelligence subtests of the BIS (Jäger et al., 1997; see also Valerius & Sparfeldt, 2014). Therefore, the extracted *g* factor meets the requirements of a "good *g*" (Jensen & Weng, 1994) and a sufficiently good indicator of general intelligence. However, and beyond the question of the relationship between intelligence and CPS, study 3 in particular highlights that there may be predictors of CPS performance other than cognitive ability that are related to specific exploration behaviors (e.g., double-checking; see class 2 in the respective study). Therefore, further studies investigating how strategic behaviors translate into CPS performance should exploratively broaden the range of predictors to include personality traits (e.g., conscientiousness) and motivational aspects.

4.2.4 Age and educational background of participants

A further limitation of the presented studies concerns the generalizability of our findings to other age groups and to different educational backgrounds. The sample (N = 495) of the three studies stemmed from two academic-tracked school types (Gymnasium and Gemeinschaftsschule) located in the federal state of Saarland and students' average age was M = 16.40 (SD = .94) years.

Concerning the age of participants, prior studies have shown that even younger students entering secondary school use the strategic behavior VOTAT (e.g., 6th grade Finish students; Wüstenberg, Stadler, et al., 2014). This is consistent with the fact that VOTAT is taught in secondary school science classes in many countries. As VOTAT is related to intelligence (Lotz et al., 2017), it is most likely to develop with students' cognitive abilities. Therefore, we expected that the findings of study 1 and 3, which incorporated intelligence, CPS performance, and VOTAT, would be transferrable to younger children, most likely starting from 5th grade on. In study 2, the strategic behavior to identify eigendynamics early was only applied by 13% to 19% of the students in our sample (10% to 21% solved the tasks correctly). These proportions can be regarded as rather low. Because the strategic behavior studied was related to intelligence and therefore can be expected to develop along with students' cognitive abilities, younger students may not be able to use this specific strategic behavior. The age of roughly 15 to 16 years investigated in our sample might be the starting point from which on students use the strategic behavior to identify eigendynamic effects early.

In a sample of older (i.e., adult) participants, we expect an increased application of the strategic behaviors compared to younger participants. For example, in a sample of university students (age: $M = 22.8$, $SD = 4.0$), VOTAT was consistently applied for each input variable of the system by 74% of participants in the first task and by 90% in the last task in a set of eight MCS tasks (Wüstenberg et al., 2012). These results suggest that there may be ceiling effects in the VOTAT application rate associated with a reduction in variance. Such ceiling effects in a sample of older participants could lead to a weaker or even non-replicable pattern of results in studies 1 and 3, in which VOTAT is an important indicator. To avoid the described reduction in variance of the VOTAT application rate in an older sample, more difficult CPS tasks should be used (see Stadler, Niepel, & Greiff, 2016). With such a set of CPS tasks of appropriate difficulty, a similar pattern of results should be found for studies 1 and 3 in an older sample. Concerning study 2, the strategic behavior to identify eigendynamics early has not yet been investigated in a sample of older participants. However, in such a sample we would expect an increased use of this strategic behavior, but no ceiling effects due to its low application rate in high school students as shown in study 2. Therefore, the general hypothesis, that students who identify eigendynamic effects early will perform better in CPS tasks than students who do so in a later step should also hold true for a sample of older participants.

Regarding the educational background of the sample, at the time of data collection in 2014, about 50% of the cohort achieved the general university entrance qualification in Germany (“Allgemeine Hochschulreife”, “Fachhochschulreife”; Bundesministerium für Bildung und Forschung, 2021), and in the state of Saarland, even 60% achieved a qualification for university entrance (Malecki, 2016). This statistic reflects the fact that the surveyed sample captured the majority of students in the cohort. However, as the sample did not include students from non-academic school types (e.g., “Realschule”), cognitively weaker students were not included. We expect that the use of strategic exploration behaviors varies by educational background as it is related to cognitive abilities (Strenze, 2007). Regarding studies 1 and 3, we would expect a lower rate of VOTAT application for students from non-academic-tracked school types, but the same pattern of results. Regarding the strategic behavior to identify eigendynamic effects early examined in study 2, students with lower levels of educational achievement would most likely use this strategic behavior very rarely, if at all.

Taken together, our findings from the three studies are generalizable to older participants, but may only be replicable in younger children beyond a certain age. As age and education are both related to cognitive abilities, they confound each other, and we expect that lower levels of education may be partially compensated by age. In future studies, students from non-academic-tracked school types should be included in the sample to ensure generalizability across all levels of cognitive ability.

4.3 Final conclusion

The overall aim of the present dissertation project was to shed light on the fuzzy relationship between the two prominent constructs intelligence and complex problem solving. Both constructs are well researched and each of them plays an important role in educational science and psychology. Typically, in CPS assessment tools, test takers’ actions can be extracted from log-file data and meaningful patterns of observable behavior can be isolated. With respect to the currently unresolved relationship between intelligence and CPS, the aforementioned patterns of behavior provide a valuable opportunity to gain insight into how intelligence translates into successful problem solving and under what conditions intelligence enhances successful problem solving. Accordingly, strategic behaviors derived from log-file data, such as VOTAT, NOTAT, and the behavior to identify eigendynamic effects early, which was investigated for the first time in this dissertation project, can be viewed as a window into the black box of cognitive processing that improves our understanding of the intelligence-CPS relationship.

In the first study, we investigated the Elshout-Raaheim hypothesis for the first time in an MCS-based assessment tool of CPS. We considered the relative frequency of VOTAT use as an indicator of prior (strategic) knowledge concerning the exploration process of CPS tasks and found supportive results showing that the pattern of correlation coefficients followed an inverted U-shaped curve in both the knowledge acquisition and the knowledge application phase across a set of CPS tasks. This moderating effect of prior knowledge on the relationship between CPS and intelligence suggests that intelligence plays a more important role in problem solving when participants have acquired medium levels of prior knowledge and a less important role when participants have acquired either rather low or rather high levels of prior knowledge.

In the second study, we investigated the strategic behavior to identify eigendynamic effects early, which was investigated for the first time and relates to how participants deal with the important feature of eigendynamic effects in CPS. The examined strategic behavior predicted the performance in both CPS phases for the respective tasks and for the task set. Moreover, the mediation models showed that intelligence manifests itself in the strategic behavior to identify eigendynamics early, which goes hand in hand with higher CPS performance in the knowledge acquisition phase. Regarding the knowledge application phase, the pattern of results suggests that more intelligent students achieved higher CPS performance scores by adequately identifying eigendynamics, but regardless of whether they did so in an early or later exploration step.

In the third study, we investigated the effectiveness and efficiency of students' exploration behavior and its relation to intelligence and CPS performance. We found an increasingly consistent use of VOTAT and, going beyond previous studies, an increasingly efficient use of strategic exploration behavior across the sequence of tasks. Using a person-centered approach, we found four distinct (latent) classes of students. With respect to these classes, the pattern of results highlights the importance of effectiveness for successful problem solving and its relationship to intelligence. However, classes of students who explored effectively but either more or less efficiently, showed comparable intelligence and CPS test scores. Thus, a pattern of increasingly efficient exploration behavior across tasks was found, but the relationship of efficiency to CPS performance and intelligence requires further research.

In the fourth study, we investigated the psychometric properties of the very time-efficient intelligence screening *mini-q* (Baudson & Preckel, 2016) and introduced a parallel version, the *mini-q B*. Both versions of the screening tool may be valuable instruments in future research projects examining the interplay between intelligence, CPS performance, and CPS process measures when only a rough measure of intelligence is needed and time for data collection is limited, as is often the case in educational contexts.

Taken together, studies 1 through 3, which form the core of this dissertation project, illustrate how observable, meaningful patterns of behavior extracted from log-file data of CPS tasks contribute to the clarification of the intelligence-CPS relationship. As demonstrated, they can serve as (statistical) moderators and (statistical) mediators of the relationship between intelligence and CPS. That is, depending on the research question, the study design, and considerations on the theoretical status of the respective indicators, they can (statistically) affect the strength of the association between the two constructs (i.e., moderation) and/or show how intelligence manifests itself in observable and meaningful behavior that translates into CPS performance (i.e., mediation). In addition, study 4 evaluated and extended a very time-efficient intelligence screening instrument, expanding the possibilities for future research in the field of CPS. Therefore, the studies of this dissertation project have made significant contributions to clarify the relationship between intelligence and CPS and have identified numerous new and promising research avenues. Both, the moderation and the mediation of the intelligence-CPS relationship, have theoretical and practical implications in psychology and educational science. Elucidating moderating effects reveals the circumstances under which students can best use their cognitive abilities to solve complex problems. The investigation of mediating effects shows which strategic behaviors contribute to the intelligence-CPS relationship. Thus, our findings provide insight into how to adapt learning environments and support students in order to promote successful problem solving.

References

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

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, *10*, 341–342. https://doi.org/10.3758/BF03331551

Baudson, T. G., & Preckel, F. (2016). mini-q: Intelligenzscreening in drei Minuten [mini-q: Intelligence screening in three minutes]. *Diagnostica*, *62*, 182–197. https://doi.org/10.1026/0012-1924/a000150

Beauducel, A., Liepmann, D., Horn, S., & Brocke, B. (2010). *Intelligence Structure Test: English version of the Intelligenz-Struktur-Test 2000 R (IST 2000 R)*. Hogrefe.

Beckmann, J. F. (1994). *Lernen und komplexes Problemlösen: Ein Beitrag zur Konstruktvalidierung von Lerntests* [Learning and complex problem solving: a contribution to the construct validation of learning tests]. Holos.

Beckmann, J. F., & Goode, N. (2014). The benefit of being naïve and knowing it: the unfavourable impact of perceived context familiarity on learning in complex problem solving tasks. *Instructional Science*, *42*, 271–290. https://doi.org/10.1007/s11251-013-9280-7

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). Erlbaum.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, *93*, 296–316. https://doi.org/10.1037/0021-9010.93.2.296

Brehmer, B. (1987). Development of mental models for decision in technological systems. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New Technology and Human Error* (p. 111–120). Wiley.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, *81*, 211–241. https://doi.org/10.1016/0001-6918(92)90019-a

Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*, 171–184. https://doi.org/10.1016/0747-5632(93)90005-d

Brody, N. (1999). What is intelligence?. *International Review of Psychiatry*, *11*, 19–25. https://doi.org/10.1080/09540269974483

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (pp. 367–407). Macmillan Publishing.

Bundesministerium für Bildung und Forschung. (2021.) *Anteil der Studienberechtigten an der altersspezifischen Bevölkerung (Studienberechtigtenquote) nach Art der Hochschulreife* [Proportion of those eligible for higher education in the age-specific population by type of university entrance qualification]. https://www.datenportal.bmbf.de/portal/de/Tabelle-2.5.85.html

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*, 404–431.

Carroll, J. B. 1993. *Human Cognitive Abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/cbo9780511571312

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1–22. https://doi.org/10.1037%2Fh0046743

Chen, Y. C., Pan, Y. T., Hong, Z. R., Weng, X. F., & Lin, H. S. (2020). Exploring the pedagogical features of integrating essential competencies of scientific inquiry in classroom teaching. *Research in Science & Technological Education*, *38*, 185−207. https://doi.org/10.1080/02635143.2019.1601075

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*, 1098–1120. https://doi.org/10.1111/1467-8624.00081

Christ, A., Becker, N., & Kröner, S. (2020). Multiple complex problem-solving scenarios: The incremental validity of ability self-concept beyond reasoning in adults. *Intelligence*, *78*, 101421. https://doi.org/10.1016/j.intell.2019.101421

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Wiley. https://doi.org/10.1002/9780470567333

Coyle, T. R., & Greiff, S. (2021). The future of intelligence: The role of specific abilities. *Intelligence*, *88*, 101549. https://doi.org/10.1016/j.intell.2021.101549

Csapó, B., & Funke, J. (2017). The development and assessment of problem solving in 21stcentury schools. In B. Csapó & J. Funke (Eds.), *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning* (pp. 19–32). Educational Research and Innovation, OECD Publishing. https://doi.org/10.1787/9789264273955-en

Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, *8*, 277958. https://doi.org/10.3389/fpsyg.2017.02022

Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist*, *52*, 1038−1045. https://doi.org/10.1037/0003-066x.52.10.1038

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, *39*, 323–334. https://doi.org/10.1016/j.intell.2011.06.004

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Gaming*, *11*, 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*, 1153. https://doi.org/10.3389/fpsyg.2017.01153

Dörner, D., & Kreuzig, H. W. (1983). Problemlösefähigkeit und Intelligenz [Problem solving ability and intelligence]. *Psychologische Rundschau, 34*, 185–192.

Dörner, D., & Schaub, H. (1994). Errors in planning and decision-making and the nature of human information processing. *Applied Psychology*, *43*, 433–453. https://doi.org/10.1111/j.1464-0597.1994.tb00839.x

Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität* [Lohhausen. On dealing with uncertainty and complexity]. Huber.

Dörner, D., Stäudel, T., & Strohschneider, S. (1986). *Moro: Programmdokumentation* [Moro: Program documentation] (Memorandum No. 23). University of Bamberg, LS Psychologie II.

Eichmann, B., Goldhammer, F., Greiff, S., Brandhuber, L., & Naumann, J. (2020). Using process data to explain group differences in complex problem solving. *Journal of Educational Psychology*, *112*, 1546–1562. https://doi.org/10.25657/02:23272

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*, 933–956. https://doi.org/10.1111/jcal.12451

Elshout, J. J. (1987). Problem solving and education. In E. DeCorte, H. Lodewijks, R. Parmentier, & P. Span (Eds.). *Learning and Instruction* (pp. 259−273). Pergamon.

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

Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* [Knowledge about dynamic systems: Acquisition, representation, and application]. Springer.

Funke, J. (1995). Experimental research on complex problem solving. In P. A. Frensch, & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 243–268). Erlbaum.

Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking & Reasoning*, *7*, 69–89. https://doi.org/10.1080/13546780042000046

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

Geiser, C. (2012). *Datenanalyse mit Mplus* [Data analysis with Mplus]. Guilford. https://doi.org/10.1007/978-3-531-93192-0

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

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*, 608–626. https://doi.org/10.25656/01:17967

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence*, *33*, 169–186. https://doi.org/10.1016/j.intell.2004.10.002

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23. https://doi.org/10.1016/ S0160-2896(97)90011-8

Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlosefähigkeit anhand minimal komplexer Systeme [Systematic research of complex problem solving ability using minimal complex systems]. *Zeitschrift für Pädagogik*, *56* (Beiheft), 216–227. https://doi.org/10.25656/01:3430

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., 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., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior*, *61*, 36–46. https://doi.org/10.1016/j.chb.2016.02.095

Greiff, S., Stadler, M., Sonnleitner, P., Wolf, 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:10.1016/j.compedu.2015.10.018

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, *36*, 189–213. https://doi.org/10.1177/0146621612439620

Grežo, M., & Sarmány-Schuller, I. (2022). It is Not Enough to be Smart: On Explaining the Relation Between Intelligence and Complex Problem Solving. *Technology, Knowledge and Learning*, *27*, 69–89. https://doi.org/10.1007/s10758-021-09498-2

Guthke, J., & Stein, H. (1996). Are learning tests the better version of intelligence tests?. *European Journal of Psychological Assessment*, *12*, 1–13. https://doi.org/10.1027/1015-5759.12.1.1

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*, 451–470. https://doi.org/10.1111/bmsp.12028

Hayes, J. R., & Simon, H. A. (1977). Psychological differences among problem isomorphs. In N. J. Castellan, D. B. Pisoni, & G. R. Potts (Eds.), *Cognitive theory* (2nd ed., pp. 21–44). Erlbaum.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Horn, J. L. (1991). Measurement of Intellectual Capabilities: A Review of Theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson technical manual* (pp. 197–232). Riverside.

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp.41–68). Guilford Press.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447–472. https://doi.org/10.1037/1076-8971.2.3-4.447

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. BIS-Test, Form 4* [Berlin Intelligence-Structure Test. Version 4]. Hogrefe.

Jensen, A. R. (1980). *Bias in mental testing*. Free Press.

Jensen, A. R. (1989). The relationship between learning and intelligence. *Learning and Individual Differences*, *1*, 37–62. https://doi.org/10.1016/1041-6080(89)90009-5

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Jensen, A. R., & Weng, L.-J. (1994). What is a good g? *Intelligence*, *18*, 231–258. https://doi: 10.1016/0160-2896(94)90029-9

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

Kaufman, A. S., Schneider, W. J., & Kaufman, J. C. (2019). Psychometric Approaches to Intelligence. In R. J. Sternberg (Ed.), *Human intelligence: An introduction* (1st ed., pp. 67–103). Cambridge University Press.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology*, *90*, 677–691. https://doi.org/10.1037/0021-9010.90.4.677

Kelley, T. L. (1928). *Crossroads in the mind of man: A study of differentiable mental abilities*. Stanford University Press.

Kersting, M. (1999). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztests* [Assessment and personnel selection with computer-simulated problem-solving scenarios? On the validity of problem-solving scenarios and intelligence tests]. Hogrefe.

Klahr D. (2000). *Exploring science: The cognition and development of discovery processes*. MIT Press.

Klauer, K. J. & Spinath, F. M. (2010). Anlage und Umwelt [Predisposition and environment]. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (4th ed., pp. 9–16). Beltz.

Klopp, E., Schneider, J. F., & Stark, R. (Eds.). (2020). *Thinking Aloud – The Mind in Action*. Bertuch.

Kluwe, R. H., Misiak, C., & Haider, H. (1991). The control of complex systems and performance in intelligence tests. In H. Rowe (Ed.), *Intelligence: reconceptualization and measurement* (pp. 227–244). Lawrence Erlbaum.

Kramer, J. (2009). Allgemeine Intelligenz und beruflicher Erfolg in Deutschland: vertiefende und weiterführende Metaanalysen [General intelligence and occupational success in Germany: In-depth and further meta-analyses]. *Psychologische Rundschau*, *60*, 82–98. https://doi.org/10.1026/0033-3042.60.2.82

Kretzschmar, A., Neubert, J. C., & Greiff, S. (2014). Komplexes Problemlösen, schulfachliche Kompetenzen und ihre Relation zu Schulnoten [Complex problem solving, school competencies and their relation to school grades]. *Zeitschrift für Pädagogische Psychologie*, *28*, 205–215. https://doi.org/10.1024/1010-0652/a000137

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

Kröner, S. (2001). *Intelligenzdiagnostik per Computersimulation* [Intelligence assessment via computer simulation]. Waxmann.

Kröner, S., Plass, J. L., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, *33*, 347–368. https://doi.org/10.1016/j.intell.2005.03.002

Kuhn, D., and Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*, 866–870. https://doi.org/10.1111/j.1467-9280.2005.01628.x

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all?. *Journal of Personality and Social Psychology*, *86*, 148–161. https://doi.org/10.1037/0022-3514.86.1.148

Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; 5. Aufl.)* [Multiple choice vocabulary intelligence test]. Spitta.

Leichner, N., Prestele, E., Matheis, S., Weis, S., Schmitt, M., & Lischetzke, T. (2022). Lehramt-Studienwahlmotivation sagt Zielorientierungen vorher, pädagogisches Wissen und selbst eingeschätzte Kompetenz aber nur teilweise [Motivation for becoming a teacher predicts goal orientation, but pedagogical knowledge and self-assessed competence only partially]. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. https://doi.org/10.1024/1010-0652/a000348

Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, *18*, 686–697. https://doi.org/10.1016/s0747-5632(02)00024-9

Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012). *IST-Screening*. Hogrefe.

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., Scherer, R., Greiff, S., & Sparfeldt, J.R. (2022). *g*'s little helpers – VOTAT and NOTAT mediate the relation between intelligence and complex problem solving. *Intelligence*, *95*, 101685. https://doi.org/10.1016/j.intell.2022.101685

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

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*, 825–845. https://doi.org/10.1007/s10869-018-9561-x

Malecki, A. (2016). *Schulen auf einen Blick* [Schools at a glance]. Statistisches Bundesamt.

Mayer, R. (2010). Problem solving and reasoning. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 273–278). Elsevier.

Mayer, R. E. (1992). *Thinking, problem solving, cognition*. WH Freeman/Times Books/Henry Holt & Co.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander, & P. H. Winne (Eds.). *Handbook of Educational Psychology* (pp. 287–303). Erlbaum. https://doi.org/10.4324/9780203874790.ch13

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136−181). Guilford.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Meuer, M., & Imhoff, R. (2021). Believing in hidden plots is associated with decreased behavioral trust: Conspiracy belief as greater sensitivity to social threat or insensitivity towards its absence? *Journal of Experimental Social Psychology*, *93*, 104081. https://doi.org/10.1016/j.jesp.2020.104081

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

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*, 2090–2101. https://doi.org/10.1037/dev0000764

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Muthen & Muthen.

Nalis, D., Schütz, A., & Pastukhov, A. (2018). The Bamberg Trucking Game: A paradigm for assessing the detection of win–win solutions in a potential conflict scenario. *Frontiers in Psychology*, *9*, 138. https://doi.org/10.3389/fpsyg.2018.00138

Naumann, J., Goldhammer, F., Rölke, H., & Stelter, A. (2014). Erfolgreiches Problemlösen in technologiebasierten Umgebungen: Wechselwirkungen zwischen Interaktionsschritten und Aufgabenanforderungen [Successful problem solving in technology-based environments: Interactions between interaction steps and task requirements]. *Zeitschrift für Pädagogische Psychologie*, *28*, 193–203. https://doi.org/10.25656/01:14830

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: knowns and unknowns. *American Psychologist*, *51*, 77–101. https://doi.org/10.1037/0003-066x.51.2.77

Nicolay, B., Krieger, F., Kuhn, J. T., Graesser, A. C., Ifenthaler, D., Baker, R., & Greiff, S. (2023). Unsuccessful and successful complex problem solvers – A log file analysis of complex problem solving strategies across multiple tasks. *Intelligence*, *101*, 101793. https://doi.org/10.1016/j.intell.2023.101793

Nicolay, B., Krieger, F., Stadler, M., Gobert, J., & Greiff, S. (2021). Lost in transition – Learning analytics on the transfer from knowledge acquisition to knowledge application in complex problem solving. *Computers in Human Behavior*, *115*, 106594. https://doi.org/10.1016/j.chb.2020.106594

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 510–520. https://doi.org/10.1037/0278-7393.14.3.510

Novick, L. R., & Bassok, M. (2005). Problem Solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning* (pp. 321–349). Cambridge University Press.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 398–415. https://doi.org/10.1037/0278-7393.17.3.398

Nylund, K., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569. https://doi.org/10.1080/10705510701575396

OECD (2013). *PISA 2012 Assessment and Analytic Framework*. Educational Research and Innovation, OECD Publishing. https://doi.org/10.1787/9789264190511-en

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin*, *136*, 65–86. https://doi.org/10.1037/a0017815

Osman, M., (2017). Problem solving: Understanding complexity as uncertainty. In B. Csapó & J. Funke (Eds.), *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning* (pp. 47–57). Educational Research and Innovation, OECD Publishing. https://doi.org/10.1787/9789264273955-en

Pollet, E., & Schnell, T. (2017). Brilliant: But what for? Meaning and subjective well-being in the lives of intellectually gifted and academically high-achieving adults. *Journal of Happiness Studies*, *18*, 1459–1484. https://doi.org/10.1007/s10902-016-9783-4

Preckel, F., & Brüll, M. (2008). *Intelligenztests* [Intelligence tests]. Reinhardt.

Putz-Osterloh, W. (1981). Über die Beziehung zwischen Testintelligenz und Problemlöseerfolg [On the relation between test intelligence and sucess in problem solving]. *Zeitschrift für Psychologie*, *189*, 79–100.

Putz-Osterloh, W. (1993). Strategies for knowledge acquisition and transfer of knowledge in dynamic tasks. *Advances in Psychology*, *101*, 331–350. https://doi.org/10.1016/S0166-4115(08)62664-3

Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.). *Advances in the Psychology of Human Intelligence* (pp. 73–97). Erlbaum.

Raven, J. C. (1998). *APM Raven's Advanced Progressive Matrices*. APA PsycTests. https://doi.org/10.1037/t10893-000

Rigas, G., & Brehmer, B. (1999). Mental processes in intelligence tests and dynamics decision making tasks. In P. Juslin, & H. Montgomerry (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 45–65). Lawrence Erlbaum.

Rigas, G., Carling, E., & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence*, *30*, 463–480. http://dx.doi.org/10.1016/s0160-2896(02)00121-6

Rollett, W. (2008). *Strategieeinsatz, erzeugte Information und Informationsnutzung bei der exploration und Steuerung komplexer dynamischer Systeme* [Strategy usage, generated information and used information in the exploration and control of complex dynamic systems]. LIT.

Rost, D. H. (2013). *Handbuch Intelligenz* [Handbook of Intelligence]. Beltz.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Ruby, J., Weise, J. J., Greiff, S., & Sparfeldt, J. R. (submitted). Double-checks or better no repeated steps? The Role of Effective and Efficient Exploration Behaviors for Successful Complex Problem Solving.

Scharfen, J., Peters, J. M., & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, *67*, 44–66. https://doi.org/10.1016/j.intell.2018.01.003

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, *32*, 102–119. https://doi.org/10.1037//0012-1649.32.1.102

Scherer, R., & Beckmann, J. F. (2014). The acquisition of problem solving competence: evidence from 41 countries that math and science education matters. *Large-scale Assessments in Education*, *2*, 1–22. https://doi.org/10.1186/s40536-014-0010-7

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., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, *48*, 37–50. https://doi.org/10.1016/j.intell.2014.10.003

Schlegel, K., & Mortillaro, M. (2019). The Geneva Emotional Competence Test (GECo): An ability measure of workplace emotional intelligence. *Journal of Applied Psychology*, *104*, 559–580. https://dx.doi.org/10.1037/apl0000365

Schleicher, A. (2017). Preface. In B. Csapó & J. Funke (Eds.), *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning* (pp. 19–32). Educational Research and Innovation, OECD Publishing. https://doi.org/10.1787/9789264273955-en

Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86*, 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*, 1479–1497. https://doi.org/10.1037/a0016810

Schmitz, F., & Wilhelm, O. (2016). Modeling mental speed: Decomposing response time distributions in elementary cognitive tasks and relations with working memory capacity and fluid intelligence. *Journal of Intelligence*, *4*, 1–23. https://doi.org/10.3390/jintelligence4040013

Schmitz, F., & Wilhelm, O. (2019). Mene mene tekel upharsin: Clerical speed and elementary cognitive speed are different by virtue of test mode only. *Journal of Intelligence*, *7*, 16. https://doi.org/10.3390/jintelligence7030016

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). The Guilford Press.

Schoppek, W. (2002). Examples, rules, and strategies in the control of dynamic systems. *Cognitive Science Quarterly*, *2*, 63–92.

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

Schubert, A. L., Löffler, C., Wiebel, C., Kaulhausen, F., & Baudson, T. G. (2024). Don't waste your time measuring intelligence: Further evidence for the validity of a three-minute speeded reasoning test. *Intelligence*, *102*, 101804. https://doi.org/10.1016/j.intell.2023.101804

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. http://dx.doi.org/10.1016/j.lindif.2012.12.011

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*, 227–237. https://doi.org/10.1016/s0959-4752(02)00022-1

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 to assess complex problem solving. *Psychological Test and Assessment Modeling*, *54*, 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, 289–305. https://doi.org/10.1016/j.intell.2013.05.002

Sparfeldt, J. R., Becker, N., Greiff, S., Kersting, M., König, C. J., Lang, J. W., & Beauducel, A. (2022). Intelligenz(tests) verstehen und missverstehen [Understanding and misunderstanding intelligence (tests)]. *Psychologische Rundschau*, *73*, 161–172. https://doi.org/10.1026/0033-3042/a000597

Spearman, C. (1904). "General intelligence", objectively determined and measured. *The American Journal of Psychology*, *15*, 107–197. https://doi.org/10.2307/1412107

Stadler, M. J., Becker, N., Greiff, S., & Spinath, F. M. (2016). The complex route to success: complex problem-solving skills in the prediction of university success. *Higher Education Research & Development*, *35*, 365–379. https://doi.org/10.1080/07294360.2015.1087387

Stadler, M., Becker, N., Godker, 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., Becker, N., Schult, J., Niepel, C., Spinath, F. M., Sparfeldt, J. R., & Greiff, S. (2018). The logic of success: The relation between complex problem-solving skills and university achievement. *Higher Education*, *76*, 1–15. https://doi.org/10.1007/s10734-017-0189-y

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

Sternberg, R. J. (1995). Expertise in complex problem solving: A comparison of alternative conceptions. In P. A. Frensch, & J. Funke (Eds.), *Complex Problem Solving: The European Perspective* (pp. 295–321). Erlbaum.

Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, *52*, 1030–1037. https://doi.org/10.1037/0003-066x.52.10.1030

Sternberg, R. J. (2013). Contemporary theories of intelligence. In W. M. Reynolds, G. E. Miller, & I. B. Weiner (Eds.), *Handbook of Psychology: Educational Psychology* (pp. 23–44). Wiley.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*, 401–426. https://doi.org/10.1016/j.intell.2006.09.004

Süß, H. M., & Kretzschmar, A. (2018). Impact of cognitive abilities and prior knowledge on complex problem solving performance – Empirical results and a plea for ecologically valid microworlds. *Frontiers in Psychology*, *9*, 626. https://doi.org/10.3389/fpsyg.2018.00626

Süß, H. M., Kersting, M., & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen [Intelligence and knowledge as predictors of success in computer simulated complex problems]. *Diagnostica, 37*, 334–352.

Süß, H. M., Oberauer, K., & Kersting, M. (1993). Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme [Intellectual abilities and the control of complex systems]. *Sprache & Kognition*, *12*, 83–97.

Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen* [Intelligence, knowledge and problem solving. Cognitive prerequisites for success in problem solving with computer-simulated problems]. Hogrefe.

Thurstone, L. L. (1936). The factorial isolation of primary abilities. *Psychometrika*, *1*, 175–182. https://doi.org/10.1007/bf02288363

Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.

Tschirgi, J. E. (1980). Sensible Reasoning: A Hypothesis about Hypotheses. *Child Development*, *51*, 1–10. https://doi.org/10.2307/1129583

Valerius, S., & Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory factor analyses using three test batteries. *Intelligence*, *44*, 120–133. https://doi.org/10.1016/j.intell.2014.04.003

Valerius, S., & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell [Correlations of general and specific intelligence factors with general and specific school performance in a nested factor model]. *Zeitschrift für Pädagogische Psychologie*, *29*, 101–108. https://doi.org/10.1024/1010-0652/a000151

Van der Maas, H. L. J., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2018). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management*, *44*, 249–279. https://doi.org/10.1177/0149206317702220

Veenman, M. V., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, *14*, 89–109. https://doi.org/10.1016/j.learninstruc.2003.10.004

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education*, *14*, 541–554. https://doi.org/10.1007/BF03172978

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

Wagener, D. (2001). *Psychologische Diagnostik mit komplexen Szenarios: Taxonomie, Entwicklung, Evaluation* [Psychological diagnostics with complex scenarios: Taxonomy, development, evaluation]. Pabst.

Weise, J. J., Becker, N., & Sparfeldt, J. R. (2024). Intelligenzmessung in drei Minuten – Evaluation des *mini-q* und Konstruktion einer Parallelversion [Assessment of intelligence in three minutes – An evaluation of the mini-q and the construction of a parallel form]. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. https://doi.org/10.1024/1010-0652/a000384

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving – Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502. https://doi.org/10.1016/j.intell.2020.101502

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2022). Focusing on eigendynamic effects promotes students' performance in complex problem solving: A log-file analysis of strategic behavior. *Computers & Education*, *189*, 104579. https://doi.org/10.1016/j.compedu.2022.104579

Weiß, R. H. (2008). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgetest – Revision (WS/ZF-R)* [General Intelligence Test Scale 2 - Revision]. Hogrefe.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment*, *20*, 23–45. https://doi.org/10.1080/10627197.2015.997614

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

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., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences*, *29*, 18–29. https://doi.org/10.1016/j.lindif.2013.10.006

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*, 172–223. https://doi.org/10.1016/j.dr.2006.12.001

Zohar, A., & Peled, B. (2008). The effects of explicit teaching of metastrategic knowledge on low-and high-achieving students. *Learning and Instruction*, *18*, 337–353. https://doi.org/10.1016/j.learninstruc.2007.07.001

Appendices

Appendix A: Publications

Study 1

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2020). The moderating effect of prior knowledge on the relationship between intelligence and complex problem solving – Testing the Elshout-Raaheim hypothesis. *Intelligence*, *83*, 101502. https://doi.org/10.1016/j.intell.2020.101502

Study 2

Weise, J. J., Greiff, S., & Sparfeldt, J. R. (2022). Focusing on eigendynamic effects promotes students' performance in complex problem solving: A log-file analysis of strategic behavior. *Computers & Education*, *189*, 104579. https://doi.org/10.1016/j.compedu.2022.104579

Study 3

Ruby, J., Weise, J. J., Greiff, S., & Sparfeldt, J. R. (submitted). Double-checks or better no repeated steps? The Role of Effective and Efficient Exploration Behaviors for Successful Complex Problem Solving. *Computers & Education*.

Study 4

Weise, J. J., Becker, N., & Sparfeldt, J. R. (2024). Intelligenzmessung in drei Minuten – Evaluation des mini-q und Konstruktion einer Parallelversion [Assessment of intelligence in three minutes – An evaluation of the mini-q and the construction of a parallel form]. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. https://doi.org/10.1024/1010-0652/a000384