Universität des Saarlandes

Fakultät für Empirische Humanwissenschaften

und Wirtschaftswissenschaft

Fachrichtung Bildungswissenschaften

Increasing Intelligence Test Scores: The Effects Of Watching Video Tutorials

Dissertation

zur Erlangung des akademischen Grades eines Doktors der

Naturwissenschaften der Fakultät HW

der Universität des Saarlandes

vorgelegt von:

M. Sc. Benedikt Schneider

aus Ulm

Saarbrücken, 2022

Dekan:

Prof. Dr. Jörn R. Sparfeldt, Universität des Saarlandes

Berichterstatter*innen:

Prof. Dr. Jörn R. Sparfeldt, Universität des Saarlandes

Prof. Dr. Robin Stark, Universität des Saarlandes

Prof. Dr. Oliver Dickhäuser, Universität Mannheim

Tag der Disputation: 05.07.2022

Acknowledgements

First and foremost, I want to thank Jörn R. Sparfeldt for the tremendous support and guidance over the years. I will never forget the chances you gave me and the steps you took in my direction. When starting the project, the path ahead seemed rocky, sometimes almost insurmountable. Nevertheless, you've always pushed me forward to my limit and beyond. Without you, all this simply would not have been possible. Thank you for always having my back when it counted.

Furthermore, I heartedly want to thank Frank M. Spinath and Nicolas Becker for their guidance and mentorship in the early years of my academic studies as well their valuable advice and help regarding lots of different aspects related to research. Additionally, I want to thank Franziska Perels for the opportunity to collect additional data.

Concerning the practical aspects of conducting research, I want to thank Luca Deeg, Annika Blum, Katrin Arendt, Stefanie Krämer (and also Julia Ruby) for their tremendous and reliable help gathering data and making sure everything was handled as intended. You've been amazing! Also, a thank you to Cristina Deeg, Khanda Adams, Birgit Schneider, and Petra Arnold for helping with all the strenuous bureaucratic necessities that otherwise would have set me straight into panic mode.

A significant thank you also belongs to my family as well as my former and current colleagues Rebecca Schneider, Christin Lotz, Verena Keimerl, Linda Sommerfeld, Julia Ruby, Sarah Schwabe, and Lisa Stark. You gave me companionship, valuable advice, comfort, and lots of things to laugh about. The little sun and boat gave me lots of reasons to smile over the years. Also, a dear thank you to Julius Weise for all the inspiring conversations and simply for being the one and only "*Julio*"!

Further, I also want to thank all the other friends supporting me over the years, including Michael Benker and Roland Jeske. In particular, I want to thank Werner Georg Seiler for his unwavering belief and constantly demonstrated friendship spanning over ten years now! Also, shoutout to Marco Koch, Maximilian Frey, and Simeon Schöne for being awesome. Importantly, I also want to thank Kai "*ACK*" Kawald for his valuable insight into the relevance of my research and the constantly stimulating philsopohical exchange about important contemporary matters including multimodal transportation vehicles, astronomic occurrences, and ornithological phenomena.

Finally, I want to thank my dearest Hans-Jörg. I’ve been pondering for years now what I might write in these acknowledgements to properly honor all the hardship we’ve shared. In the end, I think there simply is not enough space here to adequately express the gratitude I feel in light of your neverending support. You’ve always been there, even when you haven’t. Your patience has really been incredible. You have not been a rock. You’ve been my mountain. It has been a long and straining hike, but with you everything felt lighter. In the end, we did it. Love!

Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
BIS	Berlin Intelligence Structure (model or test; see text)
BOMAT	Bochumer Matrizentest
df	Degrees of Freedom
EG	Experimental Group
CG	Control Group
$CG_{Examples}$	Control Group (Inspecting and Working on Item Examples)
$CG_{Matrices}$	Control Group (Watching the Figural Matrices Video Tutorial)
DESIGMA-Advanced	Design A Matrix Advanced
GPA	Grade Point Average
IST 2000 R	Intelligenz-Struktur-Test 2000 Revision
I-S-T Screening	Intelligenz-Struktur-Test Screening
LPS	Leistungsprüfsystem
M	Mean
MANOVA	Multivariate Analysis of Variance
Mdn	Median
SD	Standard Deviation
WIT-2	Wilde Intelligenztest 2

List of Publications

The presented thesis is based on four studies comprising three articles that are published in peer-reviewed journals. Articles are available online through the respective publishing company.

Study 1 and Study 2:

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

Study 3:

Schneider, B., & Sparfeldt, J. R. (2021a). How to solve number series items: Can watching video tutorials increase test scores? *Intelligence*, *87*, 101547. https://doi.org/10.1016/j.intell.2021.101547

Study 4:

Schneider, B., & Sparfeldt, J. R. (2021b). How to get better: Taking notes mediates the effect of a video tutorial on number series. *Journal of Intelligence*, 9, 55. https://doi.org/10.3390/jintelligence9040055

In the following, I will use the term *we* when referring to these studies my co-authors and I have conducted together.

Summary

Intelligence is widely regarded as essential predictor for success in educational and organizational contexts. Correspondingly, test-takers often rely on approaches such as practice or coaching in order to maximize their intelligence test scores. However, test scores elevated by such approaches are not considered to reflect increases in dispositional cognitive ability. Accordingly, meaningful consequences can result when these elevated scores are used in selection and admission decisions. As particularly accessible and potentially very effective and efficient approach, test-takers might rely on watching video tutorials in order to prepare for an upcoming assessment. Nevertheless, the effects of watching video tutorials on intelligence test scores remain largely unknown. In this dissertation project, it was aimed to gain insight into the effects of watching video tutorials on intelligence test scores by conducting four empirical studies.

In Study 1, the effects of a video tutorials on test scores were examined using figural matrices as popular and widely used figural reasoning task type. As primary goal, we investigated the figural matrices test scores of an experimental group watching a short figural matrices video tutorial focusing on teaching the rules compared to a control group watching an irrelevant video. As secondary goal, we investigated validity aspects by examining the correlations between the figural matrices scores and the scores in an intelligence test. The results of Study 1 indicated substantially higher figural matrices test scores in the experimental group than in the control group. Moreover, the correlations revealed to be comparable and of substantial magnitude in both groups. Noteworthy, however, these correlations revealed to be higher in the separate groups than in the entire sample.

Study 2 was conducted to strengthen the evidence obtained in Study 1. Accordingly, the results were aimed to be replicated in a substantially larger and more diverse sample. Administering another version of the figural matrices task but the identical videos used in Study 1, the same result pattern emerged, revealing substantially higher test scores in the experimental group than in the control group, comparable validity coefficients in the expected magnitude in both groups, and higher validity coefficients in the separate groups than in the entire sample.

Expanding the evidence to other reasoning tasks, the effects of video tutorials were further examined in Study 3 on the basis of number series as another popular and widely used numerical task type. In addition to an experimental group watching a number series video tutorial based on an illustration model illustrating the commonalities of number series items as well as the typical processes and a control group watching a task-irrelevant video tutorial, a second control group inspecting and working on number series item examples was included to represent another approach test-takers might use when preparing for an assessment. Moreover, and besides the scores in another intelligence test, we included grades as an additional variable for investigating validity aspects. The results revealed to be in line with Study 1 & Study 2, as the experimental group watching the number series video tutorial demonstrated higher number series test scores compared to the control groups. Furthermore, and related to validity aspects, the correlations between the number series test scores and the intelligence test scores as well as the correlations between the number series test scores and grades revealed to be comparable and substantially high in all three groups.

Shifting the focus towards the effects on test-taking behavior, we conducted Study 4 to investigate whether watching a number series video tutorial also stimulates test-takers to take more notes, and whether using such notes can be assumed underlying the demonstrated test score increases. Correspondingly, we compared the number of items with notes (i.e., the notated relations between numbers) of the experimental group in Study 3 watching the number series video tutorial with the number of items with notes in the control group watching the irrelevant video tutorial. Furthermore, we inspected the correlation coefficients between the number of items with notes and the number series test scores in both groups and examined whether the number of items with notes mediated the effect of increased test scores after watching the number series video tutorial. Besides revealing the expected higher number of notes in the experimental group compared to the control group as well as substantial correlations between these notes and the number series test scores in both groups, the mediation analysis indicated that the effect of the number series video tutorial on number series tests scores was mediated by such a note-taking behavior.

To conclude, the results obtained in this dissertation project confirmed watching video tutorials as an effective and efficient approach to increase test scores. By demonstrating substantial test score increases using two widely used reasoning task-types and additionally gathering insight into validity aspects and changes in test-taking behavior, the relevance of watching video tutorials as an influence likely to be involved in a plethora of testing situations was emphasized.

Zusammenfassung

Intelligenz gilt weithin als einer der bedeutsamsten Prädiktoren für die Vorhersage von Erfolg in bildungs- und arbeitsbezogenen Kontexten. Entsprechend nutzen Testteilnehmende oft Ansätze wie Übung oder Coaching, um ihre Testwerte zu maximieren. Da derart gesteigerte Testwerte jedoch keine Zuwächse der intellektuellen Fähigkeiten kennzeichnen, können bei Verwendung solcher gesteigerten Testwerte in Auswahlkontexten bedeutsame Konsequenzen resultieren. Als besonders zugänglicher als auch zugleich potenziell sehr effektiver und effizienter Ansatz zur Testvorbereitung könnten Teilnehmende auf Video-Tutorials zurückgreifen. Dennoch sind die Auswirkungen der Nutzung eines solchen Ansatzes bislang weitgehend unbekannt. In dieser Dissertation wurde der Einfluss von Video-Tutorials auf Intelligenztestwerte anhand von vier empirischen Studien untersucht.

In Studie 1 wurden die Effekte von Video-Tutorials anhand figuraler Matrizen als weit verbreitete figurale Aufgabenart untersucht. Als primäres Ziel wurden die Testwerte in figuralen Matrizenaufgaben einer Experimentalgruppe, die vor Bearbeitung ein kurzes Video-Tutorial über die Regeln in figuralen Matrizen sah, mit den Testwerten einer Kontrollgruppe verglichen, der zuvor ein irrelevantes Video präsentiert wurde. Darüberhinaus wurden Validitätsaspekte anhand der Korrelationen zwischen den Testwerten in figuralen Matrizenaufgaben und den Testwerten in einem Intelligenztest als externe Variable betrachtet. Die Ergebnisse von Studie 1 zeigten erheblich höhere Testwerte in figuralen Matrizenaufgaben in der Experimentalgruppe als in der Kontrollgruppe. Zusätzlich zeigten sich vergleichbar hohe Korrelationen in beiden Gruppen. Beachtenswerterweise waren diese Korrelationen in den getrennten Gruppen numerisch höher als in der gesamten Stichprobe.

Studie 2 wurde durchgeführt, um die in Studie 1 gewonnenen Erkenntnisse zu stärken. Entsprechend wurden die Ergebnisse in einer substanziell größeren und heterogeneren Stichprobe repliziert. Mittels einer anderen Version der Matrizenaufgaben aus Studie 1 und den identischen Videos zeigte sich das gleiche Befundmuster von erheblich höheren Testwerten in der Experimentalgruppe als in der Kontrollgruppe, vergleichbaren Korrelationen in beiden Gruppen und höheren Korrelationen in den getrennten Gruppen als in der Gesamtstichprobe.

Um den Fokus der Effekte von Video-Tutorials weiter auszuweiten, wurden in Studie 3 Zahlenreihen als weit verbreitete numerische Aufgabenart untersucht. Zusätzlich zu einer Experimentalgruppe, die ein Video-Tutorial über die Gemeinsamkeiten von Zahlenreihen-Aufgaben und den typischen Prozessen basierend auf einem Modells zur Aufgabenbearbeitung sah, und einer Kontrollgruppe, der ein aufgaben-irrelevantes Video-Tutorial präsentiert wurde, wurde eine zweite Kontrollgruppe einbezogen, die als alternativen Ansatz der Testvorbereitung Zahlenreihen-Beispiele eigenständig inspizierte und bearbeitete. Für die Betrachtung von Validitätsaspekten wurden neben den Testwerten in einem anderen Intelligenztest auch die Abiturnoten einbezogen. Die Ergebnisse in Studie 3 zeigten analoge Befundmuster zu Studie 1 und Studie 2: In der Experimentalgruppe zeigten sich höhere Zahlenreihen-Testwerte als in den beiden Vergleichsgruppen; zusätzlich waren sowohl die Korrelationen zwischen den Zahlenreihen-Testwerten und den Testwerten im Intelligenztest als auch die Korrelationen zwischen den Zahlenreihen-Testwerten und den Abiturnoten in allen Gruppen vergleichbar hoch.

In Studie 4 wurde der Einfluss von Video-Tutorials auf das Testbearbeitungs-Verhalten fokussiert. Es wurde untersucht, ob nach Präsentation eines Zahlenreihen-Video-Tutorials die Teilnehmenden zu einer erhöhten Nutzung von Notizen angeregt wurden und ob ein derartiges Notizenverhalten als Mediator für die Testwertsteigerungen angenommen werden kann. Entsprechend wurde die Anzahl der Zahlenreihen-Aufgaben mit Notizen (d.h. notierte Beziehungen zwischen den Zahlen) der Experimentalgruppe aus Studie 3, die vor Aufgabenbearbeitung ein Zahlenreihen-Video-Tutorial sah, mit der Anzahl der Zahlenreihen-Aufgaben mit Notizen der Kontrollgruppe verglichen, der zuvor ein irrelevantes Video-Tutorial präsentiert wurde. Zusätzlich wurden die Korrelationen zwischen der Anzahl der Aufgaben mit Notizen und den Zahlenreihen-Testwerten untersucht und überprüft, ob der Effekt des Zahlenreihen-Video-Tutorials auf die Testwerte durch die Anzahl der Aufgaben mit Notizen mediiert wurde. Die Ergebnisse zeigten eine höhere Anzahl an Aufgaben mit Notizen in der Experimentalgruppe und substanziell hohe Korrelationen zwischen der Anzahl an Aufgaben mit Notizen und den Zahlenreihen-Testwerten in beiden Gruppen. Darüberhinaus zeigte die Mediationsanalyse, dass die Testwertsteigerungen in Zahlenreihen-Aufgaben nach Präsentation eines Zahlenreihen-Video-Tutorials durch das untersuchte Notizenverhalten vermittelt wurden.

Zusammenfassend kann festgehalten werden, dass die Nutzung von Video-Tutorials als effektiver und effizienter Ansatz zur Steigerung von Testwerten angesehen werden kann. Die Relevanz

eines solchen Ansatzes als Einfluss, der möglicherweise in vielen Testsituationen bereits eine Rolle spielt, wurde in dieser Dissertation anhand von substanziell hohen Testwertsteigerungen, Evidenzen zu Validitätsaspekten und Einfluss auf Testbearbeitungsverhalten näher verdeutlicht.

1. Introduction

Intelligence has been widely recognized as one of the most important predictors for success. Accordingly, a large body of evidence demonstrates substantial predictive validity for a variety of important outcomes in education (e.g., Deary et al., 2007; Rost, 2013; Roth et al., 2015) and work (e.g., Hülsheger et al., 2007; Schmidt & Hunter, 1998, 2004). Ideally, using the test scores underlying such predictions in selection or admission decisions will result in a good fit between a selected test-taker's cognitive ability and the cognitive demands for the position of interest. As key prerequisite, all participants are required to be tested under the same premise (e.g., no individual additional exposure to test material beyond what is intended in a test's manual), as violating this premise might otherwise endanger the validity of the assessment.

At the same time, there is a large body of research indicating that individuals can increase their intelligence test scores through sources such as retaking a test, practice, or coaching (e.g., Hausknecht et al., 2007; Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984; Scharfen et al., 2018). Typically, however, the correspondingly elevated test scores are not considered to reflect genuine increases of a person's underlying cognitive ability (e.g., Estrada et al., 2015; Haier, 2014; Jensen, 1998; te Nijenhuis et al., 2007). With larger differences between elevated test scores and intellectual ability, the consequences stemming from such a disparity become increasingly relevant. Ultimately, test-takers might end up in positions where the cognitive demands of the environment exceed their cognitive ability, potentially leading to adverse consequences not only for the test-taker (e.g., being overburdened by the cognitive demands), but also for the institution (e.g., a candidate not performing as expected, wasted resources) – in addition to the consequences imposed on suitable applicants that were not selected due to a limited number of vacant positions.

With the rapid development of technological advances, opportunities to access elements useful for increasing test scores have increased tremendously. In contrast to decades ago, insight into these elements can now be gained as easily as by clicking a button to start searching the internet. For example, using the keywords "*Practice Intelligence Test*" on popular search engines yields over

500.000.000 results[1], providing interested users with a broad selection of information including strategies, explanations, material, and item examples. These and more aspects relevant for practice and coaching can be approached simultaneously by watching video tutorials. Specifically searching for these video tutorials (e.g., *"Intelligence Test Video Tutorial"*) yields over 60.000.000 results[2], emphasizing the popularity of such an approach. However, little is known about the effects of watching such a video tutorial on intelligence test scores. Potentially, however, watching video tutorials before an assessment can serve as particularly valuable and attractive approach for individuals to increase their test scores. For example, different elements relevant for practice and coaching (e.g., providing explanations, showing item examples) can be combined and further enriched by various means (e.g., using illustrations). Furthermore, the information is typically presented in a compact and engaging time-frame, signifiying not only an effective, but also efficient approach of increasing test scores. Moreover, video tutorials can be accessed (and saved) according to individual preferences, thus allowing key elements to be watched (and re-watched) shortly before an assessment when it's most critical. Finally, as many video tutorials can be retrieved free of charge, traditional obstacles such as the necessity to invest resources can be overcome. In summary, watching video tutorials to prepare for an upcoming assessment might be a substantially meaningful approach to increase test scores, although little is known about the related effects.

In the presented thesis, the effects of watching such video tutorials were investigated in four studies. Besides experimentally examining the increases in intelligence test scores on the basis of different popular rule-based intelligence task types as primary goal, we also explored validity aspects and examined effects on test-taking behavior.

In the first study, we evidenced the effects of video tutorials by using figural matrices as a popular rule-based task type. Using a video tutorial explaining six widely used rules found in figural matrices, we conducted an experiment comparing the figural matrices test scores of an experimental group watching this video tutorial prior to the assessment with the figural matrices test scores of a control group watching a video about nutrition. Additionally, to investigate validity aspects, we examined the correlation coefficients between the figural matrices test scores and the scores of another intelligence test in both groups as well as in the entire sample and compared these correlation coefficients.

[1] Accessed January 8th 2022 using www.google.com
[2] Accessed March 21st 2022 using www.google.com

In the second study of the dissertation project, we conducted another experiment aiming for a more heterogeneous sample related to ability as well as a highly increased sample size to replicate and thus strengthen the results obtained in Study 1. Similar to Study 1 and using the same videos, we compared the figural matrices test scores of an experimental group watching a figural matrices video tutorial with the figural matrices test scores of a control group watching a video about nutrition and examined the correlation coefficients with another intelligence test.

In the third study, we expanded on the effects of video tutorials by focusing on number series as another popular rule-based intelligence task type. As basis for the number series video tutorial, we developed a number series illustration model derived from inductive-deductive iteration loops of inspecting in total 287 number series items found in various published tests (e.g., Heller & Perleth, 2000; Jäger et al., 1997; Kersting et al., 2008; Liepmann et al., 2007; Weiß, 2006) and further considered important theoretical aspects such as the processes typically involved in solving number series items (e.g., Holzman et al., 1983; Kotovsky & Simon, 1973, Loe et al., 2018; Verguts et al., 2002). Using this video tutorial, we conducted an experiment comparing the number series test scores of the experimental group watching the number seres tutorial prior to the assessment with the number series test scores of 1.) a control group watching an irrelevant video tutorial (figural matrices) and 2.) a control group working on item examples instead, thereby also examining the effects of a video tutorial in comparison with another approach test-takers might use when preparing for an upcoming assessment. Additionally, and to investigate validity aspects, we inspected the correlation coefficients between the number series test scores and the scores of another intelligence test as well as the correlations with grade point average (GPA).

In the fourth study, we shifted our focus towards changes in test-taking behavior after watching a number series video tutorial. Using the data of the third study, we investigated whether the experimental group watching the number series video tutorial and the control group watching the irrelevant tutorial differed in the number of notes that were taken after watching the tutorials. Furthermore, we investigated whether taking these notes represents a generally effective test-taking behavior, and whether such note-taking behavior can be interpreted underlying the corresponding test score increases after watching a number series video tutorial.

In conclusion, we pursued the aim in this dissertation project to investigate test score increases after watching video tutorials by inspecting two different and widely used rule-based tasks highly relevant for assessing intelligence. Besides evidencing test score increases after watching video tutorials as primary goal of this dissertation project, we indicated the replicability of such effects, provided insight into aspects related to validity, compared the approach of watching a video tutorial with another approach test-takers might use when preparing for an assessment, and investigated changes in test-taking behavior assumed to underly the evidenced test score increases after watching a video tutorial.

2. Theoretical Framework

In the following chapter, a brief overview is provided regarding the theoretical framework the dissertation project is based on. First, the construct of intelligence as well as the relevance of intelligence in work and education are introduced. Second, the possibility of increasing intelligence is discussed and differentiated from methods to specifically increase intelligence test scores. Finally, the section is concluded by highlighting the relevance of video tutorials and elaborating on specifically increasing test scores by watching a video tutorial before an assessment.

2.1 Intelligence

Intelligence is widely regarded as one of the most important predictors for educational and vocational accomplishments. Besides introducing the underlying construct of intelligence, the following section describes important aspects of assessing intelligence and elaborates on the relevance of intelligence in the context of work and education.

2.1.1 The Construct of Intelligence

Over the course of the last century, psychometric research has provided a well-founded insight into intelligence as "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" (Gottfredson, 1994, p. 13). These characteristics are empirically reflected in well-documented hierarchical models of intelligence (e.g., Carroll, 1993: analyzing over 450 sets of data including over 130.000 participants), typically revealing a structure with a general intelligence factor *g* at the top (also referred to as *general mental ability*) and more specific ability factors arrayed below (Neisser et al., 1996; see also Rost, 2013). This "*g* – based factor hierarchy is the most widely accepted current view of the structure of abilities" (Neisser et al., 1996, p. 81). Essentially, the key element related to *g* is often represented by problem solving or *reasoning*, while more specific facets (especially facets related to test-taking or the specific test; see Jensen, 1980) are differentiated from intelligence. Such a distinction is typically shared in many models of intelligence, although the exact structure in regard to the dimensionality and facets vary.

One model focusing reasoning while also emphasizing dimensional aspects is the integrative, faceted, and hierarchical *Berlin Intelligence Structure Model* (*BIS*; Jäger, 1982, 1984; Jäger et al. 1997; see Figure 1).

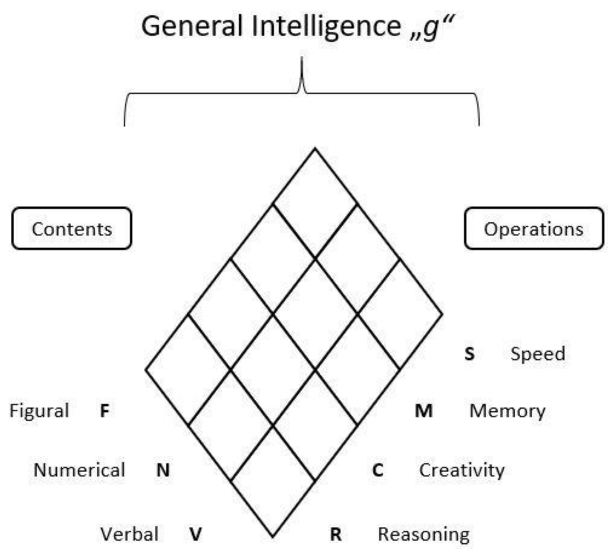

Figure 1: Berlin Intelligence Structure Model (Jäger, 1982, 1984; Jäger et al., 1997).

In this model based on empirically comparing and integrating different models of intelligence structure, general intelligence *g* is assumed at the top level as integral part of all ability components. Below, at the second level, seven higher order abilities are divided into two facets. The facet *contents* distinguishes abilities according to the applied material (F: Figural; N: Numerical; V: Verbal). The facet *operations* distinguishes abilities according to the involved cognitive processes (S: Perceptual Speed; M: Memory; C: Creativity; R: Reasoning). At the third level, intellectual performance is indicated by cross-classifying the content facet with the operations facet, resulting in 12 distinct cells. Each cell designates an intellectual performance according to the abilities involved in the corresponding cross-classification of the two facets. Thus, the BIS-model not only falls in line with the aforementioned understanding about intelligence, but also provides a classification scheme for cognitive tasks (e.g., a figural reasoning task results from combining the cognitive operation reasoning with figural content). As such, the model is also suitable as starting point for more systematic (or even experimental) investigations related to intelligence and intelligence tasks, for example by keeping an operation facet such as reasoning constant and exploring different content facets for this operation. In short, and although a uniform exact definition of intelligence has not been established so far (see Rost, 2013), a general consensus about the construct of intelligence has been achieved – in particular by assuming reasoning as its essence (see Neisser et al., 1996).

Assessing intelligence in the broad sense of *g* typically involves administering a complete intelligence test consisting of all corresponding subtests. As test economy can play a critical role regarding the feasibility of intelligence testing, short (or screening) versions comprising the most relevant subtests are often available as alternative (e.g., the *IST-Screening* [Liepmann et al., 2012] as short version of the *IST 2000R* [Liepmann et al., 2007], comprising three subtests focusing on verbal/numerical/figural reasoning). Depending on evidenced high proximity to *g*, single tasks can also be used as *one-shot* – method for assessing intelligence. As popular task type for such one-shot assessments, figural matrices such as Raven's Progressive Matrices repeatedly show the assumed high correlations with *g*: "When the Progressive Matrices test is factor analyzed with a variety of other tests, it is typically among the two or three tests having the highest *g* loadings" (Jensen, 1998, p. 38.). Thus, they are also considered "marker" tests for *g* (Jensen, 1998, p. 38). Importantly, all test-takers are required to work under the same conditions when intelligence tests are administered. When this premise is violated, meaningful conclusions based on comparing an individual's test score with a criterion, the test scores of other test-takers, or with the norms reported in a test manual are compromised. However, when this premise is considered, intelligence test scores allow for meaningful prediction of many different criteria highly relevant in contexts such as work and education.

2.1.2 The Relevance of Intelligence in Work and Education

Intelligence is considered the "most important trait or construct in all of psychology" (Schmidt, 2009, p. 4). As such, intelligence plays a crucial role for success in both work and education (Kuncel et al., 2004).

In the contexts of work, this role is illustrated by a variety of important aspects: First, general mental ability predicts job performance and training performance (e.g., Hülsheger et al., 2007; Hunter

& Hunter, 1984; Schmidt, 2009; Schmidt & Hunter, 2004). Averaging the unweighted validity coefficients of 12 meta analyses, the validity estimates revealed to be substantially large for predicting both job performance (*r* = .55) and for predicting training performance (*r* = .63; Schmidt & Hunter, 2004). Second, general mental ability predicts attained occupational level (e.g., Jensen, 1998; Schmidt & Hunter, 2004). The associated correlations were shown to reach substantially large magnitudes (up to *r* = .72), and revealed correlations of similar scale (*r* = .70) even when *g* was measured in childhood and linked to the occupational level in adulthood (see Jensen, 1998, p. 293). Third, the validity of general mental ability to predict job performance increases with higher job complexity (e.g., Hunter & Hunter, 1984; Hunter, 1986; Schmidt, 2009). Moreover, while the mean scores within an occupation increase in higher levels of occupation, the range of these mean scores becomes smaller. At the same time, while high scores are found in almost every occupation, the lowest scores rise with the level of the occupation, suggesting that *g* serves a threshold for different occupations (Jensen, 1998; see also Schmidt & Hunter, 2004, p. 164). Together, these results indicate that intelligence becomes increasingly relevant in more demanding positions, and that it is difficult for individuals with low general mental ability test scores to enter higher levels of occupation (Schmidt & Hunter, 2004). Fourth, general mental ability exerts stronger effects on both job performance and occupational level than other traits and variables such as, for example, specific aptitudes, personality traits, and job experience (Schmidt & Hunter, 2004). Regarding the relevance of *g* in the context of occupational level and level of performance within occupations, Jensen (1998, p. 294) concluded: "(…) None of these other factors is so potent as to completely override the threshold aspect of *g* in predicting an individual's probable success in a particular occupation". Such a conclusion regading the relevance of intelligence is further corroborated by Schmidt (2009, p. 14): "Higher intelligence leads to better job performance on all jobs, and the increases in job performance resulting from hiring on GMA have high economic value for organizations". In summary, intelligence plays a central role in the context of work and occupation, and intelligence test scores are thus often used for personnel selection and human resource development.

In the context of education, the role of intelligence is regarded similarly relevant. First, and as essential criterion for scholastic achievement, intelligence predicts school grades (e.g., Jensen, 1998; Gygi et al., 2017; Neisser et al., 1996; Sternberg et al., 2001; Valerius & Sparfeldt, 2015). The coefficients reported are typically substantially large, as evidenced in a recent meta-analysis including 230 samples with more than 100.000 participants (Roth et al., 2015; ρ = .54). In some studies, the coefficients revealed to be of even higher magnitude. For example, in a study investigating a large sample of over 70.000 children (Deary et al., 2007), the latent correlation between performance in cognitive ability tests assessed at age 11 and performance in national school examinations taken at age 16 was reported to be as high as r = .81; correlations of similar magnitude and with comparably large sample sizes were also reported in further studies (e.g., Calvin et al., 2010; see also Kaufman et al., 2012). Second, intelligence is often shown to predict academic achievement better than any other measurable variable (Jensen, 1998). Correspondingly, intelligence is widely considered as the single best predictor for educational success, although other variables should not be neglected (e.g., motivation: Kriegbaum et al., 2018; prior academic achievement: Soares et al., 2015). Third, numerous studies revealed a substantial predictive validity of intelligence for a wide range of different criteria relevant for school and education, including oral reading (r = .62), reading comprehension (r = .68), rank in high school graduating class (r = .62), grades in college (r = .44), and highest level of education attained by age 40 (.50 ≤ r ≤.58; Jensen, 1981, p. 31; see also Rost, 2013, pp. 319–320, for a more extensive overview). Regarding the relevance of *g* in the context of education, Jensen (1998) concluded: “If there is any unquestioned fact in applied psychometrics, it is that IQ tests have a high degree of predictive validity for many educational criteria” (p. 277). To summarize, intelligence also plays a central role in the context of education, and intelligence test scores are hence used for admission and selection purposes in educational contexts such as schools, universities and scholarship programs (e.g., the selection of medical students at the Austrian public universities of Vienna, Innsbruck, Graz, and Linz[3]; selecting candidates for the *Studienstiftung des Deutschen Volkes* as largest and oldest German sponsorship organization for the academically gifted on the basis of an admission test representing an alternate method of entrance[4]).

In conclusion, intelligence is highly relevant in both work and education. Ideally, using corresponding test scores leads to an optimal fit between a person’s intellectual capabilities and the challenges issued by the environment. If these challenges exceed a person’s ability, meaningful consequences can result. In the context of education, for example, it was shown that lower intelligence test scores significantly increased the risk of school dropout (Pagani et al., 2017). In the context of work, it was shown that individuals were more likely to move to a job of lower complexity

[3] See the official website: https://www.medizinstudieren.at/; see also Arendasy et al., 2016
[4] See the official website (also including openly available item examples): https://www.studienstiftung.de/infos-fuer-studierende/bewerbung-und-auswahl/selbstbewerbung/

when a job's complexity level exceeded that individual's general mental ability (Wilk & Sackett, 1996). Thus, predictions should rely on intelligence test scores genuinely reflecting a person's true cognitive ability to avoid overestimating an individual's capabilities. However, intelligence test scores can be increased – potentially facilitating such adverse consequences, as these increases do not necessarily reflect increases related to the underlying trait.

2.2 Increasing Intelligence Test Scores

In the last decades, a plethora of evidence has been reported on increased intelligence test scores. The following section first discourses on the possibility of increasing test scores by raising intelligence before elaborating on approaches specifically aimed at improving test performance.

2.2.1 Increasing Test Scores by Raising Intelligence?

Over the years, lots of debate has been revolving around the question whether and how intelligence can be raised. Accordingly, various influences have been investigated, ranging from attempts specifically aimed at increasing cognitive functioning (e.g., cognitive training) to influences characterized by engaging with a cognitively stimulating, complex environment over longer periods of time (e.g., attending school).

Regarding attempts specifically aimed at increasing cognitive functioning, "the central question is not whether performance on cognitive tests can be improved by training, but rather, whether those benefits transfer to other untrained tasks or lead to any general improvement in the level of cognitive functioning" (Owen et al., 2010, p. 775). Such an understanding is based on the premise that *g* is reflected by all forms of cognitive activity to some extent. Thus, the treatment effect of cognitive training should be reflected in a wide variety of tasks to demonstrate something more than specific score gains (see Jensen, 1998, pp. 333–334). As a result, to consider cognitive training successful, performance increases particularly in distantly related tasks are required: "Any true increase in the level of *g* connotes more than just narrow (or near) transfer of training; it necessarily implies *far* transfer" (Jensen, 1998, p. 334). The aspects of performance increases after cognitive training and transfer to other tasks were addressed in a recently published comprehensive review (Simons et al., 2016). In summary, although an extensive body of research indeed evidenced the expected performance increases in the trained task, the authors conclude that there is less evidence for also enhancing performance in *closely related* tasks, and even less evidence for also enhancing performance in *distantly related* tasks or improving everyday cognitive performance. Such a conclusion is further corroborated by a recent meta-analytic review concluding no performance improvements on measures of intelligence or other measures of far transfer when training working memory (Melby-Lervåg et al., 2016). In short, as evidence often fails to support the required transfer effects in untrained tasks (see also Owen et al., 2010), the benefits from cognitive training seem to be task-specific and most likely do not transfer to increases in intelligence.

Regarding influences based on engaging with a stimulating and complex environment over longer periods of time, attending school was reviewed to be substantially relevant (Ceci, 1991). As such, it was argued that intelligence is not only an important predictor for academic achievement, but that schooling also nurtures the corresponding cognitive processes. Accordingly, the evidence on schooling indeed points towards a causal factor from schooling (i.e., duration of attending school) on intelligence (e.g., Cahan & Cohen, 1989; Ceci, 1991; Stelzl et al., 1995). For example, Stelzl et al. (1995) tested 578 children as closely as possible to their 10th birthday, thereby keeping chronological age constant and school age varying with month of the birthday. The effect of 1 year of schooling on several measures of intelligence revealed to be substantial (*CFT 20*: 7 IQ points; *HAWIK Vocabulary Test:* 6 IQ points; *KFT* [progressively estimated[5]]: 12 IQ points). These results were in line with prior assumptions pointing towards an estimated loss of about five IQ points per year for delayed onset of school (Ceci et al., 1991), leading to the conclusion that “school fosters the development of general intelligence as it keeps children cognitive challenged” (Stelzl et al., 1995, p. 293). This notion of engaging in challenging and stimulating environments as necessary elements to raise intelligence is further illustrated by early educational intervention programs. Although such programs are typically multifaceted (and embedded in different approaches such as, for example, home visits or parent training) and thus the particular causal mechanisms are difficult to be discerned, the common denominator lies in enriching the environment in order to increase a child’s intelligence (besides nurturing self-regulation and social skills; see Protzko et al., 2013). Correspondingly, meta-analytic evidence (Protzko et al., 2013) supports positive effects on intelligence for infants of lower socio-

[5] As some items were omitted for a part of the sample and the estimates were calculated based on considering these items solved, the resulting effects regarding the KFT are expected to be slightly overestimated (see Stelzl et al., 1995, p. 289).

economic status enrolled in such a program (corresponding with an average IQ gain of 4 points, with greater gains in more complex interventions). Furthermore, the evidence also supports positive effects on a young child's intelligence when parents increase cognitive stimulation by engaging in interactive reading (e.g., encouraging to ask open-ended questions; corresponding with an average IQ gain of 6 points), as well as sending their child to preschool (corresponding with an average IQ gain of 4 points). In short, the results emphasize "the likelihood that environmental complexity is the prime mechanism underlying gains in IQ; however, which specific aspects of that complexity are most effective or beneficial remain unknown" (Protzko et al., 2013, p. 29).

Importantly, two aspects should be kept in mind when interpreting increases in intelligence test scores. First, it can be difficult to interpret whether the increases indicated by rising intelligence test scores also represent a genuine improvement in the underlying trait. For example, intelligence scores consistently have been reported rising over time for decades (Flynn, 1987; often referred to as the *Flynn effect*). Although some authors (e.g., Lynn, 2013) surmise the nature of the effect to indicate real increases in cognitive ability, other authors (e.g., te Nijenhuis et al., 2014) posit the alternate explanation that the effect has little to nothing to do with *g*. Re-examining all published studies on how strongly the Flynn effect truly correlates with the *g*-factor, a recent meta-analysis (te Nijenhuis & van der Flier, 2013) points toward the direction that score gains over time could indeed be "hollow" in respect to *g* (i.e. these gains relate to the non-*g*-variance; see also Flynn et al., 2014; Woodley et al., 2014). Thus, and in short, increased test scores do not necessarily reflect genuine improvements concerning the underlying cognitive processes. Generally, gains in intelligence test scores indeed often fail to show such underlying improvements, pointing towards rather test-specific aspects instead (see Jensen, 1998). Second, it is important to note that even when successful interventions demonstrate an increase in test scores, the gains diminish (or even fade out) over time. This *fadeout effect* was substantiated in a recent meta-analysis (Protzko, 2015) and can be observed "for almost every intervention when researchers followed their participants" (Protzko, 2015, p. 202). As such, the fadeout effect can be found even for well-funded and intense early educational intervention programs (e.g., changes in intelligence and academic achievement fading out three or four years after the end of the preschool intervention in several *Headstart*-programs; Brody, 1997). Finally, a recent study showed that an early intervention can indeed promote increases attributed to the *g* factor (rather than attributed to test-specific elements, as mentioned before). However, the effect was shown to fade out nonetheless – implicating that even when *g* is raised successfully, the gains will most likely not be permanent (Protzko, 2016).

In conclusion, increasing intelligence test scores can be achieved, but raising dispositional intelligence tends to be rather difficult and might at least require the long-term and active exposure to enriched and stimulating environments that most targeted programs or (short-term) approaches typically fail to provide. As a result, increased test scores typically reflect short-term increases that diminish after a while or test-specific improvements. Regardless, test-takers might specifically seek out opportunities to increase their test scores for an upcoming assessment in order to reach a desired position or for entering a desired school or university – potentially focusing on approaches specifically aimed at increasing test scores such as retaking a test, practice, or coaching.

2.2.2 Practice and Coaching

Different methods to increase test scores in an intelligence test encompass simple approaches such as practicing a test as well as more complex approaches such as relying on coaching.

Concerning simple approaches, a basic approach "involves taking the same or similar tests at least two times at various intervals, without any implication of special instructions or specific coaching in test taking" (Jensen, 1980, p. 590). Often, this is also referred to as *practice effect* (e.g., Hausknecht et al., 2007; Jensen, 1980), *retest effect* (e.g., Lievens et al., 2005; Lievens et al., 2007; Scharfen et al., 2018), or *retest bias* (e.g., Villado et al., 2016). Correspondingly, "practicing a test" often refers to the pheonomenon of simply retaking a test, although sometimes the term is used more heterogeneously. Re-examining 174 samples in a recent meta-analysis (Scharfen et al., 2018), the average effect size between the first and second test administration for such an approach revealed to be *SMCR* = 0.37 (corrected for bias *S*tandardized *M*ean *C*hange with *R*aw score standardization; interpretation similar to Cohen's *d*) for identical test forms and SMCR = 0.23 for alternate test forms. These effects also fall in line with previous meta-analytic evidence regarding practice effects (e.g., [identical/alternate forms]: *d* = 0.40/0.22, Hausknecht et al., 2007; *d* = 0.42/0.23, Kulik, Kulik, & Bangert, 1984). Thus, intelligence test scores can effectively be increased simply by retaking the same (or a similar) test.

Concerning more complex approaches, the related methods are often summarized as *coaching* and cover a "wide diversity of training interventions varying in duration, content, procedures, and objectives" (Anastasi, 1981, p. 1086). As implicated by the description, coaching is less unambiguous than the previously elaborated practice. In consequence, "coaching" is often referred to as "training" – even when *transfer* as requirement for successful training is neglected (see chapter 2.2.1). Jensen (1980) emphasized coaching as teaching test-taking strategies, including "how to analyze test questions and problems, instructions and demonstrations in working through typical test problems, distributing one's time most efficiently, and doing many typical practice problems with a tutor or manual providing immediate informative feedback" (p. 591). The effect of coaching revealed to be slightly larger than for retaking a test in meta-analyses, averaging an effect size of d = 0.43 (Kulik, Bangert-Drowns & Kulik, 1984) – with an even higher effect size of d = 0.64 when coaching was accompanied by practice (Hausknecht et al., 2007). Thus, similarly to retaking a test, intelligence test scores can effectively be increased by coaching.

Referring to the previous chapter, the evidence on score gains for practice and coaching aligns with the previously described notion of test scores typically not going in hand with an increase in dispositional intelligence. Jensen (1998, p. 307) summarized this for test score increases: "The practice effect from taking a given *g*-loaded test, as indicated by the amount of test-retest gain in score, appears to be unrelated to *g*. Test-retest gains probably reflect only the source of variance known as the test's specificity". This conclusion is supported by a meta-analysis including over N = 26.000 participants from 64 test-retest studies using IQ batteries (te Nijenhuis et al., 2007), revealing no *g* saturation for such score gains (further supported by more recent studies; e.g., Estrada et al., 2015). A similar conclusion can be drawn for coaching, as studies showed reduced *g*-loadings after different types of coaching attempts (e.g., reading a book on test orientation covering characteristics of different intelligence tests as well as strategies for solving items, or reading the same book followed by a training program of three hours; te Nijenhuis et al., 2001). In conclusion, especially by focusing on improving test performance through practice or coaching, intelligence test scores can be easily increased – although these test score increases typically reflect the aforementioned short-term or test specific improvements not related to genuine enhancements of intelligence (see Haier, 2014, for a short review). Nevertheless, test-takers might specifically pursue such approaches in order to score as high as possible in an upcoming assessment. However, relying on these rather traditional methods can require heavy investment of resources (e.g., time, money). Potentially, modern approaches can overcome these limitations while demonstrating effectivity and efficiency at the same time – for example, watching a video tutorial prior to the assessment.

2.2.3 Increasing Test Scores by Watching a Video Tutorial

Video tutorials have become an increasingly popular instrument to illustrate a wide variety of subjects (Wolf, 2015). For test-takers preparing for an upcoming assessment, watching a video tutorial about intelligence tests might represent a particularly attractive approach to increase their test scores for various reasons.

First, different elements relevant for practice and coaching can be combined in a video tutorial (e.g., item examples, strategies, task-related explanations, generally useful information). Furthermore, these elements can be further exemplified by using didactically helpful elements (e.g., visual illustrations such as arcs and arrows). Thus, using such an approach may well be (at least) similarly effective to the more traditional approaches for increasing test scores. Importantly, it should further be considered that information in video tutorials is typically condensed into a compact and engaging time-frame (e.g., 10–20 minutes; see, for example, DeVaney, 2009; He et al., 2012). Thereby, the required time investment is much lower than when compared to working through a whole test as practice or seeking more differentiated coaching programs. Regarding the design of video tutorials, focusing on particularly relevant aspects might be specifically relevant. For example, focusing on teaching the rules in reasoning tasks might provide test-takers with the essential tools to answer the test-items correctly, potentially resulting in particularly large increases. First empirical insight into such an assumption was gained in a study consisting of four experiments by Loesche et al. (2015) investigating Raven's Advanced Progressive Matrices (Raven et al., 1998). In each experiment, the participants in the experimental group were presented a video (6:41 min.) in which five rules (based on the taxonomy by Carpenter et al., 1990) were explained and illustrated with four items. Afterwards, these four items were presented to practice on for five minutes. The participants in the control group of each experiment were presented a video with instructions from the test manual (3:00 min.) and worked on 12 figural matrices afterwards (12 min.). Next, participants in both groups worked on the 26 (of 36 in total) Raven's Advanced Progressive Matrices items (Raven et al., 1998) that followed the five rules explained in the video of the experimental group. In three of the four experiments, participants in the experimental group solved significantly more figural matrices items than participants in the control group, revealing medium to large effects (Experiment 1: N = 626 5th to 8th –graders: d = 0.52; Experiment 3: N = 382 5th to 7th –graders: d = 0.56; Experiment 4: N = 47 undergraduate students: d = 0.81). Although some limitations have to be considered (e.g., partially inconsistent results [i.e., no statistically significant differences in Experiment 2: N = 353 5th to 8th –

graders: *d* = 0.14]; differences between the experimental groups and the control groups related to practice, time, and number of shown items; prior familiarization with figural matrices also in the control group; exclusion of items from the later administered published test due to insufficient alignment with the rules explained in the video), the study corroborates the previously mentioned assumption that including particularly relevant aspects in a video tutorial might lead to substantial score increases. In short, watching a video tutorial can provide test-takers with an elaboration of important aspects shown to be effective as part of practice or coaching while also considering efficiency.

Second, video tutorials are widely available. For example, using keywords such as *"Intelligence Test Video Tutorial"* or *"IQ Test Video Tutorial"* (or even keywords related to specific popular intelligence test task types, e.g. *"IQ Figural Matrices Video Tutorial"* or *"IQ Number Series Video Tutorial")* in online search engines yields an abundant amount of results. Additionally, saving the video tutorial (or the link for access) allows interested persons to re-watch passages or even the whole tutorial at one's own pace, thereby providing additional opportunities to internalize the relevant content. Moreover, tutorials can be watched timely before an upcoming assessment. Thus, and in contrast to retaking a test or relying on coaching programs, the information can be accessed or recapitulated when it's most critical (i.e., directly before the test) rather than much earlier before the assessment. As repeated exposure to information and timing were shown to be associated with higher increases related to retaking a test (e.g., larger effects on test scores for shorter test-retest intervals as well as larger effects on test scores for a higher number of test administrations; Scharfen et al., 2018), using video tutorials incorporates opportunities beyond what is offered by traditional approaches.

Third, video tutorials are easily accessible. In contrast to decades ago when information about intelligence tests was limited, accessing information nowadays requires as little as a smartphone with a working internet connection. Additionally, depending on the search engine, the results on video tutorials can often be filtered according to desired specific features (e.g., duration of the video, country), thereby facilitating the accessibility of the desired content by only including the individually most relevant attributes. Furthermore, and although there are online video tutorials locked behind paywalls (e.g., by professional test preparation services), searching for online tutorials also reveals plentiful free material. Thus, there are often little to no barriers involved, especially when compared to the logistics and resources necessary when relying on traditional approaches such as retaking a test or coaching.

In summary, watching video tutorials before an upcoming assessment might be an attractive approach for test-takers by overcoming traditional barriers while emphasizing aspects of effectivity and efficiency. Accordingly, empirically investigating the effects of watching such tutorials is highly relevant for intelligence testing. This relevance is especially emphasized when popular and widely used tasks are focused in such tutorials, as test-takers are more likely to encounter these tasks in an upcoming assessment. Furthermore, this relevance is further accentuated by particularly large increases (signifying a greater disparity between elevated scores and dispositional intellectual ability), as more severe consequences are to be expected when such elevated test scores are used in admission and selection decisions. Moreover, and keeping the aforementioned aspects of wide availability and easy accessibility in mind, increasing test scores by watching a video tutorial before an assessment most likely impacts a plethora of testing situations. Nevertheless, and besides focusing on test score increases as most essential research question, further questions should be considered such as whether validity aspects are affected when test-takers rely on watching video tutorials, or if using such an approach also stimulates test-takers to adapt their test-taking behavior. In this dissertation project, we provide answers to these questions by experimentally investigating the effects of video tutorials on intelligence test scores on the basis of two popular rule-based tasks. Specifically, we focused on evidencing the effect on test scores as primary goal, while providing insight into validity aspects and effects on test-taking behavior as secondary goals.

2.3 Research Aims of the Dissertation Project

Building on the theoretical framework and the empirical evidence presented in the previous chapters, we conducted four studies published in three articles for this dissertation project to examine the effects of video tutorials on intelligence test scores. In detail, the overarching research goals were threefold:

As primary goal, we targeted investigating the test score increases in commonly administered intelligence task types after watching corresponding video tutorials. As secondary research goal, we aimed at gathering evidence on whether watching such video tutorials affects validity aspects of the subsequential test. The third and final research goal aim centered on examining whether watching a video tutorial effects test-takers to adapt their test-taking behavior, and if such a behavior can be assumed underlying the related test score increases.

3. Empirical Studies

To carry out the aim of investigating the effects of video tutorials on intelligence test scores, we conducted four empirical studies published in three peer-reviewed articles. In the following section, the four studies are outlined by summarizing the respective theoretical background, hypotheses, methods, results, and discussion. Supplementary analyses further clarifying relevant aspects but not integrated in the corresponding publications are provided in chapter 3.4.5.

In Study 1 (Schneider et al., 2020 – Experiment 1), we experimentally investigated the effect of video tutorials by focusing on figural matrices as popular and widely used figural reasoning task type. Accordingly, we compared the figural matrices test scores of an experimental group watching a figural matrices video tutorial with the test scores of a control group watching a video about nutrition. Furthermore, we investigated validity aspects by examining the correlations between the figural matrices test scores and the scores in an intelligence test in both groups.

In Study 2 (Schneider et al., 2020 – Experiment 2), we conducted another experiment to replicate the effects obtained in Study 1, thereby strengthening the results. Additionally, we streamlined specific aspects related to the methods used in Study 1 and increased the sample size.

Study 3 (Schneider & Sparfeldt, 2021a) was conducted to investigate the effects of video tutorials by focusing on number series as another popular (but numerical) reasoning task type. Besides comparing the number series test scores of an experimental group watching a number series video tutorial with the number series test scores of A.) a control group watching a figural matrices tutorial and B.) another control group inspecting item examples instead, we also inspected the correlation coefficients between the number series test scores and the scores in an intelligence test as well as between the number series test scores and GPA.

Shifting the focus towards the effects on test-taking behavior, we conducted Study 4 to examine whether watching a number series video tutorial also stimulates test-takers to take more notes, and whether using such notes can be assumed underlying the number series test score increases demonstrated in Study 3. Accordingly, we compared the number of items with notes of participants in the experimental group watching the number series video tutorial with the number of items with notes of participants in the control group watching the figural matrices video tutorial.

Additionally, we examined the correlations between the number of items with notes and the number series test scores in both groups and analyzed whether these notes mediated the effect of the number series video tutorial on number series test scores.

3.1 Study 1 – Schneider et al. (2020): Experiment 1

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

In Study 1, we aimed to investigate the effects of a video tutorial on the basis of figural matrices as a commonly used task to assess (figural) intelligence.

3.1.1 Theoretical Background and Hypotheses

As outlined in the previous chapters, scores in intelligence tests are essential predictors and correlates for outcomes such as job performance and educational success (e.g., Deary et al., 2007; Hülsheger et al, 2007; Jensen, 1998; Roth et al., 2015; Schmidt & Hunter, 2004) and are thus frequently used in selection and admission decisions. Although scores can be evelated by, for instance, practice and/or coaching (Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984; Scharfen et al., 2018), they typically do not reflect an enhanced dispoitional intelligence (e.g., Estrada et al., 2015; Haier, 2014; Hayes et al., 2015; see also Jensen, 1998, te Nijenhuis et al., 2007). As a result, using increased test scores (rather than genuine scores) might overestimate a test-taker’s cognitive ability, leading to adverse consequences that are further exacerbated by the magnitude of the corresponding effects (e.g., a test-taker not performing as expected or being overburdened by the requirements). However, test-takers might specifically seek opportunities to increase their test scores in order to score as high as possible in an upcoming assessment. Watching a short video (as widely realized in online video tutorials) portraying important elements for popular tasks such as figural matrices might serve as particularly effective and efficient approach to achieve these higher test scores. However, little is known about the effects of watching such a video tutorial.

Figural Matrices represent popular rule-based tasks to assess figural reasoning and are hence often essential parts of broader intelligence test batteries (e.g., WAIS-IV; Wechsler, 2008). Furthermore, they show substantially high saturation in *g* (e.g., Carpenter, Just, & Shell, 1990; Carroll, 1993; Jensen, 1998; Johnson et al., 2004). Typically, in figural matrices items, matrices of fields containing symbols are presented. These symbols are connected according to one ore more rules. In order to solve an item correctly, test-takers are required to identify the rules underlying each item. Thus, focusing on teaching the relevant rules in a short video tutorial might provide test-takers with the vital information required to reach high test scores in a later administered figural matrices task. A first insight into the effects of knowing the rules in figural matrices on test scores was provided recently (Loesche et al., 2015). However, effects were shown to be inconsistent across four studies. Additionally, the control group differed from the experimental group regarding factors such as time, practice, and number of shown items. In summary, in Study 1, we wanted to investigate in an experimental design how watching a figural matrices video tutorial focusing on teaching the underlying rules in figural matrices influences 1.) mean values and 2.) correlation coefficients with another intelligence test.

Hypothesis 1. We expected that watching a figural matrices video tutorial focusing on teaching the rules leads to higher figural matrices mean scores than watching an irrelevant video.

Hypothesis 2a. We expected substantial correlation coefficients between the figural matrices scores and the scores in another intelligence test in the experimental group.

Hypothesis 2b. We expected substantial correlation coefficients between the figural matrices scores and the scores in another intelligence test in the control group.

Hypothesis 2c. We assumed comparable correlation coefficients in the experimental and the control group.

3.1.2 Methods

Participants and Procedure. The sample in Study 1 comprised *N* = 112 psychology students recruited in a first-semester course about personality theories at a medium sized German university (29 male, 83 female; mean age [based on 109 participants]: *M* = 21.09 years; *SD* = 2.87). In a first session at the beginning of the semester (t1), participants completed an intelligence test administered by the same experimenter (30 min.). In a second session (about 6–8 weeks later; t2), the students were randomly assigned to either the experimental group (n = 56) or the control group (n = 56). After working on tasks irrelevant for this investigation on individual laptops, the participants were presented either a figural matrices video tutorial (experimental group; representing test-takers with rule knowledge) or a control video (control group; representing test-takers without rule knowledge). Afterwards, all participants worked on figural matrices tasks. Up to the second session, neither "intelligence" nor "intelligence assessment" were part of the curriculum.

Video Tutorials. The figural matrices video tutorial in the experimental group focused on explaining in text form as well as orally the six rules relevant for the subsequential figural matrices tasks. The explanation of each rule was illustrated by a structured step-by-step example how an item using the specific underlying rule can be solved (see Figure 2 for a screenshot; for a more detailed depiction exemplified using the rule "addition", see Appendix B). After the six rules had been explained individually, an example of an item containing two rules at once was presented. Finally, in the last part of the video, a synopsis was provided recapitulating the most vital aspects. Item examples used in the video were specifically constructed for the video and contained modified symbols not part of the later administered figural matrices test. The video tutorial had a duration of 13:19 minutes and was validated in a pilot investigation using N = 19 participants, indicating the explanations were thorough and helpful to solve related items. The control video used in the control group was similar regarding main aspects of the figural matrices video tutorial used in the experimental group (e.g., duration of 12:22 minutes, visual and auditory information, pictorial information and text) but different in content. Specifically, the basics of nutrition were elaborated and supplemented by information on how using a particular website can help reaching dieting goals.

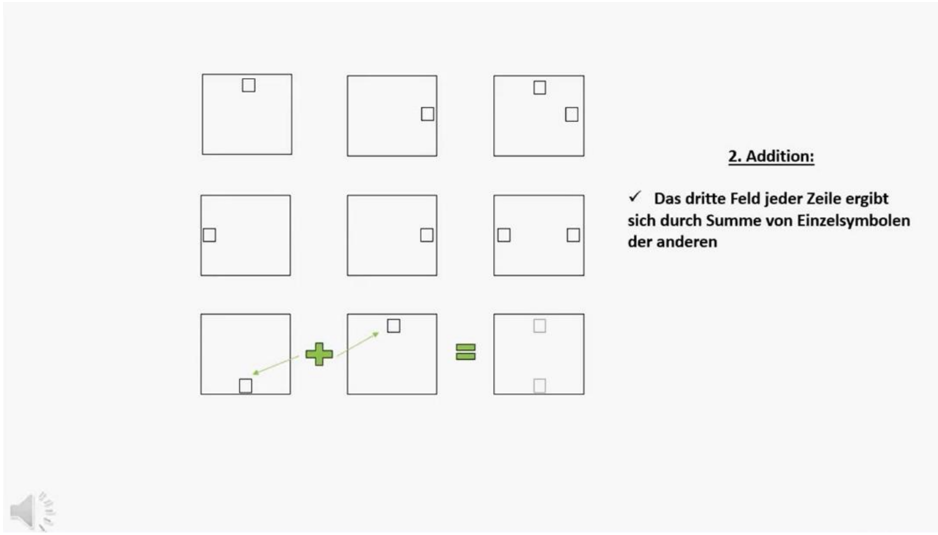

Figure 2: Original Screenshot of the Figural Matrices Video Tutorial used in Study 1.

Instruments. Figural matrices performance was assessed using two variants of the *DESIGMA-Advanced* (Becker et al., 2016; Becker et al., 2014; Becker & Spinath, 2014). One variant used a computerized *construction* response format, requiring the test-taker to construct the solution for an item by selecting individual elements out of an array of all potentially relevant elements. The other variant used a computerized *distraction* response format and required the test-taker to select the correct solution out of nine choices for the solution. In total, the DESIGMA-Advanced contains 38 items (each with a time limit of 90 seconds), and for each item stem the two previously described formats were available. To consider both response formats as well as all 38 item stems for each participant, we created two parallel item-sets (each containing 19 items) based on item parameters reported in the test manual as well as a comparable number and type of rules, yielding four test versions: (1) item-set A in construction followed by item-set B in distraction, (2) item-set B in construction followed by item-set A in distraction, (3) item-set A in distraction followed by item-set B in construction, and (4) item-set B in distraction followed by item-set A in construction. In short, test versions were balanced across all participants, and each participant worked on all 38 item stems by first working on one response format and item set and afterwards working on the remaining item set and response format not yet administered. To exclude practice of specific rules, the instructions were modified and contained an item example portraying the same element in all fields of the 3×3 – matrix except the solution field, besides providing basic information (e.g., how to operate the interface).

Intelligence was assessed using four subtests of the *Leistungsprüfsystem* 2 (*LPS-2*; Kreuzpointner et al., 2013) that is aimed at measuring general intelligence *g* based on the three-stratum intelligence model (Carroll, 1993, 2005). The subtests we used were *general knowledge* (crystallized intelligence; participants had to mark the incorrect letter in words; 60 items, 3 minutes), *numerical sequences* (fluid intelligence; participants had to mark the incorrect numbers in number sequences; 30 items, 5 minutes), *mental rotation* (broad visual perception; participants had to mark the mirror-inverted element in numbers or letters; 40 items, 2 minutes), and *addition* (broad cognitive speediness; participants had to add numbers and mark the single digit of the numbers' sum; 80 items, 6 minutes). As indicator for general intelligence, we used the sum score of correctly solved items.

Analyses. As preliminary analyses to examine whether the randomizing procedure was successful, we calculated independent *t*-tests using group as independent variable and the LPS-2 sum score as well as the subtest scores as dependent variables. Regarding our analyses related to Hypothesis 1, we computed a 2 × 2 × 2 MANOVA with two dependent variables (1: sum score of correctly solved construction figural matrices [item-set A: $\alpha = .95$; item-set B: $\alpha = .92$]; 2: sum score of correctly solved distraction figural matrices [item-set A: $\alpha = .87$; item-set B: $\alpha = .90$]) and three independent variables (item set; format order; group). Specifically, to test the hypothesis whether watching a figural matrices video tutorial increased mean values, we inspected the main effect group as well as the corresponding interaction terms. In case of statistically significant results, we calculated univariate ANOVAs for both response formats ($p < .05$) and inspected the effect sizes (d, η^2). Regarding our analyses related to Hypotheses 2a – 2c, we examined the correlations between figural matrices scores and intelligence scores in (2a) the experimental group and (2b) the control group, and (2c) compared these coefficients.

3.1.3 Results

Preliminary Analyses. Regarding the preliminary analyses, the results indicated no statistically significant differences between the experimental group and the control group (LPS-2 sum score: $t(110) = 1.10$, $p = .27$; general knowledge: $t(110) = 0.79$, $p = .43$; numerical sequences: $t(110) = 0.46$, $p = .65$; mental rotation: $t(110) = 1.46$, $p = .15$; addition: $t(110) = 0.28$; $p = .78$), suggesting no significant a priori differences in intelligence (see Table 1 for descriptive statistics).

Table 1: Descriptive Statistics for the Intelligence Test used in Study 1.

	Entire Sample			Experimental Group			Control Group		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
LPS-2 Sum score	92.58	18.76	112	90.63	17.22	56	94.54	20.16	56
LPS-2 General knowledge	35.25	8.79	112	34.59	7.86	56	35.91	9.66	56
LPS-2 Numerical sequences	20.74	3.50	112	20.59	3.65	56	20.89	3.37	56
LPS-2 Mental rotation	21.96	7.17	112	20.98	6.69	56	22.95	7.54	56
LPS-2 Addition	14.63	5.97	112	14.46	6.00	56	14.79	5.98	56

Mean Differences. Descriptive statistics for the figural matrices tasks are depicted in Table 2. The results for the 2 × 2 × 2 MANOVA as well as for the corresponding ANOVAS in each response format are depicted in Table 3. Regarding the first hypothesis related to the effect of the video tutorial in Study 1, the MANOVA revealed the expected statistically significant main effect group (p < .01, η^2 = .31). Additionally, the MANOVA showed a statistically significant interaction effect format order × group (p < .01, η^2 = .09), but no statistically significant interaction effect Item-set × group (p = .39) or three-way interaction (p = .36). In univariate analyses, the corresponding effect format order × group showed to be statistically significant neither for the construction format (p = .06), nor the distraction format (.92). However, the figural matrices score differences between the experimental group and the control group still revealed to be statistically significant for the construction format (p < .01, η^2 = .30 corresponding with d = 1.31) as well as for the distraction format (p < .01, η^2 = .26 corresponding with d = 1.19), supporting the results of the MANOVA and thus hypothesis 1.

Table 2: Descriptive Statistics for the Figural Matrices Tasks used in Study 1.

	Construction Figural Matrices			Distraction Figural Matrices		
	M	*SD*	*n*	*M*	*SD*	*n*
Experimental Group						
Item set A						
Construction matrices first	13.29	4.66	14	14.50	4.35	14
Distraction matrices first	14.21	4.61	14	13.71	4.05	14
Item set B						
Construction matrices first	13.36	3.99	14	14.57	4.26	14
Distraction matrices first	13.50	4.31	14	13.86	3.72	14
Control Group						
Item set A						
Construction matrices first	4.71	4.78	14	8.29	4.29	14
Distraction matrices first	9.93	6.96	14	8.50	4.11	14
Item set B						
Construction matrices first	5.57	4.40	14	10.43	6.12	14
Distraction matrices first	8.64	5.77	14	9.07	4.81	14

Table 3: Results for the Multivariate and Univariate Analyses in Study 1.

	MANOVA				*ANOVA (Construction format)*				*ANOVA (Distraction format)*			
	df	*F*	η^2	*p*	*df*	*F*	η^2	*p*	*df*	*F*	η^2	*p*
Item-set	103	0.66	.01	.52	104	0.08	.00	.78	104	0.74	.01	.39
Format order	103	15.58	.23	<.01	104	6.09	.06	.02	104	0.60	.01	.44
Group	103	22.63	.31	<.01	104	45.20	.30	<.01	104	35.57	.26	.00
Item-set × format order	103	2.14	.04	.12	104	0.60	.01	.44	104	0.19	.00	.66
Item-set × group	103	0.96	.02	.39	104	.00	.00	.96	104	0.54	.01	.47
Format order × group	103	5.22	.09	<.01	104	3.62	.03	.06	104	0.01	.00	.92
Item-set × format order × group	103	1.02	.02	.36	104	0.13	.00	.72	104	0.23	.00	.63

Correlations with the Intelligence Test. Regarding the correlations between the figural matrices test scores and the LPS-2 sum scores related to Hypothesis 2a–2c, our analyses revealed substantially high correlations in the experimental group (construction: r = .53, p < .01; distraction: r = .58, p < .01; supporting Hypothesis 2a) as well as in the control group (construction: r = .56, p < .01; distraction: r = .53; supporting Hypothesis 2b). These correlation coefficients did not differ statistically significantly between the groups – neither regarding the construction format (p = .82), nor regarding the distraction format (p = .72), supporting Hypothesis 2c. When inspecting the correlations between the figural matrices scores and the subtest scores, no statistically significant differences were found in any comparison, as well (.20 ≤ p ≤ .81). For a full correlation matrix of all variables, see Table 4.

Table 4: Correlation Matrix for all Variables in Study 1.

	Entire Sample (*N* = 112)						**Experimental Group (*n* = 56)**						**Control Group (*n* = 56)**					
	2	**3**	**4**	**5**	**6**	**7**	**2**	**3**	**4**	**5**	**6**	**7**	**2**	**3**	**4**	**5**	**6**	**7**
1	.83*	.40*	.35*	.33*	.17	.36*	.82*	.53*	.42*	.50*	.19	.46*	.73*	.56*	.49*	.37*	.35*	.45*
2		.42*	.30*	.41*	.23*	.36*		.58*	.42*	.60*	.22	.51*		.53*	.38*	.42*	.44*	.37*
3			.80*	.68*	.72*	.71*			.75*	.72*	.67*	.70*			.83*	.65*	.75*	.72*
4				.39*	.36*	.37*				.40*	.24	.33*				.39*	.44*	.41*
5					.36*	.53*					.39*	.50*					.34*	.57*
6						.31*						.26						.36*
7																		

Note. * = p <.05. 1 = Construction Figural Matrices. 2 = Distraction Figural Matrices. 3 = LPS-2 Sum Score. 4 = Subtest General Knowledge. 5 = Subtest Numerical Sequences. 6 = Subtest Mental Rotation. 7 = Subtest Addition.

3.1.4 Discussion

The main findings of Study 1 were threefold. First, the results showed statistically significant mean differences between the figural matrices test scores in the experimental group watching the figural matrices video tutorial and the control group watching the control video about nutrition. The effects revealed to be substantially large (construction: d = 1.31; distraction: d = 1.19), going considerably beyond the magnitude of effects typically found regarding practice and coaching effects in meta-analyses (Hausknecht et al., 2007; Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984; Scharfen et al., 2018) as well as beyond the effects reported for providing rule knowledge (Loesche et al., 2015). Second, the correlation coefficients were substantially large in the experimental group as well as in the control group (.53 ≤ r ≤ .58) and (at least) similar in magnitude to the correlation coefficients between figural matrices test scores and intelligence test scores reported in prior studies ($r_{DESIGMA\text{-}Advanced/LPS}$ = .39, Becker et al., 2014; $r_{BOMAT/LPS\text{-}2}$ = .45, Kreuzpointner et al., 2013). Thus, the results suggested that a relevant validity aspect is still retained when watching such a figural matrices video tutorial. Third, when comparing the correlation coefficients between the experimental group and the control group, no statisticially significant differences were revealed. Hence, the results indicate that such an essential validity aspect is retained regardless whether test-takers watch a video tutorial providing rule knowledge before an assessment or not. Noticeably, however, the correlation coefficients between the figural matrices test scores and the intelligence sum scores were numerically slightly lower when knowledgeable and un-knowledgeable test-takers are not separated as basis for the correlations (as indicated by using the entire sample as basis for the correlations instead of separate groups: r = .40 /r = .42, as shown in Table 4). Ideally, in the context of a real-life assessment, test-takers are hence best kept homogeneous regarding their rule knowledge. In summary, Study 1 showed that watching a video tutorial on the basis of figural matrices increases figural matrices test scores considerably, and that using the related test scores still retains essential validity aspects. However, one has to keep in mind that the sample size was rather low to detect statistically differences regarding the correlations, and that aspects not directly related to video tutorials were also considered (i.e. format order, item set). Furthermore, we used psychology students for the investigation, resulting in a rather homogeneous sample related to cognitive ability. Besides investigating other tasks as basis for video tutorials, further research to replicate the effects obtained in Study 1 can strengthen the conclusions drawn from watching a figural matrices video tutorial.

3.2 Study 2 – Schneider et al. (2020): Experiment 2

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

Study 2 was conducted to replicate the effects obtained in Study 1 and thus strengthen the results.

3.2.1 Theoretical Background and Hypotheses

On the basis of figural matrices and by focusing on teaching the underlying rules as essential element, Study 1 evidenced the effect of watching a video tutorial on intelligence test scores by showing substantially large mean difference between an experimental group watching such a video tutorial when compared to a standard control condition. Furthermore, Study 1 showed that validity aspects in the sense of the correlations with another intelligence test remained comparable for test-takers watching the video tutorial and for test-takers in a control condition.

Nevertheless, there were several aspects worth re-examining. For example, the sample used in Study 1 consisted of psychology students and was therefore rather homogeneous. Furthermore, additional influences such as response format or different item sets were considered that were not directly related to investigating effects of video tutorials. Additionally, a bigger sample size can further strengthen the results. Thus, Study 2 was conducted to replicate the main findings and address these issues by removing additional influences not directly related to video tutorials, using a different and more heterogeneous sample, as well as substantially increasing the sample size for the investigation. Essentially, the related Hypotheses were identical to the Hypotheses evaluated in Study 1:

Hypothesis 1. We expected that watching a figural matrices video tutorial focusing on teaching the rules leads to higher figural matrices mean scores than watching an irrelevant video.

Hypothesis 2a. We expected substantial correlation coefficients between the figural matrices scores and the scores in another intelligence test in the experimental group.

Hypothesis 2b. We expected substantial correlation coefficients between the figural matrices scores and the scores in another intelligence test in the control group.

Hypothesis 2c. We assumed comparable correlation coefficients in the experimental and the control group.

3.2.2 Methods

Participants and Procedure. The sample used in Study 2 consisted of N = 229 teacher-education students recruited in the lectures “educational assessment” and “basics of educational sciences” of a medium sized German university (mean age: M = 21.98, SD = 4.49; gender based on on 228 participants: 29% male, 71% female). Similar to Study 1, neither intelligence nor intelligence assessment was part of the curriculum before the study commenced. All data were collected in one testing session. Regarding the procedure, the participants were first randomly assigned to either the experimental group (n = 114) or the control group (n = 115). Then, each group was accompanied to a different lecture hall by a corresponding experimenter. Afterwards, the study continued according to a standardized protocol. First, a questionnaire assessing demographics was administered. Next, the participants worked on three of the LPS-2 subtests used in Study 1. Afterwards, the videos were presented on a projector screen. Finally, after the respective video had concluded, the figural matrices items were presented.

Video Tutorials. For Study 2, we used the identical videos we used in Study 1 (i.e., EG: figural matrices video tutorial about six rules underlying figural matrices, 13:22 minutes; CG: video about nutrition, 12:22 minutes; see Study 1).

Instruments. Figural matrices performance was assessed using a newly developed paper-pencil version of the DESIGMA-Advanced (Becker et al., 2014; Becker et al., 2016; Becker & Spinath, 2014). The paper-pencil version included 28 items following the same six rules as the versions included in Study 1, but depicted different elements. Similar to the construction response format in Study 1, participants had to mark every necessary element required to construct the correct solution for an item in a response area containing all 20 possible elements. Instructions were presented and read to the participants by the respective experimenter of the group and included three item examples. Time to complete the test was limited to 20 minutes.

Intelligence was assessed using three of the four LPS-2 scales used in Study 1 (general knowledge: 60 items, 3 minutes; numerical sequences: 40 items, 5 minutes; mental rotation: 40 items, 2 minutes).

Analyses. As preliminary analyses, we again compared the intelligence test scores between the experimental group and the control group using independent t-tests. Regarding our analyses related to Hypothesis 1, we conducted an independent t-test with group as independent variable and the sum score of correctly solved figural matrices items as dependent variable ($p < .05$). Additionally, we inspected the effect size d according to Cohen (1988) in case of statistically significant results. Regarding our analyses related to Hypotheses 2a–2c, we examined the correlations between figural matrices scores and intelligence scores in (2a) the experimental group and (2b) the control group, and (2c) compared these coefficients.

3.2.3 Results

Preliminary Analyses. Similarly to Study 1, our analyses revealed no statistically significant a priori differences between the groups (LPS-2 sum score: $t(227) = -0.33$; $p = .75$; general knowledge: $t(227) = -0.28$; $p = .78$; numerical sequences: $t(227) = -0.39$; $p = .69$; mental rotation: $t(227) = -0.13$; $p = .90$).

Mean Differences. Descriptive statistics for the figural matrices task (as well as the intelligence test) are depicted in Table 5. Regarding the first hypothesis related to the effect of the video tutorial in Study 2, the independent *t*-test revealed significantly higher figural matrices scores in the experimental group than in the control group ($t(df_{\text{adjusted}} = 218) = 9.17$; $p < .01$) corresponding with an effect size of $d = 1.21$. Thus, the effect of watching a figural matrices video tutorial in Study 2 was similar to the effects revealed in Study 1.

Table 5: Descriptive Statistics for the Figural Matrices and Intelligence Scores in Study 2.

	Experimental Group			**Control Group**		
	M	*SD*	*n*	*M*	*SD*	*n*
Figural Matrices	17.63	6.02	114	9.43	7.46	115
LPS-2 Sum score	79.87	15.79	114	79.22	14.53	115
LPS-2 General knowledge	36.04	8.87	114	35.73	7.77	115
LPS-2 Numerical sequences	21.57	4.10	114	21.37	3.77	115
LPS-2 Mental rotation	22.26	8.12	114	22.12	8.03	115

Correlations with the Intelligence Test. Regarding the correlations between the figural matrices test scores and the LPS-2 sum scores related to Hypothesis 2a–2c in Study 2, our analyses again revealed substantially high correlations in the experimental group (r = .45, p < .01, supporting Hypothesis 2a) as well as in the control group (r = .38, p < .01, supporting Hypothesis 2b). Similar to Study 1, there was no statistically significant difference concerning these correlation coefficients (p = .52), supporting Hypothesis 2c. When inspecting the correlations between the figural matrices scores and the subtest scores, no statistically significant differences emerged related to any subtest, again falling in line with the results of Study 1 (.25 ≤ p ≤ .87). For a full correlation matrix of Study 2, see Table 6.

Table 6: Correlation Matrix for all Variables in Study 2.

	Entire Sample (*N* = 229)				**Experimental Group (*n* = 114)**				**Control Group (*n* = 115)**			
	2.	3.	4.	5.	2.	3.	4.	5.	2.	3.	4.	5.
1. Figural Matrices	.36*	.26*	.42*	.20*	.45*	.37*	.50*	.22*	.38*	.23*	.46*	.24*
2. LPS-2 Sum Score		.78*	.62*	.77*		.80*	.67*	.73*		.76*	.57*	.81*
3. LPS-2 General Knowledge			.34*	.27*			.44*	.24*			.22*	.31*
4. LPS-2 Numerical Sequences				.33*				.32*				.35*
5. LPS-2 Mental Rotation												

Note. * = p < .05.

3.2.4 Discussion

In Study 2, the main findings obtained in Study 1 were replicated. First, related to Hypothesis 1, the differences between the experimental group watching the figural matrices video tutorial and the control group watching the control video in Study 2 revealed to be comparably high (Study 2: d = 1.21; Study 1: d = 1.19 and d = 1.31), supporting the previously in Study 1 demonstrated effects of large magnitude (according to Cohen, 1988) on figural matrices test scores after watching a video tutorial that go beyond the average effects typically found for practice and coaching in meta-analyses (Hausknecht et al., 2007; Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984; Scharfen et al., 2018). Next, related to Hypotheses 2a–c, the correlations between the figural matrices scores and the sum scores of the intelligence test again revealed to be of substantial and comparable magnitude (Study 2: experimental group: r = .45, control group: r = .38; Study 1: .53 ≤ r ≤ .58), falling in line with the previously mentioned earlier evidence on similar correlations ($r_{\text{DESIGMA-Advanced/LPS}}$ = .39, Becker et al., 2014; $r_{\text{BOMAT/LPS-2}}$ = .45, Kreuzpointner et al., 2013). When inspecting the correlation coefficient between the figural matrices scores and the intelligence sum score in the entire sample rather than in the separate groups, the coefficient once more revealed to be slightly lower (r = .36, as shown in Table 6), indicating – at least numerically – again a benefit when keeping the influence of rule knowledge in figural matrices test homogeneous.

Essentially, in Study 2 we replicated the results obtained in Study 1 in a more diverse sample of larger size. Thereby, we strengthened the conclusion that watching a figural matrices video tutorial focusing on teaching the underlying rules increases figural matrices test scores by a large margin and that an important validity aspect is retained (even more so when rule knowledge is kept homogeneous).

Taken together, in Study 1 and Study 2 the effects of video tutorials on intelligence test scores were evidenced on the basis of figural matrices as a popular figural rule-based task and by comparing a video tutorial condition with a standard control condition. However, open questions still remained regarding the effects of video tutorials focusing other popular reasoning tasks (e.g., numerical tasks such as number sequences). Furthermore, adding additional control conditions representing similar approaches test-takers might use when preparing for an assessment might shed additional light regarding the relevance of watching video tutorials. Thus, Study 3 was conducted to consider these aspects.

3.3 Study 3 – Schneider & Sparfeldt (2021a)

Schneider, B., & Sparfeldt, J. R. (2021a). How to solve number series items: Can watching video tutorials increase test scores? *Intelligence*, *87*, 101547. https://doi.org/10.1016/j.intell.2021.101547

In Study 3, we aimed to investigate the effects of video tutorials on the basis of number series as a popular task to assess (numerical) intelligence. Additionally, we expanded the experimental design by including an additional control group representing another approach test-takers might use when preparing for an assessment.

3.3.1 Theoretical Background and Hypotheses

Watching video tutorials might be an effective and efficient approach for test-takers to increase their test scores in an upcoming assessment. In Study 1 and 2 of this dissertation project, we demonstrated these assumed higher test scores in figural matrices tasks for test-takers in an experimental group watching a video tutorial focusing on teaching the underlying rules of figural matrices compared to test-takers in a standard control condition watching a video about nutrition. Additionally, validity aspects in the sense of the correlations with another intelligence test remained comparable in the experimental group and in the control group. However, an open question remained whether watching a video tutorial focusing on another popular reasoning task such as number series also reveals test score increases and comparable validity coefficients. Furthermore, an open question remained if watching such a video tutorial leads to higher test scores than similar approaches test-takers might use when preparing for an assessment such as relying on item examples instead.

Number series represent popular rule-based tasks to assess numerical reasoning and are thus frequently part of broader intelligence test batteries (e.g., Heller & Perleth, 2000; Jäger et al., 1997; Kersting et al., 2008; Liepmann et al., 2007; Lohman, 2011; Sauerland et al., 2008; Schrank et al., 2014; Weiß, 2006; Wonderlic, 1992). In number series tasks, test-takers are required to identify the rules underlying sequences of numbers and then indicate the next number in the series according to these rules (e.g., a series of 1 – 4 – 5 – 8 – 9 – 12 – 13 – ? is solved by alternatingly adding 3 and adding 1 to the numbers and then indicating „16“ as the correct solution). Inspecting a large number of items from several published tests revealed commonalities in the structure of such number series items that can be further illustrated by the processing phases involved when solving an item (e.g., Holzman et al., 1983; Kotovsky & Simon, 1973; see also Holzman et al., 1982; LeFevre & Bisanz, 1986; Loe et al., 2018; Verguts et al., 2002). Building on these item inspections of published tests as well as these postulated processing phases, we focused on illustrating these commonalities related to the structure of number series items and the typical processes in a number series video tutorial. In summary, as main goals of Study 3, we aimed to investigate whether watching such a number series video tutorial leads to 1.) higher test scores than watching an irrelevant (figural matrices) video tutorial and 2.) higher test scores than inspecting and working on number series item examples instead. Additionally, we gathered further evidence on validity aspects in the groups by inspecting the correlations of the number series sum scores with the scores in another intelligence test as well as with high school grade point average (GPA).

Hypothesis 1. We expected that watching a number series video tutorial leads to higher number series mean scores than watching an irrelevant video tutorial about figural matrices.

Hypothesis 2. We expected that watching a number series video tutorial leads to higher number series mean scores than inspecting and working on item examples.

Additional Analyses. Concerning the correlations between the number series sum scores and the sum scores in another intelligence test as well as between the number series sum scores and GPA, we assumed comparable correlation coefficients in the typical size and direction in all groups.

3.3.2 Methods

Participants and Procedure. The sample in Study 3 consisted of *N* = 192 teacher-education students who attended a weekly third semester educational assessment lecture (mean age [based on *N* = 173 participants]: *M* = 22.00; *SD* = 4.26; gender [based on *N* = 175 participants]: 31 % male, 69 % female; one additional student did not strictly work according to instructions and was excluded from the analyses). Similar to Study 1 and Study 2, neither intelligence nor intelligence assessment were part of the curriculum before the study commenced. Regarding the procedure, the students were first randomly assigned to one of three groups. Afterwards, each group was accompanied to a different lecture hall by a corresponding trained experimenter. Next, participants in each group watched an introduction sequence portraying the same person introducing the purpose of the investigation (i.e., information about a task would be presented to help performing better in future assessments). In the experimental group (EG; *n* = 63), this introduction sequence (33 seconds) was followed by the number series video tutorial (see below; 14:25 minutes). In the figural matrices video tutorial control group ($CG_{Matrices}$; *n* = 65), participants first watched the identical introduction sequence, but were presented a figural matrices video tutorial afterwards (14:25 minutes). In the control group focusing on item examples ($CG_{Examples}$; *n* = 64), the participants first watched a similar introduction sequence of comparable length (34 seconds) elucidating that they would have the chance to inspect and work on item examples as well as to reflect about solving these items on their own. Afterwards, the participants of this group were handed a sheet. On the first page of this sheet (and besides instructions to familiarize with the items and to freely switch between pages), the same five number series item examples that were included in the number series video tutorial of the experimental group were presented (without solution). On the second page, the solved versions of these number series items were depicted along with a dedicated space for additional notes. The available time was equal to the duration of the video tutorials presented to the other groups (14:25 minutes). After the respective video tutorial (or inspecting and working on the item examples) had concluded, the procedure continued identically for all three groups with the presentation of number series items. Additional data (intelligence test scores, GPA) were available from a data collection three months prior to the investigation (intelligence – EG: *n* = 59; $CG_{Matrices}$: *n* = 57; $CG_{Examples}$: *n* = 59; GPA – EG: *n* = 59; $CG_{Matrices}$: *n* = 56; $CG_{Examples}$: *n* = 58) with comparably high participation rates (EG/$CG_{Matrices}$/$CG_{Examples}$: intelligence – 94%/88%/92%; GPA – 94%/86%/91%).

Video Tutorials. Concerning the number series video tutorial, we developed a number series illustration model derived from inductive-deductive iteration loops of inspecting in total 287 number series items found in various published tests (Heller & Perleth, 2000; Jäger et al., 1997; Kersting et al., 2008; Liepmann et al., 2007; Weiß, 2006) and further considering theoretical aspects such as the four processing phases underlying number series items (e.g., Holzman et al., 1983; Kotovsky & Simon, 1973, Loe et al., 2018; Verguts et al., 2002). This illustration model (see Table 7) was used as a basis to explain the commonalities of number series items in the video tutorial.

Table 7: Number Series Illustration Model Developed for Study 3.

Component	Explanation	Item Example	Item Description	Model Illustration
Period	The recurring pattern within an item; start and end of a period is signified by brackets.	2 4 7 9 12 14 17 **?**	Period Length: 2 Complexity: Low One Period	[+2 +3]
Period length	Length of the recurring pattern (number of recurring operations within a period); signified by the number of operations in square brackets.	4 6 8 10 12 14 16 **?**	Period Length: 1 Complexity: Low One Period	[+2]
Period Complexity	Complexity of a period (low or high); periods with low complexity remain consistent throughout the item, periods with high complexity are connected by an additional operation in between periods.	3 5 8 13 19 27 36 **?**	Period Length: 2 Complexity: High One Period	$[+2\ +3]_{+3}$
Parallel Periods	Interwoven periods not related to each other (e.g., relating to every other number instead to adjacent numbers).	5 7 10 13 15 19 20 **?**	Period Length: 1 Complexity: Low Two Periods	[+5] [+6]

Besides explaining the components of the illustration model, the four processing phases were introduced in detail to guide test-takers step by step through the solution process (1. *relation detection*: finding relations between numbers; 2. *discovery of periodicity*: discovering recurring patterns; 3. *completion of pattern description*: discovering a pattern over the whole sequence of numbers; 4. *extrapolation*: using this pattern to indicate the solution). Additionally, we further exemplified these processes (e.g., by using arcs to indicate related numbers) and portrayed five item examples specifically constructed for the purpose of the video. At the end of the video, a summary regarding the most important aspects was provided. An initial version of the video was improved after a small pilot study in which students not participating in the investigation provided qualitative feedback. The final number series video tutorial (see Figure 3; for a more detailed depiction, see Appendix C) had a duration of 14:25 min.

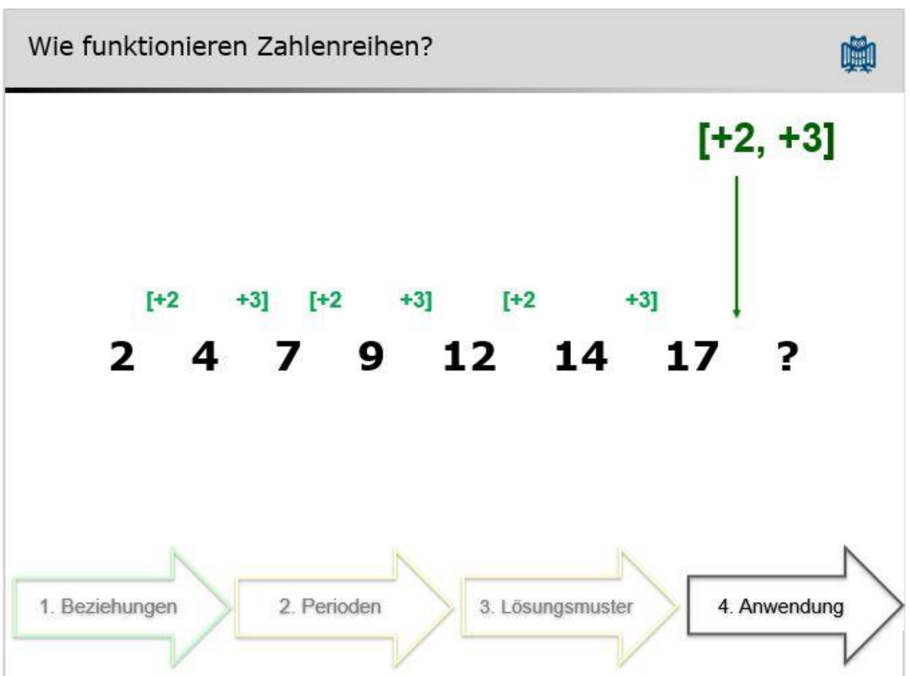

Figure 3. Original Screenshot of the Number Series Video Tutorial used in Study 3.

Regarding the figural matrices video tutorial, we used an updated version of the video tutorial used in Study 1 and Study 2 that was further improved regarding aspects such as audio quality, design, and individual explanations. Accordingly, the number series video tutorial and the figural matrices video tutorial resulted to be similar in structural and didactical aspects (e.g., guiding step by step through the solution process, showing example items, illustrating explanations using arrows, presenting a summary at the end). The updated figural matrices video tutorial used in Study 3 (see Figure 4) had a duration of 14:25, as well. Both video tutorials are openly available to the scientific community (see Appendix A for the link).

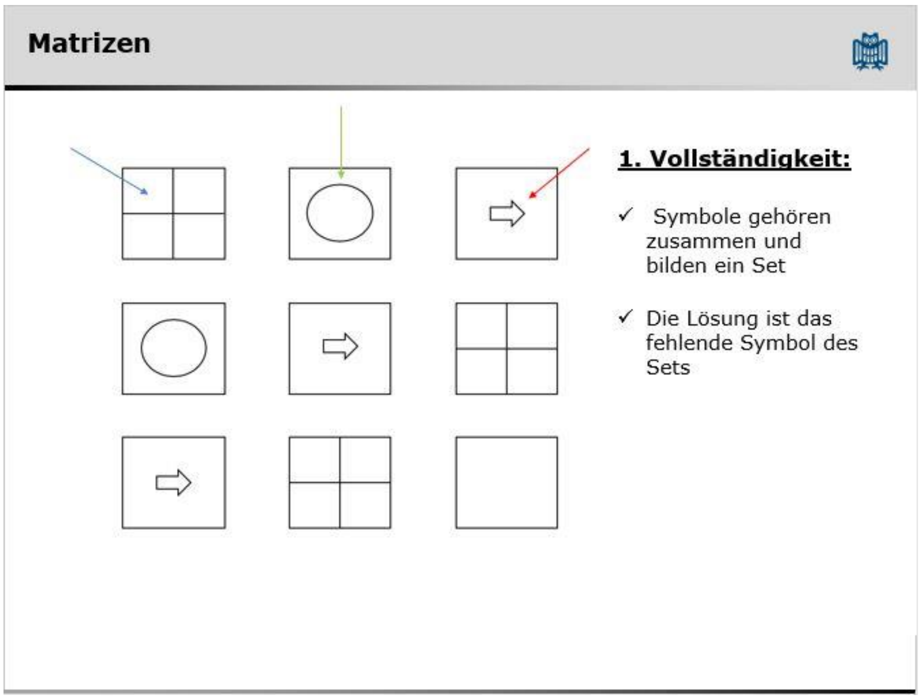

Figure 4. Original Screenshot of the Updated Figural Matrices Video Tutorial used in Study 3.

Instruments. Number series performance was assessed using the sum score of 71 number series items. Precisely, we presented the 20 number series items of the *I-S-T 2000 R* (10 minutes; Liepmann et al., 2007), the 20 number series items of the *WIT-2* (10 minutes; Kersting et al., 2008), as well as 31 newly constructed number series items (16 minutes; one further item was excluded due to a typographical error).

Intelligence in the sense of *g* was assessed using the sum score of the *IST-Screening* (Liepmann et al., 2012), comprising three subtests assessing verbal (*verbal analogies*; 20 items, 6 minutes), figural (*figural matrices*; 20 items, 10 minutes) and numerical (*number series*; 20 items, 10 minutes) reasoning. Grade point average (GPA) of the high school graduation certificate (“Abiturnote”; ranging from “4”, worst grade, to “1”, best grade) was used to indicate scholastic achievement. GPA was reversely scored so that higher numbers corresponded with better achievement.

Analyses. As preliminary analyses to test whether the intelligence sum scores differed statistically significantly between the experimental group and the control groups before the experiment took place, we ran preliminary analyses using two independent *t*-tests (EG vs. $CG_{Matrices}$; EG vs. $CG_{Examples}$). Related to Hypothesis 1, we analyzed whether the participants of the experimental group watching the number series video tutorial solved more number series items correctly than the participants of the control group watching the figural matrices video tutorial. Hence, we conducted a directed linear contrast (EG = 1, $CG_{Matrices}$ = – 1, $CG_{Examples}$ = 0; one-tailed; α = .05) and examined the effect size *d* (Cohen, 1988). Additionally, we ran the analyses for each subtest. Related to Hypothesis 2, we analyzed whether the participants of the experimental group watching the number series video tutorial solved more number series items correctly than the participants inspecting and working on item examples. Accordingly, we again conducted a directed linear contrast (EG = 1, $CG_{Matrices}$ = 0, $CG_{Examples}$ = – 1; one-tailed; α = .05) and examined the effect size *d* (Cohen 1988) in addition to running these analyses for each subtest.

Related to the correlations between the number series scores and the intelligence scores as well as the correlations between the number series scores and GPA, we calculated the Pearson correlations coefficients in every group and tested whether the respective correlation differed from zero (one-tailed). Furthermore, we statistically compared the correlation coefficients in the experimental group with the respective correlation coefficients in each control group (two-tailed).

3.3.3 Results

Preliminary Analyses. Regarding the preliminary analyses, our results revealed no statistically significant a priori mean differences in intelligence (EG vs. $CG_{Matrices}$: $t(df_{adjusted} = 104.22) = -0.44$, $p = .66$; EG vs. $CG_{Examples}$: $t(df_{adjusted} = 111.37) = -0.24$, $p = .80$). Descriptive Statistics are depicted in Table 8.

Table 8: Descriptive Statistics of the Intelligence Test in Study 3.

	Whole Sample (*N* = 175)		EG (*n* = 59)		$CG_{Examples}$ (*n* = 59)		$CG_{Matrices}$ (*n* = 57)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
IST-Screening sum	47.39	5.07	47.17	6.00	47.42	4.88	47.60	4.22
IST-Screening verbal	16.11	1.62	16.08	1.70	16.07	1.68	16.19	1.48
IST-Screening numerical	16.34	2.93	16.34	3.26	16.15	3.09	16.54	2.41
IST-Screening figural	14.94	2.22	14.75	2.74	15.20	2.04	14.86	1.78

Mean Differences. Related to Hypothesis 1, the participants in the experimental group watching the number series video tutorial solved statistically significantly more number series items than the participants in the control group watching the figural matrices video tutorial ($t(189) = 2.46$, $p < .01$, $d = 0.44$). These results were also demonstrated when comparing the subtest scores (I-S-T 2000 R: $t(189) = 2.21$, $p = .02$, $d = 0.41$; WIT-2: $t(189) = 2.41$, $p < .01$, $d = 0.43$; newly constructed items: $t(189) = 2.08$, $p = .02$, $d = 0.37$). Thus, the results support the hypothesis that watching a number series video tutorial increases number series test scores. Related to Hypothesis 2, the participants in the experimental group watching the number series video tutorial solved statistically significantly more number series items than the participants in the control group inspecting and working on item examples ($t(189) = 1.68$, $p <.05$, $d = 0.30$). When comparing the subtest scores individually, the results showed no statistically significant differences (I-S-T 2000 R: $t(189) = 1.47$, $p = .07$; $d = 0.26$; WIT-2: $t(189) = 1.62$, $p = .05$; $d = 0.30$; newly constructed items: $t(189) = 1.48$, $p = .07$, $d = 0.25$). Thus, the results support the Hypothesis that watching a number series video tutorial leads to higher test scores than inspecting item examples instead, although a higher number of items was required to demonstrate such an effect statistically (as reflected in the statistically significant group differences regarding the sum score consisting of all 71 items, which failed to reach statistical significance on the subtest level of 20/20/31 items [I-S-T 2000 R/WIT-2/newly constructed]). Descriptives are presented in Table 9.

Table 9: Descriptive Statistics of the Number Series Scores in Study 3.

	Whole Sample (N = 192)		EG (n = 63)		$CG_{Examples}$ (n = 64)		$CG_{Matrices}$ (n = 65)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sum score	49.24	12.67	52.33	12.13	48.59	12.98	46.89	12.46
I-S-T 2000 R	15.77	4.22	16.68	3.64	15.59	4.55	15.05	4.30
WIT-2	14.02	4.05	14.98	3.57	13.83	4.12	13.28	4.29
newly constructed	19.46	5.73	20.67	6.20	19.17	5.64	18.57	5.21

Correlations with Intelligence and Grades. Concerning the correlations between number series and general intelligence, the results revealed statistically significant and substantially large correlations between the number series sum scores and the intelligence sum scores in all groups (EG: $r = .63$, $p < .01$; $CG_{Matrices}$: $r = .57$, p <.01; $CG_{Examples}$: $r = .56$, $p < .001$). Furthermore, these correlations did not differ statistically significantly between the experimental group and the control groups (EG vs. $CG_{Matrices}$: $p = .62$; EG vs. $CG_{Examples}$: $p = .58$). Concerning the correlations between number series and GPA, the results revealed statistically significant correlations in the typical range and direction in all groups (EG: $r = .38$, $p < .01$; $CG_{Matrices}$: $r = .25$, $p = .03$; $CG_{Examples}$: $r = .27$, $p = .02$). These correlations did not differ statistically significantly between the experimental group and the control groups, as well (EG vs. $CG_{Matrices}$: $p = .44$; EG vs. $CG_{Examples}$: $p = .50$). For the full correlation matrices between all variables, see Tables 10–13.

Table 10: Correlations of All Variables in Study 3: Entire Sample.

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. IST-Screening verbal	—								
2. IST-Screening numerical	.26*	—							
3. IST-Screening figural	.28*	.40*	—						
4. IST-Screening sum score	.59*	.84*	.76*	—					
5. Number Series (IST 2000R)	.22*	.51*	.38*	.53*	—				
6. Number Series (WIT 2)	.22*	.47*	.35*	.50*	.78*	—			
7. Number Series (newly constructed)	.22*	.47*	.40*	.52*	.61*	.72*	—		
8. Number Series (sum score)	.25*	.54*	.43*	.58*	.87*	.92*	.90*	—	
9. GPA	.18*	.28*	.30*	.35*	.23*	.32*	.28*	.31*	—

Note. $N = 173$; * $p < .05$ (one tailed)

Table 11: Correlations of All Variables in Study 3: Control Group (Matrices).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. IST-Screening verbal	—								
2. IST-Screening numerical	.33*	—							
3. IST-Screening figural	.31*	.35*	—						
4. IST-Screening sum score	.66*	.83*	.73*	—					
5. Number Series (IST 2000R)	.17	.49*	.22	.43*	—				
6. Number Series (WIT 2)	.31*	.55*	.23*	.51*	.72*	—			
7. Number Series (newly constructed)	.35*	.57*	.26*	.56*	.57*	.69*	—		
8. Number Series (sum score)	.32*	.62*	.27*	.57*	.85*	.90*	.88*	—	
9. GPA	−.01	.18	.17	.17	.13	.25*	.27*	.25*	—

Note. $N = 56$; * $p < .05$ (one tailed)

Table 12: Correlations of All Variables in Study 3: Control Group (Examples).

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. IST-Screening verbal	—								
2. IST-Screening numerical	.14	—							
3. IST-Screening figural	.26*	.33*	—						
4. IST-Screening sum score	.54*	.82*	.72*	—					
5. Number Series (IST 2000R)	.26*	.51*	.47*	.61*	—				
6. Number Series (WIT 2)	.19	.37*	.39*	.46*	.84*	—			
7. Number Series (newly constructed)	.14	.32*	.47*	.45*	.60*	.72*	—		
8. Number Series (sum score)	.21	.44*	.50*	.56*	.89*	.93*	.88*	—	
9. GPA	.22*	.25*	.27*	.35*	.21	.28*	.25*	.27*	—

Note. $N = 58$; * $p < .05$ (one tailed)

Table 13: Correlations of All Variables in Study 3: Experimental Group.

	1.	2.	3.	4.	5.	6.	7.	8.	9.
1. IST-Screening verbal	—								
2. IST-Screening numerical	.32*	—							
3. IST-Screening figural	.29*	.50*	—						
4. IST-Screening sum score	.59*	.86*	.81*	—					
5. Number Series (IST 2000R)	.23*	.54*	.47*	.57*	—				
6. Number Series (WIT 2)	.21	.54*	.46*	.56*	.76*	—			
7. Number Series (newly constructed)	.22*	.56*	.46*	.57*	.65*	.75*	—		
8. Number Series (sum score)	.24*	.61*	.51*	.63*	.86*	.91*	.93*	—	
9. GPA	.29*	.38*	.46*	.50*	.34*	.43*	.30*	.38*	—

Note. $N = 59$; * $p < .05$ (one tailed)

3.3.4 Discussion

Study 3 revealed three main findings. First, the experimental group watching the number series video tutorial solved more number series items than the control group watching the figural matrices video tutorial (d = 0.44). Second, the experimental group watching the number series video tutorial also solved more number series items than the control group inspecting and working on numer series item examples (d = 0.30). Thereby, Study 3 not only demonstrated that number series test scores can be increased by watching a number series tutorial before a corresponding assessment, but that using such an approach is also indicated to be more effective than practicing by oneself using item examples. Noteworthy, these increases fall in line with the effect sizes typically reported for practice and coaching effects (e.g., practice: d = 0.40, Hausknecht et al., 2007; d = 0.42, Kulik, Kulik, & Bangert, 1984; *SMCR* = 0.37, Scharfen et al., 2018; coaching: d = 0.43, Kulik, Bangert-Drowns, & Kulik, 1984; practice and coaching: d = 0.64, Hausknecht et al., 2007), but revealed to be smaller than the effects evidenced for figural matrices in this dissertation project (see section 4.2.1 for a more detailed elaboration on this issue). Third, inspecting the correlation coefficients between the number of solved number series items and the intelligence sum scores as well as between the number of solved number series items and grades revealed substantial and comparable correlations in the typical range and direction in all groups (intelligence: EG: r = .63; $CG_{Matrices}$: r = .57; $CG_{Examples}$: r = .56; grades: EG: r = .38; $CG_{Matrices}$: r = .25; $CG_{Examples}$: r = .27). These correlations (at least) fall in line with previously reported evidence on correlations between number series and intelligence (e.g., number series of the I-S-T-2000 R and the IST-Screening: r = .66, Liepmann et al., 2012; number series of the WIT-2 and the Wonderlic intelligence-test: r = .56, Kersting et al., 2008) as well as between number series and grades (e.g., number series and GPA: r = –.18/–.23, Kempf & Meder, 1993; number series and grades of subjects German/English/Mathematics: r = –.10/–.12/–.24, Liepmann et al., 2012; higher negative values indicate better achievement). Thus, like in Study 1 and Study 2, Study 3 suggests that important validity aspects are retained.

In summary, Study 3 expanded on Study 1 and Study 2 by demonstrating increased test scores after watching a video tutorial focusing on another popular (but numerical) reasoning task (and by demonstrating higher test scores than using item examples). However, the mechanisms underlying these test score increases still remained largely unknown. Furthermore, and besides effecting test score increases, an open question remained whether watching a video tutorial influences other aspects relevant when assessing intelligence. Potentially, watching a video tutorial stimulates test-takers to correspondingly adapt their test-taking behavior, for example, by notating essential elements illustrated in the video tutorial. Thus, Study 4 was conducted to shed light on these questions by examining note-taking behavior after watching a number series video tutorial.

3.4 Study 4 – Schneider & Sparfeldt (2021b)

Schneider, B., & Sparfeldt, J. R. (2021b). How to get better: Taking notes mediates the effect of a video tutorial on number series. *Journal of Intelligence*, 9, 55. https://doi.org/10.3390/jintelligence9040055

In Study 4, shifting the focus towards the effects on test-taking behavior, we investigated whether watching a number series video tutorial also stimulates test-takers to take more notes, and whether using such notes mediates the demonstrated test score increases.

3.4.1 Theoretical Background and Hypotheses

In Study 1–3 of this dissertation project, increased test scores after watching video tutorials were demonstrated. However, various elements relevant for increasing test scores were combined in these video tutorials (e.g., teaching a step-by-step approach, providing relevant information about the tasks, explicating rules or processes). Thus, the specific mechanisms underlying the demonstrated test score increases still remain largely unknown. Potentially, watching a video tutorial also stimulates test-takers to use more effective as well as efficient approaches to tackle a test. As one effective and efficient method to approach number series tests, test-takers might notate relevant aspects shown in a corresponding video tutorial to help solving number series items.

Particularly relevant notes that can be easily integrated into the solution process might include elements of the number series processing phases (see, e.g., Holzman et al., 1982, 1983; Kotovsky & Simon, 1973). Specifically, jotting down the numerical relation and the mathematical operation between numbers allows test-takers to externalize essential elements related to the first processing phase (*relations detection*). Thereby, important cognitive functions and goals such as identifying and focusing relevant parts of an item and and enabling an easier transfer into working memory are considered (Weinstein & Mayer, 1986). Furthermore, subsequent proccesses such as discovering periodicity or finding the solution pattern are facilitated, as well – ultimately allowing an easier extrapolation of the pattern to the missing number in a series. Test-takers should display such a behavior more frequently after watching a number series video tutorial exemplifying such notes, and more items reflecting such a behavior should also be generally associated with higher scores in number series tests. Finally, such notes might explain the corresponding test scores increases after watching a number series video tutorial. In summary, in Study 4, we aimed to investigate the effects of watching a number series video tutorial on note-taking behavior by comparing the notes of an experimental group watching such a video tutorial with the notes of a control group watching an irrelevant video tutorial, thereby clarifying a mechanism assumed to underlie the associated test score increases in number series tasks.

Hypothesis 1. We expected that participants of the experimental group watching a number series video tutorial took notes for more number series items than participants of the control group watching an irrelevant video tutorial about figural matrices.

Hypothesis 2. We expected positive and substantial correlation coefficients between the amount of number series items with notes and the number series sum scores in each group.

Hypothesis 3. We expected the effect on number series test scores after watching a number series video tutorial to be mediated by the amount of number series items with notes.

3.4.2 Methods

Participants and Procedure. For Study 4, we relied on the data obtained in Study 3. Specifically, we relied the participants of the experimental group watching the number series video tutorial (EG: n = 63) as well as on the participants of the control group watching the figural matrices video tutorial (CG: n = 65), resulting in N = 128 teacher-education students attending a weekly second-year lecture on educational assessment (age based on N = 115 participants: M = 22.11, SD = 4.18; sex based on N = 116 participants: 38 % male, 62 % female). The analyses were based on the N = 110 participants indicating at least one note for at least one number series item in every subtest (86 % of the entire sample: EG: n = 58, CG: n = 52). Thereby, we ensured the participants included in the analyses were aware taking notes was allowed and motivated enough to show such a behavior.

Video Tutorials. Participants watched the video tutorials previously described in Study 3 (i.e., EG: number series video tutorial; CG: figural matrices video tutorial; duration of both tutorials: 14:25 minutes).

Instruments. Participants worked on the 71 number series items previously described in Study 3 (i.e., 20 items of the I-S-T 2000 R, Liepmann et al., 2007; 20 items of the WIT-2, Kersting et al., 2008; 31 structurally similar items newly constructed by the authors). Performance was indicated by the sum score of all correctly solved items (as well as the subtest sum scores). An item was considered to indicate a note when the mathematical relation between two numbers (i.e., number as well as operator; e.g., “+3”) was notated at least once.

Analyses. As preliminary analysis, we tested a priori intelligence differences between the experimental group and the control group by conducting an independent *t*-test using the IST-Screening sum scores obtained three months prior to the experiment. As basis for the hypothesis-related analyses, we analyzed whether the students of the experimental group solved more number series items than participants of the control group by conducting an independent *t*-test with the number series sum score as dependent variable and group as independent variable (one-tailed; α = .05) in addition to inspecting Cohen’s *d* (Cohen, 1988). Concerning Hypothesis 1, we analyzed whether the students of the experimental group took notes for more items than the participants of the control group by conducting an independent *t*-test with the number of number series items with notes as dependent variable and group as independent variable (one-tailed; α = .05). Furthermore, we inspected Cohen’s *d*, as well. Concerning Hypothesis 2, we inspected the Pearson correlations between the number series sum scores and the number of items with notes in both groups and analyzed whether the respective coefficients differed statistically significantly from zero (one-tailed; α = .05). Concerning Hypothesis 3, we conducted a mediation analysis (using PROCESS 3.5; Hayes, 2018) with predictor x = group (EG = 1, CG = 0), mediator m = number of items with notes, and criterion y = number series sum score. Specifically, we examined the 95 % - bootstrapping confidence intervals of the indirect effect (as recommended by Hayes, 2018; number of samples = 10,000). Additionally, we conducted the analyses for each subtest separately by using the number of items with notes in the respective subtest as well as using the corresponding subtest score.

3.4.3 Results

Preliminary Analyses. Using the sum scores of the IST-Screening (Liepmann et al., 2012) available for N = 99 participants, the preliminary analysis revealed no statistically significantly mean differences between the students of the experimental group and the students of the control group: $t(\mathrm{df}_{adjusted} = 93.39) = .42$; $p = .67$.

Mean Differences. Regarding the differences in number series test scores as basis for the subsequent analyses, the participants of the experimental group watching the number series video tutorial solved statistically significantly more items than the participants of the control group watching the figural matrices video tutorial ($t(108) = 2.52$, $p < .01$, $d = 0.48$). This result was also demonstrated when inspecting each subtest seperately (I-S-T 2000 R: $t(108) = 2.40$, $p < .01$, $d = 0.46$; WIT-2: $t(108) = 2.54$, $p < .01$, $d = 0.48$; newly constructed ($t(\mathrm{df}_{adjusted} = 105.56) = 1.98$, $p = .03$, $d = 0.37$). Regarding the analyses related to Hypothesis 1, the independent t-tests showed that the students in the experimental group took notes for a higher number of items than participants in the control group ($t(\mathrm{df}_{adjusted} = 92.67) = 2.09$, $p = .02$, $d = 0.41$), supporting Hypothesis 1. When inspecting the subtests, a similar pattern was demonstrated regarding the I-S-T 2000 R ($t(\mathrm{df}_{adjusted} = 92.09) = 2.38$, $p = .01$, $d = 0.46$) as well as the WIT-2 ($t(108) = 2.34$, $p = .01$, $d = 0.45$), but not regarding the newly constructed items ($t(108) = 1.06$, $p = .15$, $d = 0.20$). Descriptive Statistics are depicted in Table 14.

Table 14: Descriptive Statistics for Study 4.

	Experimental Group (*n* = 58)		**Control Group (*n* = 52)**	
	M	***SD***	***M***	***SD***
Items with notes				
Sum	57.50	9.48	52.92	12.95
I-S-T 2000 R	16.91	3.39	15.04	4.68
WIT-2	17.10	2.83	15.62	3.81
Newly constructed	23.48	5.41	22.27	6.60
Number series scores				
Sum	51.53	12.06	45.77	11.90
I-S-T 2000 R	16.47	3.69	14.63	4.33
WIT-2	14.88	3.65	12.96	4.27
Newly constructed	20.19	6.03	18.17	4.63

Correlations. Regarding the analyses related to Hypothesis 2, the number of solved number series items and the number of items with notes correlated statistically significantly and substantially in the experimental group (r = .66, p < .01) as well as in the control group (r = .75, p < .01), thus supporting Hypothesis 2. Similar result were revealed for each subtest in the experimental group (I-S-T 2000 R: r = .54, p < .01; WIT-2: r = .72, p <.01; newly constructed: r = .60, p <.01) as well as in the control group (I-S-T 2000 R: r = .53, p <.01; WIT-2: r = .87, p <.01; newly constructed: r = .66, p <.01) Correlations between all variables are depicted in Table 15.

Table 15: Correlations between all Variables in Study 4.

	1	2	3	4	5	6	7	8
Experimental Group (*n* = 58)								
1. Notes: Sum	–	.69 *	.87 *	.86 *	.66 *	.57 *	.63 *	.58 *
2. Notes: I-S-T 2000 R		–	.58 *	.28 *	.36 *	.54 *	.32 *	.19
3. Notes: WIT-2			–	.64 *	.69 *	.61 *	.72 *	.57 *
4. Notes: Newly constructed				–	.56 *	.34 *	.52 *	.60 *
5. Scores: Sum					–	.85 *	.92 *	.92 *
6. Scores: I-S-T 2000 R						–	.77 *	.63 *
7. Scores: WIT-2							–	.76 *
8. Scores: Newly constructed								–
Control Group (*n* = 52)								
1. Notes: Sum	–	.80 *	.83 *	.92 *	.75 *	.72 *	.66 *	.66 *
2. Notes: I-S-T 2000 R		–	.51 *	.57 *	.43 *	.53 *	.34 *	.29 *
3. Notes: WIT-2			–	.68 *	.86 *	.72 *	.87 *	.72 *
4. Notes: Newly constructed				–	.68 *	.62 *	.54 *	.66 *
5. Scores: Sum					–	.88 *	.92 *	.90 *
6. Scores: I-S-T 2000 R						–	.74 *	.65 *
7. Scores: WIT-2							–	.75 *
8. Scores: Newly constructed								–

Note. * p < .05 (two-tailed). Notes = Items with notes. Scores = Number series scores.

Mediation Analysis. Regarding the analyses related to Hypothesis 3, the mediation analysis revealed an indirect effect of 3.41 (SE = 1.74), with bootstrapping intervals not including zero (95% CI: 0.26, 7.06). Thus, the results indicate that the effect of watching the number series video tutorial on number series test scores was mediated by the number of items with notes. Regarding each subtest, the results revealed to be similar for the I-S-T 2000 R (indirect effect = 0.99, SE = 0.47; 95% CI: 0.18, 2.00) and the WIT-2 (indirect effect = 1.43, SE = 0.63; 95% CI: 0.25, 2.73), but not for the newly constructed items (indirect effect = 0.67, SE = 0.65; 95%CI: –0.56, 1.99). No direct effects of watching the number series video tutorial on number series test scores revealed to be statistically significant (see Figure 5 for the mediation model).

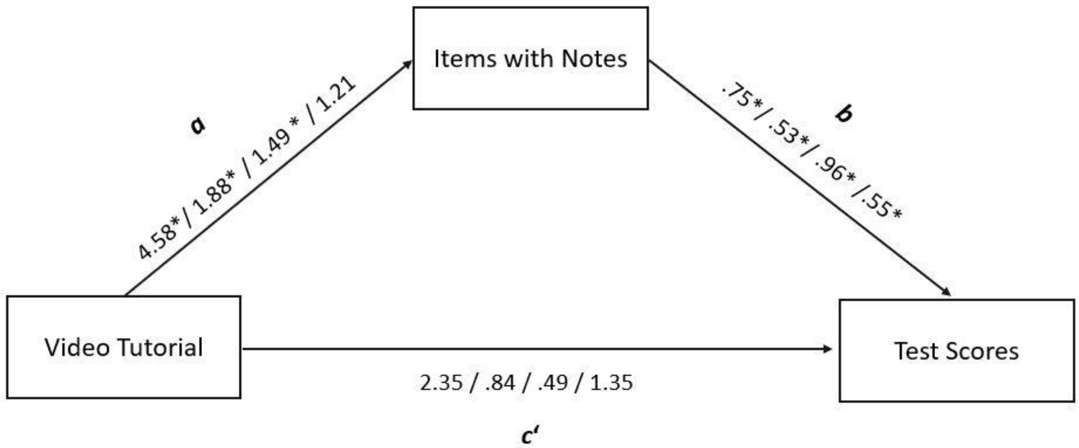

Figure 5. Mediation Model in Study 4. Sum/I-S-T 2000 R/WIT-2/Newly Constructed. **p* < .05.

3.4.4 Discussion

Study 4 revealed three main findings. First, test-takers watching a number series video tutorial took notes for significantly more number series items than test-takers watching an irrelevant video tutorial (*d* = 0.41). Thus, the results suggest that watching such a video tutorial stimulates test-takers to adapt their test-taking behavior by more frequently relying on the illustrated elements when solving number series items. Second, the number of items with notes correlated substantially with the number series sum scores in both the experimental group (*r* = .66) and the control group (*r* = .75). Accordingly, notating the relations between numbers is suggested to be a generally effective behavior for solving number series items. However, it should be noted that this conclusion is based on correlations, precluding causality. Additional studies investigating this aspect more systematically and experimentally (e.g., using an experimental group encouraged to take notes and a control group prohibited to take notes; inspecting log files to consider when and how such notes are interwined into the solution process) can help clarify this aspect further. Third, concerning the mediation analysis, the bootstrapping intervals suggest the effect of watching a number series video tutorial on number series test scores to be mediated by the number of items with notes. Hence, it is indicated that taking notes serves as mechanism for the associated test score increases after watching a number series video tutorial. As such, notating relations between numbers in a series externalizes important information into a scaffolding structure that can help free up cognitive capacity. Thereby, subsequent processes such as discovering periods and finding the solution pattern are facilitated, ultimately helping with solving number series items. Nevertheless, the results were revealed regarding the analyses for all number series items and for the first two subtests, but not for the newly constructed items. Potentially, motivation to use notes was reduced for this last subtest, as psychometric properties of the item set still revealed to be adequate (e.g., Cronbach's α =.85; correlations with the other subtest scores of $.63 \leq r \leq .76$, thus similar in magnitude to the correlation of $r = .73$ reported in the WIT-2 test manual regarding the correlation between the number series items of the I-S-T 2000 R and the number series items of the WIT-2; Kersting et al., 2008, p. 127). For additional insight exploring such a possibility, additional analyses were conducted to evaluate whether note-taking behavior decreased across subtests. Accordingly, the demonstrated number of items with notes in each subtest was calculated relatively to the theoretical maximum number of items with notes in that subtest (I-S-T 2000 R: 20; WIT-2: 20; newly constructed: 31). Dependent *t*-test analyses revealed statistically significantly less number of items with notes regarding the newly constructed items in comparison to the I-S-T 2000 R ($t(109) = -3.14$; $p < .01$) and regarding the newly constructed items in comparison to the WIT-2 ($t(109) = -5.63$; $p < .01$), but not regarding the WIT-2 in comparison to the I-S-T 2000 R ($t(109) = 1.09$; $p = .14$). Thus, and although note-taking behavior was nevertheless shown to be fairly consistent across all three subtests ($\alpha = .76$ using the sum of demonstrated items with notes in each subtest), the assumption of reduced motivation to take notes specifically for the last set of items seems possible.

In summary, Study 4 demonstrated taking specific notes to be an effective (and efficient) behavior to increase number series test scores. Moreover, Study 4 showed that watching a number series video tutorial stimulates such a behavior that was also shown to mediate the number series test score increases. However, it still remains unknown whether such a behavior is also demonstrated more frequently in other tasks (e.g., figural matrices) after watching a corresponding video tutorial.

3.4.5 Supplemental Analyses: Notes for Figural Matrices

As it was shown in Study 4 that watching a number series video tutorial stimulates test-takers to take more notes in number series tasks, the question arises whether note-taking behavior is also stimulated in figural matrices tasks after watching a figural matrices video tutorial. Similarly to number series, test-takers might efficiently notate important elements shown in the video tutorial such as symbols to free up cognitive resources and facilitate finding the solution. As additional material and data related to figural matrices were available from the data collection of Study 3 and Study 4, supplemental analyses were conducted to provide first insight into such a possibility.

Specifically, after the participants had finished working on the number series items, participants were presented 10 paper-pencil figural matrices items of the DESIGMA-Advanced in a distraction format (Becker et al., 2016; Becker et al., 2014; Becker & Spinath, 2014). The items were selected based on reported item difficulty in the test manual as well as balanced rules and number of rules to closely represent a shorter version of the test. Regarding presentation order, the items were arranged accordingly and with increasing difficulty (e.g., first, four items including one rule; then, three items including two rules; next, two items including three rules; finally, one item including four rules; available time to complete the items was 10 minutes). For the analyses, the n = 63 participants watching the number series video tutorial and the n = 65 participants watching the figural matrices video tutorial were inspected regarding the number of solved figural matrices items as well as the number of items with notes. An item was considered to include a note when at least one symbol was indicated in (or next to) the empty solution field of the 3×3 matrix.

Regarding the number of solved figural matrices items, the participants in the group watching the figural matrices video tutorial (M = 7.74, SD = 1.65) solved significantly more figural matrices items than the participants in the group watching the number series video tutorial (M = 4.46, SD = 2.12; $t[df_{adjusted} = 117.23] = 9.73$, $p < .01$, d = 1.73). Thereby, the supplemental analyses further strengthened the evidence on increased test scores after watching a figural matrices video tutorial reported in Study 1 and Study 2 of this dissertation project.

Regarding the number of figural matrices items with notes, an interesting pattern occurred when inspecting the amount of participants taking such notes: Whereas in the group watching the figural matrices video tutorial a substantial percentage of the participants indeed relied on using such notes (38 of 65 participants, equaling 58.46 %), only a very small percentage indicated such a behavior in the group watching the number series video tutorial (3 of 63 participants, equaling 4.76 %). Accordingly, a more conservative approach was followed to examine statistically significant differences in note-taking. Specifically, a Chi²-test was conducted using a 2×2 cross table with the variables group and note-taking behavior (group: figural matrices video tutorial or number series video tutorial; note-taking behavior: one or more notes were taken or no notes were taken). No expected cell frequencies were below 5. The results showed a statistically significant correlation between group and note-taking behavior ($\chi^2(1) = 42.37$, $p < .01$, $\varphi = .58$) large in effect size according to Cohen (1988). When examining the (one-tailed) pearson correlation between the number of correctly solved figural matrices items and the number of figural matrices items with notes for the $n = 38$ participants taking at least one note in the group watching the figural matrices video tutorial, no statistically significant result was revealed ($r = .17$, $p = .15$). As only $n = 3$ participants took one or more notes in the group watching the number series video tutorial, results were not statistically interpretable for this group.

Taken together, the supplemental analyses complemented the previously investigated effects of video tutorials in several instances. First, the previously reported test score increases in figural matrices tasks after watching a corresponding figural matrices video tutorial were demonstrated once more (yielding an even larger effect size), further strengthening the most essential target of this dissertation project. Second, and besides number series tasks, the supplemental analyses showed that note-taking behavior is also demonstrated more frequently in figural matrices tasks after watching a corresponding tutorial. However, unlike in number series tasks, the analyses indicated a large number of participants not taking any notes for figural matrices items, particularly for participants not watching the relevant tutorial. Potentially, explicating symbols in the solution field is less intuitive than simply notating the relationship of two numbers in number series tasks, or requires further and more explicit instruction. Furthermore, the correlation between the number of items with notes and the number of solved figural matrices items failed to be statistically significant (a posteriori calculated power: $1–\beta = 0.27$). Nevertheless, one should keep in mind that several factors might be involved in contributing to such a result (e.g., reduced/small sample size; relying on only 10 figural matrices items; exhaustion effects after having worked on 71 number series items; broader and more ambiguous possibilities of explicating notes). Considering these limitations, revealing at least a numerically small correlation in the expected direction suggests that higher-powered further studies specifically dedicated to clarifying this aspect seem promising – as taking notes more frequently after watching a video tutorial as starting point for such investigations has already been demonstrated in these supplemental analyses.

4. General Discussion

The aim of this dissertation project was to investigate the effects of watching video tutorials on intelligence test scores. This fourth and last overarching chapter first provides a summary of the main findings of the dissertation project before elaborating on the theoretical and practical implications as well as associated future research directions. Finally, the chapter reflects on critical acknowledgements surrounding key aspects of the project.

4.1 Summary of Main Findings

Across the four studies conducted for this dissertation project, the relevance of video tutorials as approach to increase test scores has been demonstrated. In Study 1, we evidenced this relevance by showing substantially higher figural matrices test scores ($d \geq 1.19$) for participants in an experimental group watching a short figural matrices video tutorial focusing on teaching the underlying rules compared to a control group watching an irrelevant video about nutrition. Furthermore, correlation coefficients between the figural matrices test scores and the test scores of an intelligence test remained comparibly high in both groups ($.53 \leq r \leq .58$), indicating that an important validity aspect is still retained. Noteworthy, these correlation coefficients were higher when inspecting the separate groups rather than the entire sample consisting of both groups ($r = .40$ /$r = .42$). Taken together, Study 1 showed that test scores of a prominent and widely used figural reasoning task type can be increased by a large margin when watching a video tutorial shortly before an assessment.

In Study 2, we illustrated the relevance of video tutorials further by evidencing the replicability of the findings obtained in Study 1. In a more diverse sample of higher sample size, we again showed higher figural matrices test scores in an experimental group watching a figural matrices video tutorial compared to a control group watching a video about nutrition ($d = 1.21$). Furthermore, and regarding validity aspects, results similar to Study 1 were also revealed when inspecting the correlations between the figural matrices test scores and the test scores of an intelligence test (EG: $r = .45$, CG: $r =$

.38; entire sample: *r* = .36). In short, in Study 2 we replicated and thus further strengthened the evidence on increased intelligence test scores after watching a video tutorial obtained in Study 1.

In Study 3, we shifted our focus towards number series as another popular (but numerical) reasoning task and designed the video tutorial focusing on commonalities instead of focusing specific rules. Additionally, we expanded our study design by adding a control group inspecting item examples instead of watching a video tutorial. The analyses revealed higher number series test scores in the experimental group watching the number series video tutorial than in 1.) the control group watching an irrelevant figural matrices video tutorial (*d* = 0.44) and 2.) the control group group inspecting item examples (*d* = 0.30). Thus, Study 3 showed that watching a number series video tutorial increases number series test scores by a medium margin when compared to a standard control group, and by a small margin when compared to another relevant method test-takers might use. Regarding validity aspects, the correlation coefficients between the number series test scores and the scores in another intelligence test revealed to be comparably high in all groups (.56 ≤ *r* ≤ .63). Furthermore, the correlation coefficients between the number series test scores and high school GPA were shown to be comparably high and in the expected range in all groups, as well (.25 ≤ *r* ≤ .38), indicating that important validity aspects are also retained regarding number series tasks. Taken together, in Study 3 we demonstrated that the effects of watching video tutorials do not pertain to only one specific reasoning content domain (numerical in Study 3, figural in Study 1 & 2) or focus when designing a corresponding video tutorial (focusing commonalities in Study 3, teaching specific rules in Study 1 & 2). Additionally, we further illustrated the practical relevance of watching video tutorials by demonstrating higher test scores when using such an approach compared to another approach test-takers might use during test preparation.

In Study 4, we investigated the effects of video tutorials on test-taking behavior. Specifically, and by using the data of the experimental group watching the number series tutorial and the control group watching the irrelevant figural matrices video tutorial of Study 3, we inspected the notes the participants took when working on the number series items. The results revealed that participants in the experimental group took notes for significantly more number series items than participants in the control group (*d* = 0.41). Furthermore, a higher number of items with notes was associated with higher number series test scores in both groups (EG: *r* = .66/CG: *r* = .75), and the effect of watching a number series video tutorial on number series test scores was mediated by the number of items with notes. Essentially, in Study 4 we demonstrated that watching a video tutorial stimulates changes in test-taking behavior, and indicated such a behavior as mechanism for the underlying number series test score increases after watching a number series video tutorial.

In summary, and by using two widely used reasoning tasks of different reasoning content domains, we demonstrated in this dissertation project that watching a short video tutorial before an intelligence assessment is sufficient to increase related test scores substantially. Moreover, we verified the replicability of these effects. We furthermore showed that different approaches to designing video tutorials are successful in revealing such increases, and also illustrated the relevance of video tutorials by demonstrating higher test scores when watching such a tutorial compared to inspecting item examples. Finally, we provided evidence indicating that important validity aspects are still retained, and demonstrated that watching a video tutorial promotes changes in test-taking behavior that can be assumed underlying the associated test score increases in number series tasks.

4.2 Implications and Future Research

In the following section, important implications that can be drawn from the evidence obtained in this dissertation project are discoursed. Moreover, the associated opportunities for future research are emphasized.

4.2.1 Effects on Test Scores

As most essential and primary goal of this dissertation project, we repeatedly demonstrated that test scores can be increased substantially by watching a video tutorial prior to the assessment. According to Cohen (1988), the revealed effects can be interpreted as large in size for the figural matrices video tutorial (Study 1: $d = 1.19$ and $d = 1.31$; Study 2: $d = 1.21$; Supplemental Analyses: $d = 1.73$) and small to medium in size for the number series video tutorial (Study 3: EG vs. $CG_{Examples}$: $d = 0.30$; EG vs. $CG_{Matrices}$: $d = 0.44$). These effects sizes were at least in the range of the magnitude typically reported in meta-analyses investigating practice and coaching effects, and concerning figural matrices even exceeded the typically reported parameters by a substantial margin (e.g., practice: $d = 0.40$, Hausknecht et al., 2007; $d = 0.42$, Kulik, Kulik, & Bangert, 1984; *SMCR* = 0.37, Scharfen et al., 2018; coaching: $d = 0.43$, Kulik, Bangert-Drowns, & Kulik, 1984; practice and coaching: $d = 0.64$, Hausknecht et al., 2007; see Chapter 2.2.2).

Notably, the effects on number series test scores after watching the number series video tutorial were shown to be smaller than for figural matrices test scores after watching the figural matrices video tutorial. Concerning the nature of these differences, several factors might contribute. First, the differences might be related to the nature of the tasks. As such, figural matrices can be solved by scanning items following a basic dichotomy of “rule is present/rule is not present”. After watching a figural matrices video tutorial teaching the rules, test-takers might follow a more systematic scanning process related to this dichotomy. For number series, however, the dichotomy might be trivialized by what counts as a “rule” (i.e., addition, subtraction, multiplication, and division could be considered “rules” that participants are probably already aware of without watching a tutorial). Furthermore, a quantitative element is introduced in number series tasks, going beyond the previously mentioned simple dichotomy of “rule is present/rule is not present”. Second, the differences in effect sizes might be related to the design of the video tutorials. For the figural matrices video tutorial, we focused on teaching the specific rules relevant for the subsequent assessment. Thus, the alignment between video tutorial and figural matrices task was quite high. For the number series video tutorial, however, we followed a more general approach focused on number series commonalities according to our illustration model as well as the shared processes typically involved in number series tasks (e.g., Holzman et al., 1982; 1983). Thereby, the alignment between video tutorial and later administered task was lower, potentially resulting in lower effect sizes. Third, the differences might be related to the videos presented in the corresponding control groups. In Study 1 and Study 2 regarding figural matrices, participants in the experimental group watched the relevant tutorial while participants in the control group watched a video about nutrition. In Study 3 regarding number series, however, participants in the experimental group watched the relevant video tutorial while participants in the control group watched the figural matrices video tutorial. Thus, as regarding number series the video presented to the control group was specifically designed to differ only in content from the video presented to the experimental group, the videos were much more similar. Nevertheless, it should be noted that in supplemental analyses participants still reached substantially higher figural matrices test scores (d = 1.73) after watching the figural matrices video tutorial than after watching the number series video tutorial. Although several reasons might play a role for this finding (see Chapter 3.4.5), it is indicated that the first and second suggested contributing factor probably play a more paramount role as to the nature of the differences in effect sizes than the third suggested contributing factor related to the videos presented to the control groups.

Importantly, this dissertation project was specifically designed to closely represent real life scenarios: in order to score as high as possible, test-takers might watch a video tutorial shortly before an assessment and subsequently increase their test scores in the portrayed task. By using control groups watching A.) a completely different video (as realized in Study 1 and Study 2) or B.) even a different video tutorial (as realized in Study 3), we simulated real-life scenarios where test-takers watch A.) either something completely irrelevant or B.) something very similar and potentially relevant, but ultimately not helpful for the task presented during the subsequent assessment. Furthermore, by adding the control group inspecting item examples in Study 3, we expanded on these scenarios by considering test-takers seeking and reflecting on item examples on their own. Practically, using such an approach might be particularly accessible for test-takers when preparing for an assessment, as a quick internet search (e.g., "figural matrices item examples intelligence" or "number series item examples intelligence) and sorting by "pictures" reveals plentiful material as basis for such an approach. Additionally, one has to keep in mind that the duration of the video tutorials was kept to less than 15 minutes. As simply watching a video tutorial using this time-frame is less time-consuming than retaking a whole test (representing practice effects; e.g., Scharfen 2018) or participating in test preparation programs (representing coaching effects; e.g., te Nijenhuis, 2001), it is suggested that by watching video tutorials test scores can not only be increased effectively, but also efficiently when compared to traditional approaches such as practice and/or coaching as well as similar more modern approaches such as specifically seeking for available item examples. In short, by demonstrating higher test scores after watching a video tutorial compared to control conditions closely representing real-life scenarios, the practical relevance of the test score increases demonstrated in this dissertation project is emphasized.

Taken together, the repeatedly demonstrated higher test scores after watching a task-relevant video tutorial allow for a conclusion with high fidelity: watching a video tutorial works as approach to increase test scores. It works with different design ideas in mind when constructing such a tutorial (e.g., focusing on teaching specific rules and focusing on commonalities related to items and processes), and for two widely used intelligence task types. As such, it seems plausible to assume that watching video tutorials can increase test scores for various types of rule-based intelligence tasks. Although future research might explore these possibilities more thoroughly (e.g., investigating tasks such as verbal analogies task to explore the content-domain not yet investigated in the BIS-model;

see chapter 2.1.1), different approaches for further investigations seem particularly critical (e.g., focusing on aspects related to practical relevance).

On a practically relevant note, test-takers typically do not know which test or task type to expect in an assessment. For example, regarding figural matrices, a large pool of rules can be used to construct figural matrices items (for an overview, see Preckel, 2003, pp. 64–71). Accordingly, the rules taught in a video tutorial might differ from the rules included in the administered figural matrices test. When the alignment between rules taught in a video tutorial and the rules included in the later administered figural matrices items decreases, lesser test score increases are to be expected. When the alignment is too low, test-takers might waste time unnecessarily looking for rules not included, ultimately even jeopardizing their assessment. Although we focused on particularly prominent rules in this dissertation project, we also focused on a perfect overlap between the rules taught in the video tutorial and the rules included in the later administered task, as previously mentioned. Thus, the effects of such a transfer to other tests is pending to be investigated empirically (e.g., by systematically varying the degree of alignment between rules taught in the video and the rules taught in the assessment; see Krautter et al., 2021, for first insight pointing toward the particular relevance of difficult figural matrices rules). Regarding number series, because we focused on the aforementioned commonalities on the basis of a large pool of items from published tests, the related impact should be expected to be less crucial than for figural matrices. Finally, video tutorials could be constructed focusing on critical information applicable to a wide variety of tasks (e.g., including elements related to test wiseness such as making efficient use of time, encouraging guessing for difficult items, or eliminating known incorrect options; see, for example, Millman et al., 1965; McPhail, 1981; Rogers & Yang, 1996). Investigating the effects of such a widely task-independent video tutorial and evaluating the effects on a wide variety of tasks seems interesting to extract a set of core elements that can potentially be added to any task-specific video tutorial.

In summary, the repeatedly demonstrated higher test scores after watching video tutorial allow for a conclusion with high fidelity: test-takers can increase their intelligence test scores substantially by watching a video tutorial. For further insight, researchers are encouraged to see the openly available video tutorials used in this dissertation project (see Appendix A).

4.2.2 Evidence Related to Validity Aspects

Regarding effects on validity aspects as our secondary goal, we demonstrated comparably high correlation coefficients between the figural matrices test scores and the intelligence sum scores in the experimental group watching the figural matrices tutorial and in the control group watching the video about nutrition in Study 1 ($.53 \leq r \leq .58$) as well as in Study 2 ($r = .45/.38$). These coefficients fall in line with the correlations between matrices test scores and intelligence test scores reported in prior studies ($r_{\text{DESIGMA-Advanced/LPS}} = .39$, Becker et al., 2014; $r_{\text{BOMAT/LPS-2}} = .45$, Kreuzpointner et al., 2013). Similarly, in Study 3 regarding number series, the correlation coefficients between the number series test scores and the intelligence sum scores revealed to be comparably high ($.56 \leq r \leq .63$), falling as well in line with previously reported evidence (e.g., correlations between number series of the I-S-T-2000 R and the IST-Screening: $r = .66$, Liepmann et al., 2012; number series of the WIT-2 and the Wonderlic intelligence-test: $r = .56$, Kersting et al., 2008). Additionally, concerning grades as second investigated validity aspect, correlation coefficients of comparable magnitude were also revealed when inspecting the correlations between number series test scores and GPA ($.25 \leq r \leq .38$). Like the forementioned correlations, these coefficients also fall in line with the correlations reported in earlier studies (e.g., number series and GPA: r = –.18/–.23, Kempf & Meder, 1993; number series and grades of subjects German/English/Mathematics: r = –.10/–.12/–.24, Liepmann et al., 2012).

Taken together, these repeatedly demonstrated comparable coefficients in the expected magnitude indicate that test validity does not seem to be aversely affected by watching a video tutorial. As the rank order of participants is suggested to be retained (and comparably high) in different groups, it is implied that meaningful predictions can be made regardless of whether all participants watch a relevant video tutorial, an irrelevant video, or inspect item examples. Such an assumption is consistent with the scarce empirical evidence on whether practice or coaching impede a test's validity (e.g., comparable predictive validity and no evidenced prediction bias for two randomly allocated groups coached/uncoached despite higher test scores in the coached group, Allalouf & Ben-Shakhar, 1998; no significant difference between the first and second administration of cognitive ability test scores when related to GPA, Lievens et al., 2005). Furthermore, this assumption is also in line with the vastly different approaches of including aspects of practice and coaching within the instructions included among published tests. For figural matrices, for example, this can be conceptualized by providing very limited (e.g., presenting only three item examples in the figural matrices subtest of the CFT 20-R without any explanation or further instruction besides selecting the choice that "fits best"; Weiß, 2006) or even very detailed information (e.g., administering a practice booklet in the BOMAT-advanced including information about items and logical principles, item examples illustrating these principles, practice items to work on that are explained afterwards; Hossiep et al., 1999). Such a diversity is also found in published number series tests. For instance, the number series subtest of the BIS-test (Jäger et al., 1997) does not include any item example or practice item and instead relies on the minimal instruction of "Which number replaces the question mark? Write the number on the line at the end of the number series!". At the other hand, in the number series subtest of the CFT 20-R (Weiß, 2006), four item examples are included. Besides providing more detailed information regarding the solution of one item example, further information about the test and even additional help is provided (e.g., indicating that for some items multiplication and/or division is required; encouraging to skip an item and come back later when an item seems too difficult; opportunity of using the item-sheet for calculations).

Although different approaches are shown to be viable, they rely on the critical aspect that all participants are tested under the same premise (i.e., equal exposure to test-relevant material and information among all participants). Specifically regarding video tutorials in this dissertation project, the violation of this aspect is empirically illustrated in Study 1 by the numerically lower correlation coefficients between the figural matrices sum scores and the intelligence sum scores in the heterogeneous entire sample (construction matrices: r = .40; distraction matrices: r = .42) when compared to the correlation coefficients in the samples of the homogeneous groups (EG – construction matrices: r = .53, EG – distraction matrices: r = .58; CG – construction matrices: : r = .56, CG – distraction matrices: r = .53) and the similar (but less accentuated) result pattern revealed in Study 2 (entire sample: r = .36; homogeneous samples – EG: r = .45; CG: r = .38). In Study 3, as depicted in Tables 10–13 albeit not specifically part of the research questions, the correlation between the number series sum score and the intelligence sum score in the entire sample (r = .58) revealed to be almost identical to the coefficients revealed in the homogeneous control groups ($CG_{Matrices}$: r = .57; $CG_{Examples}$: r = .56), but lower than the correlation in the homogeneous experimental group (EG: r = .63). Correspondingly, the higher coefficients in the experimental groups indicate that watching a video tutorial might even have a positive impact in an assessment when all test-takers are presented such information.

Such an assumption is in line with the results obtained in a recently published study (Levacher et al., 2021). Accordingly, the study revealed higher figural matrices test scores in an experimental group (*n* = 461) receiving one page of written information about the relevant figural matrices rules before working on a figural matrices test in comparison to a control group (*n* = 421) receiving no such information (*d* = 0.92; falling in line with the effects revealed when such information is presented in a video tutorial, as demonstrated). Related to validity aspects, and further corroborating the results obtained in this dissertation project, the correlations between the figural matrices test scores and the scores in the *Hamburg Natural Sciences Test* (Meyer et al., 2019) revealed to be meaningful in both groups (EG: $r = .28$; $p < .01$; CG: $r = .24$, $p < .01$), but did not differ significantly ($p = .48$). Importantly, the item properties of the figural matrices test were further suggested not to be influenced when receiving brief written information about the rules. Thus, the authors conclude that providing all test-takers with such information might be a useful approach to reduce a priori differences related to rule knowledge in figural matrices, at the very least increasing the fairness of the test. Correspondingly, presenting a video tutorial to all test-takers before an assessment as integral part of a test could thus also be a valuable asset to intelligence testing, as it seems exceedingly difficult to create a homogeneous testing situation with no participant possessing additional information about a test – especially keeping in mind the availability and accessibility of video tutorials elaborated on in section 2.2.3.

As the primary goal of this dissertation project was to examine mean effects on intelligence test scores by watching a video tutorial, we targeted gathering additional evidence on validity aspects as secondary goal. Accordingly, future research could build upon the gathered tentative (but mainly consistent) evidence by focusing validity aspects more specifically. For example, we focused on the scores in another intelligence test and grade point average as two substantially relevant variables. Particularly related to practical implications, investigating the impact of watching video tutorials using additional important criteria more closely related to a specific important outcome (e.g., scores in high stakes exams in the context of education; job performance and training success in the context of work) could further illustrate the related practical relevance. Relatedly, it seems especially relevant to keep in mind that, as elaborated on in chapter 2, the test score increases from watching a video tutorial are not assumed to reflect genuine enhancements in dispositional intellectual ability. Given the demonstrated potentially very substantial differences in tests scores between those test-takers watching a tutorial and those who do not, selecting a test-taker with lower dispositional cognitive ability on the basis of such elevated scores might lead to a severe overestimation of that person's cognitive ability. Ultimately, and keeping the large effect sizes revealed in this dissertation project in mind, such a false positive selection could result in severe negative consequences for the test-taker (e.g., being overburdened by the requirements) as well as the institution (e.g., underperforming candidate, wasted resources) in addition to negative consequences for otherwise suitable applicants that were ultimately not selected. As it was already shown that lower intelligence test scores increased the risk of school dropout (Pagani et al., 2017) and that individuals were more likely to move to a job of lower complexity when a job's complexity level exceeded that individual's general mental ability (Wilk & Sackett, 1996), such adverse consequences seem probable. Gathering longitudinal data for test-takers that are selected under different conditions and considering additional further variables could thus reveal particularly valuable insight into the impact of overestimating someone's ability as related to practice and coaching as well as specifically related to watching a video tutorial. Finallly, regarding validity aspects, future directions are best concluded by emphasizing the current state on how retesting impacts validity: "Very little empirical research has examined how retesting may alter validity inferences concerning important criteria such as job performance or success in graduate school." (Hausknecht et al., 2007, pp. 376–377). As there is still much to discover regarding practice even though retesting has been subject of empirical investigations for decades, there is even more to be discovered regarding video tutorials, specifically.

4.2.3 Effects on Test-Taking Behavior

As another secondary goal of this dissertation project, we investigated whether watching video tutorials affects aspects of test-taking behavior. Specifically, we demonstrated in Study 3 that test-takers took notes for more number series items after watching a number series video tutorial (d = 0.41), that these notes were related to the number series test scores (EG: r = .66; CG: r = .75), and that taking such notes served as the intermediary link between watching a number series video tutorial and the corresponding number series test score increases (indirect effect: 3.41, SE = 1.74; 95% CI: 0.26, 7.06). Taken together, the results indicated that watching a number series video tutorial stimulated test-takers to more frequently show manifest behavior that reflects the first of the internally organized processes illustrated in the number series video tutorial (i.e., relations detection; e.g., Holzman et al., 1983). As these results on test-taking behavior were shown for number series, the question arises whether similar results could be expected for other tasks after watching corresponding video tutorials. Approaching this question, the obtained results were in line with the general rationale that by writing down important elements, an effective external structure is created that can be used to allocate cognitive resources for finding the correct solution more efficiently. As such, the notated relations can be interpreted as manifest product of antecedent processes that allows for easier processing of subsequent processes. Although we focused on a highly relevant process-product for number series (i.e., notations concerning relations detection), the general rationale and thus the usefulness of notes might be similar for a wide variety of tasks – important information is externalized into an external storage, allowing for an easier solution of an item. Regarding figural matrices, for example, such a usefulness can be exemplified on the basis of Figure 6 as a classic figural matrices item (taken from Becker et al., 2016). In the illustrated item, the two rules *addition* and *rotation* are included, with the rule *addition* relating to the circles and the rule *rotation* relating to the arrows. In order to solve the item, a test-taker is required to deduce that in the empty field of the matrix, an arrow pointing to the left and four circles have to be complemented. If, for instance, a test-taker has correctly deduced that an arrow to the left has to be complemented, the arrow could be drawn into the solution field as a note. With the arrow notated (and thus one of the two rules externalized and "stored" in the solution field), the test-taker can more easily concentrate on the remaining rule referring to circles to finally induce alternative *A* as the correct solution to the item. Thereby, cognitive functions such as identifying and focusing attention on relevant parts as well as an easier transfer into working memory are considered (e.g., Weinstein & Mayer, 1986).

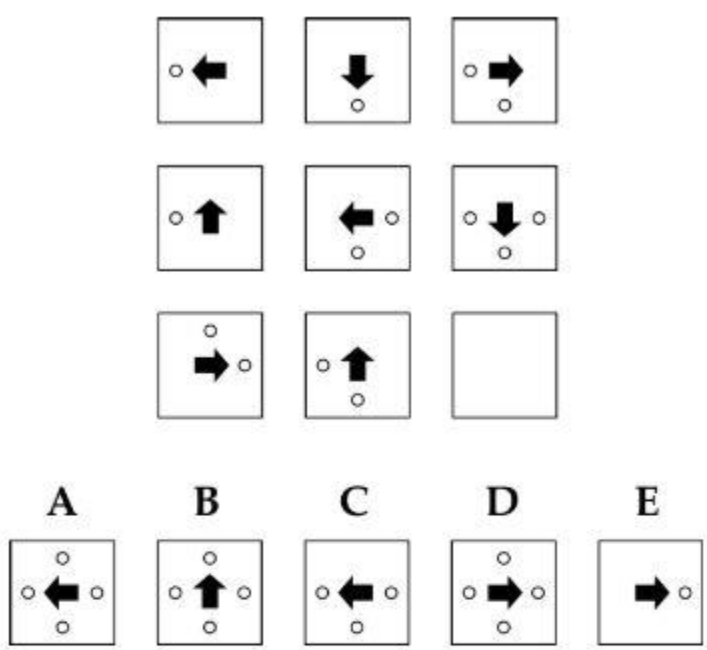

Figure 6: Classic Figural Matrices Item (taken from Becker et al., 2016)

When figural matrices items are presented in a distraction format, incorrect alternatives (such as *B*, *D*, and *E* in the presented example after notating the arrow) can additionally be excluded more easily after notating the rule-relevant elements, as well, ultimately further increasing the probability of solving an item correctly.

With video tutorials not only explaining the rules, but also illustrating how to solve items systematically, the benefits from using such an approach should become more apparent to the test-taker. Accordingly, the supplemental analyses inspecting the number of items with notes for figural matrices showed that significantly more test-takers took notes after watching the figural matrices video tutorial, making such an assumption plausible.

Nevertheless, notes should be considered a product of antecedent processes. As such, taking notes might make solving an item easier, but ultimately might not be the driving mechanism for the test score increases after watching video tutorials itself. For example, when investigating figural matrices in Study 1 and Study 2 of this dissertation project, the opportunity to take notes was not part of the investigation (besides – as in Study 1 the items were presented on a computer, an efficient use of such notes would not have been feasible). However, the large test score increases after watching a video tutorial were shown nonetheless. At the same time, and although for number series the mediation analysis showed the items with notes to mediate the effect on test scores after watching the number series video video tutorial, it seems implausible to assume that watching the video tutorial would not help with increasing test scores simply because participants cannot use notes. Rather, as both video tutorials focused on systematically explaining important elements, a more organized and systematical processing structure related to the tasks (e.g., scanning the items systematically according to the explained rules, following the relevant processes more closely) might generally play a substantial role that also manifests in increased note-taking behavior.

Future research could disentangle these possibilities more systematically. For example, video tutorials for different tasks could be presented before administering all tasks to all participants in an experimental setting where the use of notes is either prohibited or encouraged. When investigations are conducted using computerized versions of the tasks and inspecting the log-files (possibly using additional methods such as eye-tracking or thinking-aloud), additional valuable information related to potential other changes in test-taking behavior besides taking notes could be gathered, as well (e.g., time spent on items, focus of attention). Additionally, further conditions could be added to discern the effects of "rule knowledge" or "process knowledge" from the effects of systematic, illustrated explanations of these rules or processes in a video tutorial (e.g., by presenting a sheet with information instead of presenting a video tutorial). Thereby, the effects of watching video tutorials specifically beyond the effects of instructional or informational value could be differentiated. Finally, one should keep in mind that related to taking notes, no conclusions can be drawn from this dissertation project concerning the question whether test administrators should allow test-takers the use of notes. Future research could expand on this question by adding further variables to investigate validity aspects in order to consider the resulting practical implications adequately.

4.3 Critical Acknowledgements

In the following chapter, a critical reflection on potential limitations of the conducted studies is provided.

4.3.1 Design of the Project

In this subsection, limitations regarding the design of the dissertation project are addressed. Specifically, the adequacy of the samples is discussed as well as the (between-subjects) design of the studies.

Sample. This dissertation project was based on the samples of three data collections. In Study 1, the sample comprised N = 112 psychology students (EG: n = 56; CG: n = 56). In Study 2, the sample consinsted of N = 229 teacher-education students (EG: n = 114; CG: n = 115). In Study 3, the sample comprised N = 192 teacher-education students (EG: n = 63; $CG_{Matrices}$: n = 65; $CG_{Examples}$: n = 64). Finally, for Study 4, a subsample from Study 3 was used (i.e., N = 110 participants that took at least one note in every number series subtest; EG: n = 58; $CG_{Matrices}$: n = 52). All samples were selected with a priori considerations related to the experimental design, the implemented number of groups, experiences regarding participating students in former cohorts, aspects of accessibility, and estimated statistical power regarding statistical tests (e.g., correlations of medium to large size; one-tailed mean comparisons with about medium effect size). In consequence, the sample size of at least n = 50 participants in each study was a priori deemed sufficiently large for the main purposes of this dissertation project (i.e., show mean differences between the experimental and the control groups related to the tasks portrayed in the video tutorials; correlations between the test scores and other variables in each group). Accordingly, the sample sizes targeted in each study corresponded with, for example, a test-power of $1–\beta = .80$ for one-tailed *t*-test comparisons with $d = 0.5$, $\alpha = .05$, and a sample size of $n_1/n_2 = 50$ (see G*Power; Faul et al., 2007; 2009). As result, and as repeatedly shown across all studies, statistical power was sufficient to demonstrate the expected statistically significant mean score differences as primary goal of this dissertation project (in addition to the correlations related to validity aspects differing from zero within each group).

Nevertheless, the comparisons between the experimental group and the corresponding control groups related to validity aspects were not shown to be statistically significantly different in any study. Arguably, besides assuming comparable correlations for the groups, the sample sizes might have not been sufficiently large to detect these differences. In consequence, it is difficult to conclude whether there were indeed no differences between the correlations, or if there were smaller differences that the studies failed to detect. Indeed, for the sample sizes realized in this dissertation project, the range of the individual correlation coefficients as precise validity estimates was quite high (e.g., for a correlation of $r = .58$ based on 56 participants, a 95 % CI ranging from $r = .38$ to $r = .73$ would result). Thereby, such a conclusion seems principally possible. However, the same result pattern was consistently demonstrated in several analyses across all samples (i.e., correlation coefficients not differing statistically significantly; see results regarding the correlations for Studies 1 –3). Additionally, in Study 2 as a replication of Study 1, these analyses were based on a sample more than double in size when compared to the sample used in Study 1. Moreover, comparable validity coefficients fall in line with the scarce empirical evidence on validity aspects related to practice and coaching, as mentioned (Allalouf & Ben-Shakhar, 1998; Lievens et al., 2005). Thus, comparable validity coefficients seem plausible, nonetheless. Importantly, we targeted gathering data on the correlations with other important variables as additional evidence that might provide a first insight on whether watching a video tutorial affects validity aspects. Specifically focusing on statistically detecting such differences would require large sample sizes heavily exceeding the targeted scope of our studies. For example, keeping the correlations between the number series sum score and the intelligence test sum score in mind (EG/ $CG_{Matrices}$/$CG_{Examples}$: $r = .63/.57/.56$), to detect significant differences between two independent correlation coefficients of $r = .63$ and $r = .56$ (two-tailed, $\alpha = 0.5$, $1–\beta = .80$), a sample size of $N = 2650$ would have been necessary (i.e., $n = 1325$ in each group with two groups equal in sample size; see G*Power, Faul et al., 2009). As detecting these differences was not the focus of this dissertation project (and we rather focused on gathering first evidence future studies might build upon), future studies specifically targeting these research questions with adequate sample sizes should elaborate on this topic more conclusively.

As second aspect related to the samples, it could be argued that psychology students used in Study 1 were not the best sample as basis to inquire validity aspects for intelligence testing. Arguably, an intelligence-related range restriction to the right could have resulted. Furthermore, the results could have been impacted by some test-takers being already knowledgeable on intelligence testing, both aspects ultimately limiting the interpretation of the obtained evidence. However, several indications contradict such an assumption. First, the correlation-based coefficients in Study 1 revealed to be high in all groups ($.53 \leq r \leq .58$) and were at least in the range reported in prior studies (see also Becker et al., 2014). In case of severe variance reduction, substantially lower coefficients would have resulted. Second, the results were replicated in Study 2 using teacher-education students. Although slightly lower correlation coefficients resulted compared to Study 1 ($r = .45$/$r = .38$), these coefficients were still in line with similar coefficients reported in prior studies (e.g., $r_{\text{DESIGMA-Advanced/LPS}} = .39$, Becker et al., 2014; $r_{\text{BOMAT/LPS-2}} = .45$, Kreuzpointner et al., 2013; besides, as we only used three of the four subscales of the LPS-2 in Study 2, slightly lower correlations compared to Study 1 regarding the sum score were to be expected). Third, when comparing the standard deviations of the LPS-2 subtests in the experimental and control group in Study 1 and Study 2, no statistically significant differences are indicated (Study 1: LPS-2 sum score – $p = .25$; general knowledge – $p = .13$; numerical sequences – $p = .56$; mental rotation – $p = .38$; addition – $p = .89$. Study 2: LPS-2 sum score – $p = .38$; general knowledge – $p = .16$; numerical sequences – $p = .37$; mental rotation – $p = .91$). Finally, the correlations between the number series sum scores and the intelligence sum scores in Study 3 ($r = .63/.57/.56$) when relying on teacher-education students and the correlations between the figural matrices sum scores and the intelligence sum scores in Study 1 ($.53 \leq r \leq .58$) when relying on psychology students revealed to be of similar magnitude. Related to the question whether some participants were already knowledgeable in intelligence testing, the sample in Study 1 consisted of psychology students attending a course for new students in the first weeks of the first semester. Correspondingly, neither intelligence nor intelligence testing were part of the curriculum until the data collection of the study ended (the same held true for the samples of Study 2–4). Therefore, it seems unlikely that specific knowledge affected the results. Besides, as the participants were randomly assigned to the groups, individual participants possessing specific knowledge were expected to be evenly distributed among the groups (see below). To conclude, neither range restriction nor specific knowledge is indicated to severely limit the interpretation of the results.

Study Design. To investigate the effects of video tutorials, the studies in this dissertation project were realized as between-subjects designs to compare experimental groups watching a task-relevant video tutorial with control groups representing comparable conditions such as watching an irrelevant video or video tutorial or inspecting item examples. Accordingly, we realized such a design to specifically focus on internal validity while also considering similarities to real-life scenarios. However, as with every between subjects-design, a correct interpretation of the results is dependent on a successful randomizing procedure ensuring that the experimental group and the control groups were comparable before the experimental manipulation took place. Correspondingly, we conducted preliminary analyses to compare a priori group differences using the intelligence test data collected prior to the experimental manipulation. As indicated by these preliminary analyses (see sections 3.1.3, 3.2.3, 3.3.3, and 3.4.3), no statistically significant mean differences were revealed, indicating a successful randomizing procedure and thus comparable groups across all studies. Nevertheless, additionally presenting the tasks portrayed in the relevant video before the videos were presented could have allowed to draw additional conclusions as to who profits from watching such a video tutorial (possibly revealing compensation or magnification effects; e.g., Wenzel & Reinhard, 2019). First tentative analyses based on the data of Study 1 and Study 2 did not indicate such effects (e.g., creating a factor intelligence by selecting the participants with the 33 % highest and the 33 % lowest LPS-2 scores and conducting an ANOVA with the number of solved figural matrices as dependent variable and group/intelligence as independent variables revealed no significant interaction effects [Study 1: construction format/distraction format: p =.22/p =.27; Study 2: p = .68]). To further explore the possibility of compensation or magnification effects, and as number series were also included in the intelligence battery administered three months prior to administering the number series tasks in Study 3, we calculated the score differences between the first number series task presented after the experimental manipulation and the number series task included in the intelligence battery administered three months earlier. Afterwards, we calculated the Pearson correlations of these difference scores and the intelligence battery sum scores (excluding the number series subtest) in each group. The results were in line with the first tentative analyses based on Study 1 and Study 2: No statistically significant correlations were revealed for the experimental group watching the number series video tutorial (*r* = .00; *p* >.99) or the control group watching the figural matrices video tutorial (*r* = –.01, *p* = .96). However, a statistically significant correlation of moderate magnitude was revealed in the control group inspecting item examples (*r* = .29, *p* = .02). Taken together, neither compensation nor magnification effects were revealed concerning watching video tutorials, although further studies might be able to clarify this research question more conclusively (particularly building on the revealed correlation in the control group inspecting item examples). In summary, although future studies realizing different study designs seem valuable to build on the obtained evidence, the study design realized in this dissertation project seemed adequate to investigate our targeted research questions.

4.3.2 Instruments

In the following subsection, limitations regarding the instruments used in this dissertation project are addressed.

Figural matrices. One potential limitation refers to the figural matrices tasks used in Study 1 and Study 2. Specifically, instead of using Raven's Advanced Progressive Matrices (e.g., Raven et al., 1998) as hallmark and most widely known figural matrices test, we used the lesser known DESIGMA to investigate figural matrices. Nevertheless, using the DESIGMA instead of Raven's Matrices seemed favourable for several reasons. First, a number of critical arguments regarding Raven's Matrices can be found in the literature regarding, for instance, dimensionality (e.g., Bors & Vigneau, 2003: questionable single-factor model fit; DeShon et al., 1995: item subsets dependent on different processes; see also Gignac, 2015) and item context effects (Vigneau & Bors, 2005). Second, as Raven's Matrices are widely known, previously acquired specific test knowledge would have been more likely to influence the results (e.g., by participants in the control group being already familiar with the rules, thus reducing the differences between the experimental group and the control group effected by the video tutorial). By using a lesser known test, we minimized such an impact. Third, first evidence on rule-based training on the basis of Raven's Advanced Progressive Matrices was previously shown (Loesche et al., 2015). However, several aspects had to be critically considered when interpreting the results (e.g., varying effects across the studies [.14 ≤ *d* ≤ .81]; experimental and control groups differing regarding factors such as time, practice, and number of shown items; see chapter 2.2.3). By focusing on internal validity in our corresponding design and using the DESIGMA-Advanced, we considered these aspects while also showing that related effects are not limited to Raven's Matrices.

Finally, as some items were excluded in those previous studies and we used all items included in the test, we furthermore emphasized the relevance of the effects in light of real-life scenarios. In conclusion, the figural matrices of the DESIGMA-Advanced seemed adequate for the purpose of our studies. However, future studies might include several different figural matrices tasks to further investigate related aspects such as transfer (as mentioned).

Number Series. Similarly, concerning number series, it could be argued that using more popular and empirically validated pre-existing measures would have been more suitable than using newly constructed number items. Essentially, regarding our choice of number series tasks, we foremost relied on two popular measures that had been validated in our country for our targeted age group (i.e., I-S-T 2000R; WIT-2). As such, we established a meaningful and practically relevant basis for corresponding investigations based on these two widely used tasks. As we inspected the commonalities of in total 287 number series items found in various published tests (e.g., Heller & Perleth, 2000; Jäger et al., 1997; Kersting et al., 2008; Liepmann et al., 2007; Weiß, 2006) and further considered important theoretical aspects (e.g., Holzman et al., 1983; Kotovsky & Simon, 1973, Loe et al., 2018; Verguts et al., 2002) in order to establish our illustration model, we pursued the goal of creating new items following these commonalities in order to focus on number series as a task type (rather than an individual scale) and to enlarge our item-pool. Accordingly, the resulting 32 newly constructed number series items were targeted to be structurally very similar to the validated number series items of the I-S-T 2000 R and the WIT-2. Although an elaborate psychometric validation of the newly constructed items is pending future studies, the correlations of the newly constructed items with other variables (e.g., IST-verbal; IST-numerical; IST-figural; IST-sum score; I-S-T 2000R number series) in comparable magnitude to the correlations of the validated tasks with these variables indicate that our targeted approach was successful (e.g., regarding the entire sample of Study 3: WIT-2/Newly Constructed: $r = .22/.22$; $r = .47/.47$; $r = .35/.40$; $r = .50/.52$; $r = .78/.61$; correlation with each other: $r = .72$; see Tables 10–13). As such, insufficient psychometric properties are not indicated to underly the statistically insignificant subtest analyses specifically related to the newly constructed items in Study 4. In conclusion, although future studies might investigate the newly constructed number series items more elaborately or investigate additional validated number series tasks, the number series items chosen for the investigation seemed adequate for our targeted purpose.

Intelligence Tests (LPS-2, IST-Screening). Related to investigating validity aspects, using a complete intelligence test-battery (or a higher number of subtests) could have resulted in more precise validity estimates. Nevertheless, we favoured feasibility aspects related to the data collection (e.g., available time, group testing) in order to stay in accordance with focusing mean differences after watching video tutorials as primary goal and gathering evidence on validity aspects as secondary goal. The selected intelligence tests seemed particularly suitable considering not only these feasibility aspects, but also aspects related to relevance (i.e., proximity to *g*, validated subtests) and usability (i.e., being usable in the targeted sample of university students). Besides aligning well with elaborated models of intelligence (e.g., IST-Screening: figural, numerical, and verbal reasoning, fitting particularly well to the BIS-model [Jäger, 1982, 1984; Jäger et al. 1997]; LPS-2: different subtests reflecting important second-stratum factors of Carroll's three-stratum-model [Carroll, 1993, 2005]) and being administerable in a short timeframe, the tests were usable for our targeted sample of university students and have been sufficiently validated. For example, regarding the sum score of the IST-Screening that was also used in this dissertation project, the test manual reports favourable reliability coefficients (Cronbachs α = .87; Split-Half: r = .90; Test-Retest: r = .87; Parallel-Test: r = .88; see Liepmann et al., 2012, p. 17) as well as validity coefficients (e.g., correlations with the matrices of the CFT 20: r = .63; correlations with Raven matrices: r = .69; correlations with the knowledge dimension of the HAWIE-R: r = .46; see Liepmann et al., 2012, p. 24). Regarding the subtests we used of the LPS-2, the test manual similarly reports favorable reliability coefficients (general knowledge/numerical sequences/mental rotation/addition: Cronbachs α = .89/.75/.93/.88; Split-Half: r_{corr}= .91/.80/.96/.94; see Kreuzpointner et al., 2013, p. 50; see also p. 32) in addition to diverse validity coefficients (e.g., correlations with working memory tasks such as operation span: r = .32/.21/.10/.21, see p. 62; correlations with BOMAT-matrices of r = .20/.32/.32/.06, see p. 61; various correlations with six WIT-2 subtests, see p. 63). Thus, although a higher number of subtests and a greater range of diverse subtests would have been favourable to represent an indicator of a "good *g*" (Jensen & Weng, 1994; Ree & Earles, 1991; see also Reeve & Blacksmith, 2009; Lotz et al., 2016), the selected instruments were suitable as a proxy and in accordance with gathering evidence on validity aspects as secondary goal. However, future studies might focus on such a good *g* more specifically in addition to a highly increased sample size, opening further research possibilities such as investigating changes related to the factor structure whe watching video tutorials.

4.3.3 Methodological Aspects

In this subsection, limitations pertaining to methodological aspects are addressed.

Excluding Participants Inspecting Item Examples in Study 4. In Study 4 for investigating note-taking behavior, we excluded the group inspecting item examples from the analyses. Such an exclusion was based on the rationale that for Study 3, the control group inspecting item examples was specifically included to represent a second control group reflecting another approach test-takers might use before an assessment: dealing with number series on their own. In accordance with this rationale, these test-takers were instructed to freely switch between a page on the testing booklet with the number series items and another page depicting the solutions. Furthermore, test-takers in this group were encouraged within the instructions to use notes that might help with solving the items, and were also provided a dedicated space to write down such notes. As a result, these test-takers closely represented the aspired real-life scenario aimed at for Study 3. However, specifically related to note-taking for Study 4, several influences interacted in this group (e.g., dedicated space for notes, encouragemenet within the instructions, freely switching between items and solutions), resulting in substantial qualitative differences concerning note-taking and the related instructions in this group compared to the other two groups. In order to interpret the results of Study 4 in light of maximized similarity in every aspect but the relevant dimension (content: number series vs. figural matrices), we thus only focused on the experimental group watching the number series video tutorial and the control group watching the figural matrices video tutorial. However, future studies might disentangle aspects specifically related to note-taking when inspecting and working on item examples.

Participants Taking Notes. As second methodological aspect related to participants in Study 4, we only included participants that took at least one note in every number series subtest. Thereby, we ensured that test-takers were at least aware about the opportunity to take notes and were motivated enough to demonstrate such a behavior in a minimal amount. However, as taking notes was not required within the instructions, the reasons for not taking notes could have been very diverse (e.g., not understanding or not being aware taking notes was allowed, lack of motivation, or lack of necessity). As such, the results should be interpreted in light of these factors (e.g., when participants are aware of the opportunity and motivated enough to show such a behavior). Although the included number of participants was sufficiently high (86 % of the entire sample), future studies could replicate the results under different conditions to gain a deeper understanding into participants not taking any notes (who were not numerous enough to draw meaningful evidence from in Study 4).

Printing Error. Finally, as mentioned in Study 3, one item of the newly constructed items had to be excluded from the analyses due to a printing error (i.e., item 3: 2 15 28 41 54 67 **70** ?). Thus, it could be argued that test-takers could have been confused by such an occurrence, impacting the performance on subsequential items. However, it should be considered that the corresponding printing error was largely obvious (i.e., a simple “+13” rule was realized in the item). Thus, such an impact is not likely to substantially affect the results. Besides, as the last subtest, the previously administered two number series subtests would not have been impacted even if individual test-takers were influenced by such an occurrence. Finally, such individual test-takers would have been evenly distributed among the three groups by the randomizing procedure. Taken together, excluding the item from the analyses seemed sufficient to address the corresponding consequences.

4.4 Final Conclusions

In this dissertation project, it was aimed to investigate the effects of watching video tutorials on intelligence test scores.

In Study 1 and Study 2, on the basis of figural matrices as popular and widely used task type, large effects on test scores were revealed after watching a relevant video tutorial focusing on teaching the rules compared to watching an irrelevant video about nutrition. Additionally, the correlations with an intelligence test revealed to be substantial in both groups. When inspected separately in each condition, these correlations revealed to be higher than in the entire sample. In light of these findings, it is indicated that watching a video tutorial serves as effective and efficient approach to increase figural matrices test scores. Although it is suggested that important validity aspects are retained when all test-takers share the same amount of task-related knowledge, a benefit for the assessment procedure is indicated when test-takers can be kept homogenous in this regard.

Expanding beyond figural-matrices, the effects of watching video tutorials were further illustrated on the basis of number series as another popular and widely used task type in Study 3. Besides demonstrating higher number series test scores after watching a number series video tutorial focusing on number series commonalities compared to watching an irrelevant video tutorial and compared to inspecting item examples, the correlations between the number series test scores and and the scores in an intelligence test as well as the correlations between the number series test scores and grades revealed to be substantial. In conclusion, Study 3 suggested that watching a number series video tutorial is an effective and efficient approach to increase number series test scores, and more effective than simply inspecting item examples. Further falling in line with Study 1 and Study 2, it is indicated that the validity of number series tasks is not impeded – nevertheless, similarly to figural matrices, the benefits of keeping test-takers homogeneous in an assessment should be considered.

Shifting our focus towards changes in test-taking behavior in Study 4, we revealed a higher number of notes after watching a number series video tutorial compared to watching an irrelevant video tutorial, as well as substantial correlations between these notes and the number series test scores. Moreover, it was revealed that such a note-taking behavior mediated the associated number series test score increases. Thereby, we expanded beyond test score increases and validity aspects, providing insight from a different angle important for assessments as well as clarifying a mechanism assumed to underly the associated test score increases after watching a number series video tutorial.

In sum, the results obtained in this dissertation project emphasized watching video tutorials as an effective and efficient approach to increase test scores. By demonstrating up to large increases using two widely used task-types, we illustrated the potential impact of such an influence on intelligence testing. Correspondingly, we demonstrated that intelligence assessments benefit from test-takers being homogeneous in their task-related prior knowledge in order to maximize validity aspects. Taken together, by illustrating the subject matter from different angles, we provided extensive empirical insight into a highly relevant influence for intelligence assessment that is likely to be already involved in a plethora of testing situations.

References

Allalouf, A., & Ben-Shakhar, G. (1998). The effect of coaching on the predictive validity of scholastic aptitude tests. *Journal of Educational Measurement, 35*, 31–47. https://doi.org/10.1111/j.1745-3984.1998.tb00526.x

Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist, 36*, 1086–1093. https://doi.org/10.1037/0003-066X.36.10.1086

Arendasy, M. E., Sommer, M., Gutiérrez-Lobos, K., & Punter, J. F. (2016). Do individual differences in test preparation compromise the measurement fairness of admission tests? *Intelligence, 55*, 44–56. https://doi.org/10.1016/j.intell.2016.01.004

Becker, N., Preckel, F., Karbach, J., Raffel, N., & Spinath, F. M. (2014). Die Matrizenkonstruktionsaufgabe: Validierung eines distraktorfreien Aufgabenformats zur Vorgabe figuraler Matrizen [the construction task: Validation of a distractor-free item format for the presentation of figural matrices]. *Diagnostica, 61,* 22–33. https://doi.org/10.1026/0012-1924/a000111

Becker, N., Schmitz, F., Falk, A., Feldbrugge, J., Recktenwald, D., Wilhelm, O., Preckel, F., & Spinath, F. M. (2016). Preventing response elimination strategies improves the convergent validity of figural matrices. *Journal of Intelligence, 4*, 2. https://doi.org/10.3390/jintelligence4010002

Becker, N., & Spinath, F. (2014). *Design a Matrix–Advanced. Ein distraktorfreier Matrizentest zur Erfassung der allgemeinen Intelligenz [DESIGMA-Advanced]*. Göttingen, Germany: Hogrefe.

Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist, 52*, 1046–1050. https://doi.org/10.1037/0003-066X.52.10.1046

Bors, D. A., & Vigneau, F. (2003). The effect of practice on Raven's Advanced Progressive Matrices. *Learning and Individual Differences, 13,* 291–312. https://doi.org/10.1016/S1041-6080(03)00015-3

Cahan, S., & Cohen, N. (1989). Age versus Schooling Effects on Intelligence Development. *Child Development, 60,* 1239–1249. https://doi.org/10.2307/1130797

Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence, 38*, 424–432. https://doi.org/10.1016/j.intell.2010.04.005

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. https://doi.org/10.1037/0033-295X.97.3.404

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 69–76). New York, NY: Guilford Press.

Ceci, S. J. (1991). How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence. *Developmental Psychology, 27*, 703–722. https://doi.org/10.1037/0012-1649.27.5.703

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9780203771587

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence, 21,* 135-155. https://doi.org/10.1016/0160-2896(95)90023-3

DeVaney, T. A. (2009). Impact of video tutorials in an online educational statistics course. *Journal of Online Learning and Teaching, 5*, 600–608.

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

Estrada, E., Ferrer, E., Abad, F. J., Roman, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence, 50,* 93–99. https://doi.org/10.1016/j.intell.2015.02.004

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191. https://doi.org/10.1037/0033-2909.101.2.171

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The g beyond Spearman's g: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence, 44,* 1–10. https://doi.org/10.1016/j.intell.2014.01.009

Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence, 52*, 71–79. https://doi.org/10.1016/j.intell.2015.07.006

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.

Gygi, J. T., Hagmann-von Arx, P., Schweizer, F., & Grob, A. (2017). The predictive validity of four intelligence tests for school grades: a small sample longitudinal study. *Frontiers in Psychology, 8*, 375. https://doi.org/10.3389/fpsyg.2017.00375

Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in Systems Neuroscience, 8*. https://doi.org/10.3389/fnsys.2014.00034

Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92,* 373–385. https://doi.org/10.1037/0021-9010.92.2.373

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48,* 1–14. https://doi.org/10.1016/j.intell.2014.10.005

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional process analysis: A Regression-Based approach* (2nd ed.). New York: The Guilford Press.

He, Y., Swenson, S., & Lents, N. (2012). Online video tutorials increase learning of difficult concepts in an undergraduate analytical chemistry course. *Journal of chemical education, 89,* 1128–1132. https://doi.org/10.1021/ed200685p

Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT4-12+R) [Cognitive ability test for grade 4-12, revision].* Weinheim, Germany: Beltz.

Holzman, T. G., Pelligrino, J. W., & Glaser, R. (1982). Cognitive dimensions of numerical rule induction. *Journal of Educational Psychology, 74*, 360–373. https://doi.org/10.1037/0022-0663.74.3.360

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology, 75*, 603–618. https://doi.org/10.1037/0022-0663.75.4.603

Hossiep, R., Turck, D., & Hasella, M. (1999). *BOMAT-advanced – Bochumer Matrizentest [Bochum matrices test]*. Göttingen, Germany: Hogrefe.

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment, 15*, 3–18. https://doi.org/10.1111/j.1468-2389.2007.00363.x

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98. https://doi.org/10.1037/0033-2909.96.1.72

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340–362. https://doi.org/10.1016/0001-8791(86)90013-8

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells [Multimodal classification of intelligence performance: Experimentally controlled development of a descriptive intelligence structure model]. *Diagnostica, 28,* 195–226.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende modelle, neue Entwicklungen, Perspektiven [Intelligence structure research: competing models, new developments, perspectives]. *Psychologische Rundschau, 35,* 21–35.

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

Jensen, A. R. (1980). *Bias in mental testing*. New York, NY: Free Press.

Jensen, A. R. (1981). *Straight talk about mental tests*. London, GB: Methuen.

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

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

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32,* 95–107. https://doi.org/10.1016/S0160-2896(03)00062-X

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence, 40*, 123–138. https://doi.org/10.1016/j.intell.2012.01.009

Kempf, W., & Meder, G. (1993). Theoretische und empirische Untersuchungen zur Restriktivität der Klassischen Testtheorie [theoretical and empirical investigations concerning restrictivity of classical test theory]. *Diagnostica, 39,* 1–21.

Kersting, M., Althoff, K., & Jäger, A. O. (2008). *Wilde-Intelligenz-Test 2 (WIT-2) [Wilde-Intelligence-Test 2]*. Göttingen, Germany: Hogrefe.

Krautter, K., Lehmann, J., Kleinort, E., Koch, M., Spinath, F. M., & Becker, N. (2021). Test Preparation in Figural Matrices Tests: Focus on the Difficult Rules. *Frontiers in psychology, 12,* 962. https://doi.org/10.3389/fpsyg.2021.619440

Kriegbaum, K., Becker, N., & Spinath, B. (2018). The relative importance of intelligence and motivation as predictors of school achievement: A meta-analysis. *Educational Research Review, 25,* 120–148. https://doi.org/10.1016/j.edurev.2018.10.001

Kotovsky, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4,* 399–424. https://doi.org/ 10.1016/0010-0285(73)90020-0

Kreuzpointner, L., Lukesch, H., & Horn, W. (2013). *LPS-2. Leistungsprüfsystem 2 [Achievement test system 2]*. Göttingen, Germany: Hogrefe.

Kulik, J. A., Bangert-Drowns, R. L., & Kulik, C.-L. C. (1984). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95,* 179–188. https://doi.org/10.1037/0033-2909.95.2.179

Kulik, J. A., Kulik, C.-L. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21,* 435–447. https://doi.org/10.3102/00028312021002435

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

LeFevre, J. A., & Bisanz, J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. *Memory & Cognition, 14,* 287–298. https://doi.org/10.3758/BF03202506

Levacher, J., Koch, M., Hissbach, J., Spinath, F. M., & Becker, N. (2021). You Can Play the Game Without Knowing the Rules-But You’re Better off Knowing Them: The Influence of Rule Knowledge on Figural Matrices Tests. *European Journal of Psychological Assessment, 38,* 15–23. https://doi.org/10.1027/1015-5759/a000637

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *Intelligenz-Struktur-test 2000 R (I-S-T 2000 R) [Intelligence structure test 2000 R].* Göttingen, Germany: Hogrefe.

Liepmann, D., Beauducel, A., Brocke, B., & Nettelnstroth, W. (2012). *Intelligenz-Struktur-Test Screening (IST-Screening) [intelligence structure test Screening]*. Göttingen, Germany: Hogrefe.

Lievens, F., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58,* 981–1007. https://doi.org/10.1111/j.1744-6570.2005.00713.x

Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology, 92,* 1672–1682. https://doi.org/10.1037/0021-9010.92.6.1672

Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an automated number series item generator using linear logistic test models. *Journal of Intelligence, 6,* 20. https://doi.org/10.3390/jintelligence6020020

Loesche, P., Wiley, J., & Hasselhorn, M. (2015). How knowing the rules affects solving the Raven advanced progressive matrices test. *Intelligence, 48,* 58–75. https://doi.org/10.1016/j.intell.2014.10.004

Lohman, D. F. (2011). *Cognitive Abilities Test (Form 7) (CogAT).* Rolling Meadows, IL: Riverside.

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

Lynn, R. (2013). Who discovered the Flynn effect? A review of early studies of the secular increase of intelligence. *Intelligence, 41*, 765–769. https://doi.org/10.1016/j.intell.2013.03.008.

McPhail, I. P. (1981). Why teach test wiseness? *Journal of Reading, 25,* 32-38.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of “Far Transfer”: Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science, 11,* 512–534. https://doi.org/10.1177/1745691616635612

Meyer, H., Zimmermann, S., Hissbach, J., Klusmann, D., & Hampe, W. (2019). Selection and academic success of medical students in Hamburg, Germany. *BMC Medical Education, 19,* 1–15. https://doi.org/10.1186/s12909-018-1443-4

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25,* 707-726.

Neisser, U., Boodoo, G., Bouchard, T. J . 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

te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E. O. W. (2014). Are Headstart gains on the g factor? A meta-analysis. *Intelligence, 46,* 209–215. https://doi.org/10.1016/j.intell.2014.07.001

te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence, 41*, 802–807. https://doi.org/10.1016/j.intell.2013.03.001

te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g- loaded tests: No g. *Intelligence, 35,* 283–300. https://doi.org/10.1016/j.intell.2006.07.006

te Nijenhuis, J., Voskuijl, O. F., & Schijve, N. B. (2001). Practice and Coaching on IQ Tests: Quite a Lot of g. *International Journal of Selection and Assessment, 9,* 302–308. https://doi.org/10.1111/1468-2389.00182

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … & Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465,* 775–778. https://doi.org/10.1038/nature09042.

Pagani, L. S., Brière, F. N., & Janosz, M. (2017). Fluid reasoning skills at the high school transition predict subsequent dropout. *Intelligence, 62,* 48–53. https://doi.org/10.1016/j.intell.2017.02.006

Preckel, F. (2003). *Diagnostik intellektueller Hochbegabung [diagnostics of intellectual giftedness]*. Göttingen, Germany: Hogrefe.

Protzko, J., Aronson, J., & Blair, C. (2013). How to Make a Young Child Smarter: Evidence From the Database of Raising Intelligence. *Perspectives on Psychological Science, 8,* 25–40. https://doi.org/10.1177/1745691612462585

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence, 53*, 202–210. https://doi.org/10.1016/j.intell.2015.10.006

Protzko, J. (2016). Does the raising IQ-raising g distinction explain the fadeout effect? *Intelligence, 56,* 65–71. https://doi.org/10.1016/j.intell.2016.02.008

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales.* San Antonio, TX: Pearson.

Ree, M. J., & Earles, J. A. (1991). The stability of g across different methods of estimation. *Intelligence*, *15*, 271–278. https://doi.org/10.1016/0160-2896(91)90036-D

Reeve, C. L., & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence, 37,* 487–494. https://doi.org/10.1016/j.intell.2009.06.002

Rogers, W. T., & Yang, P. (1996). Test-wiseness: Its nature and application. *European Journal of Psychological Assessment, 12,* 247-259.

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.

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

Sauerland, M., Stegt, S., & Trost, G. (2008). *Test für Ausländische Studierende (TestAS) [Test for Academic Studies].* Gesellschaft für Akademische Studienvorbereitung und Testentwicklung e.V. (g.a.s.t.) c/o TestDaF-Institut.

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

Schmidt, F. L. (2009). Select on Intelligence. In E. A. Locke (Ed.), *Handbook of Principles of Organizational Behavior: Indispensable Knowledge for Evidence-Based management* (2nd ed., pp. 3–18). New York: Wiley.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin, 124,* 262. https://doi.org/10.1037/0033-2909.124.2.262

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

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock–Johnson IV Tests of Cognitive Abilities (WJ IV COG)*. Rolling Meadows, IL: Riverside.

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

Schneider, B., & Sparfeldt, J. R. (2021a). How to solve number series items: Can watching video tutorials increase test scores? *Intelligence*, *87*, 101547. https://doi.org/10.1016/j.intell.2021.101547

Schneider, B., & Sparfeldt, J. R. (2021b). How to get better: Taking notes mediates the effect of a video tutorial on number series. *Journal of Intelligence, 9, 55*. https://doi.org/10.3390/jintelligence9040055

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "Brain-Training" Programs Work? *Psychological Science in the Public Interest, Supplement, 17,* 103–186. https://doi.org/10.1177/1529100616661983.

Soares, D. L., Lemos, G. C., Primi, R., & Almeida, L. S. (2015). The relationship between intelligence and academic achievement throughout middle school: The role of students' prior academic performance. *Learning and Individual differences, 41,* 73–78. https://doi.org/10.1016/j.lindif.2015.02.005

Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence, 21,* 279–296. https://doi.org/10.1016/0160-2896(95)90018-7

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of IQ. *Merrill-Palmer Quarterly, 47,* 1–41. https://doi.org/10.1353/mpq.2001.0005

Valerius, S., & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer intelligenzfaktoren mit allgemeinen und spezifischen schulleistungen im nested-factor-modell. *Zeitschrift Für Padagogische Psychologie, 29,* 101–108. https://doi.org/10.1024/1010-0652/a000151

Verguts, T., Maris, E., & De Boeck, P. (2002). A dynamic model for rule induction tasks. *Journal of Mathematical Psychology, 46,* 455–485. https://doi.org/10.1006/ jmps.2001.1400

Vigneau, F., & Bors, D. A. (2005). Items in context: Assessing the dimensionality of Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement, 65,* 109–123. https://doi.org/10.1177/0013164404267286

Villado, A. J., Randall, J. G., & Zimmer, C. U. (2016). The effect of method characteristics on retest score gains and criterion-related validity. *Journal of Business and Psychology, 31,* 233–248. https://doi.org/10.1007/s10869-015-9408-7

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson Assessment.

Weinstein, C.E., & Mayer, R.E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 315–327). New York, NY: Collier.

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R) [Culture fair intelligence test – Revision with additional tests vocabulary and number series - revision].* Göttingen, Germany: Hogrefe.

Wenzel, K., & Reinhard, M. A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence, 77,* 101405. https://doi.org/10.1016/j.intell.2019.101405

Wilk, S. L., & Sackett, P. R. (1996). Longitudinal analysis of ability-job complexity fit and job change. *Personnel Psychology, 49,* 937–967. https://doi.org/10.1111/j.1744-6570.1996.tb02455.x

Wolf, K. (2015). Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung [Video tutorials and explanation videos as object, method, and goal of media and film education]. In A. Hartung, T. Ballhausen, C. Trültzsch-Wijnen, A. Barberi, and K. Kaiser-Müller (Eds.), *Filmbildung im Wandel* (pp. 121–131). Vienna: New Academic Press.

Wonderlic, E. F. (1992). *Wonderlic Personnel Test (WPT).* Libertyville, IL: Wonderlic and Associates, lnc.

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from g: The return of the Brand effect. *Intelligence, 43,* 27–34. https://doi.org/https://doi.org/10.1016/j.intell.2013.12.004

Appendices

Appendix A: Link to the Video Tutorials

The figural matrices video tutorial (as used in Study 3 as updated version of the video used in Study 1) as well as the number series video tutorial are available at:

https://osf.io/vwq5k/

Appendix B: Graphical Figural Matrices Illustrations

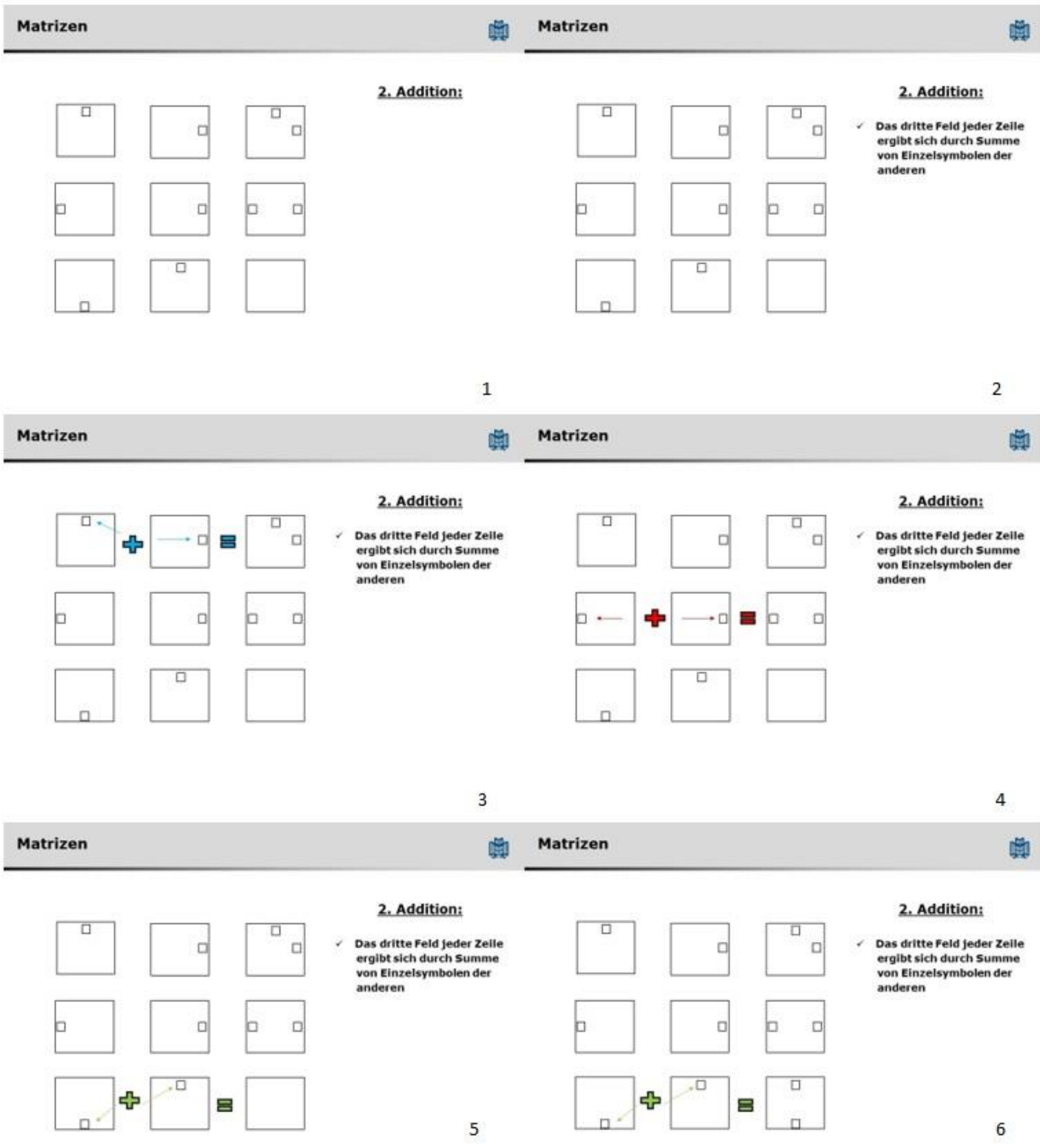

Figure 7 (Appendix B): Graphical Illustrations in the Figural Matrices Video Tutorial.

Appendix C: Graphical Number Series Illustrations

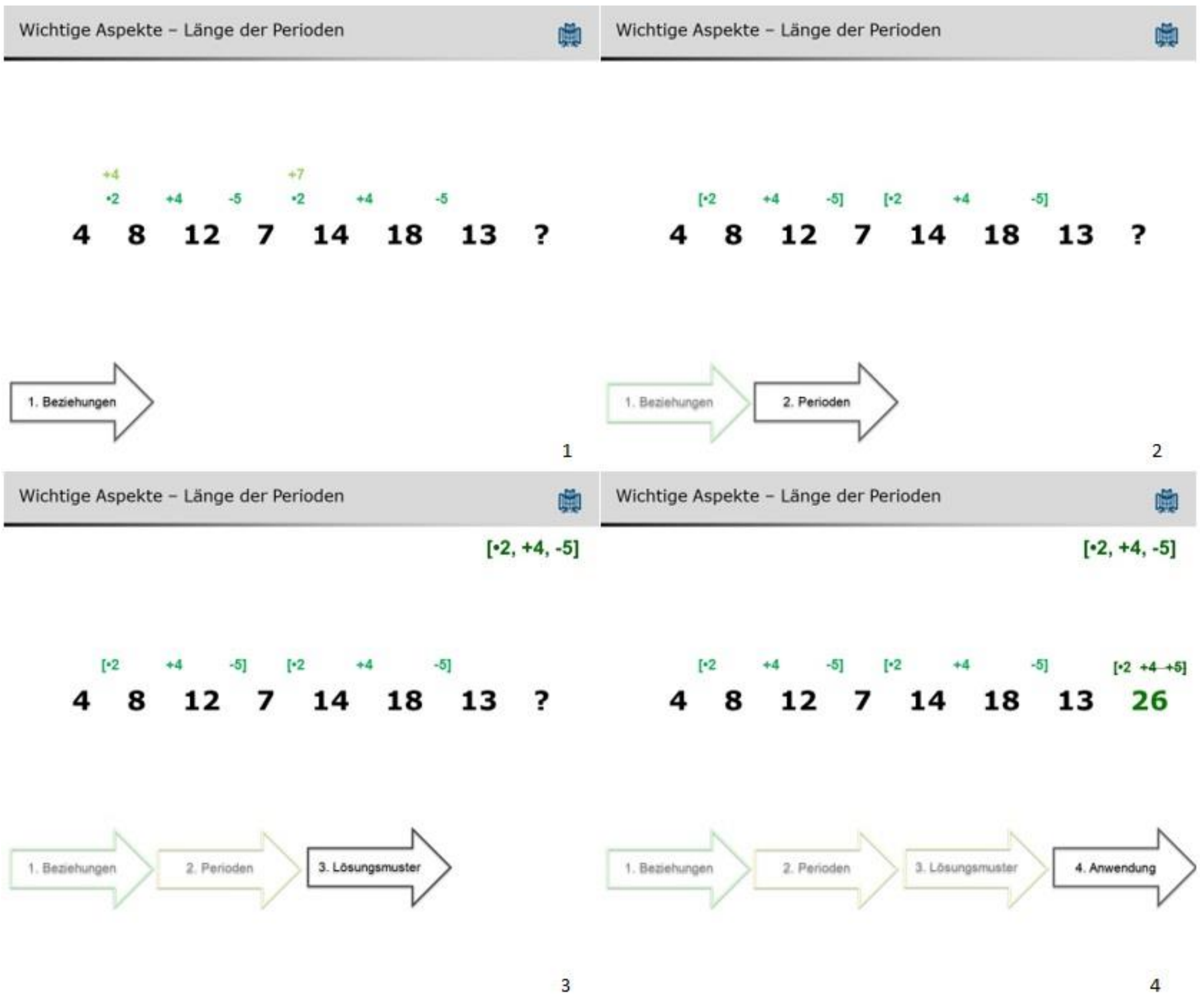

Figure 8 (Appendix C): Graphical Illustrations in the Number Series Video Tutorial.

Appendix D: Publications

Publication 1 (Study 1 & Study 2):

Schneider, B., Becker, N., Krieger, F., Spinath, F. M., & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, *82*, 101473. https://doi.org/10.1016/j.intell.2020.101473

Publication 2 (Study 3):

Schneider, B., & Sparfeldt, J. R. (2021a). How to solve number series items: Can watching video tutorials increase test scores? *Intelligence*, *87*, 101547. https://doi.org/10.1016/j.intell.2021.101547

Publication 3 (Study 4):

Schneider, B., & Sparfeldt, J. R. (2021b). How to get better: Taking notes mediates the effect of a video tutorial on number series. *Journal of Intelligence*, 9, 55. https://doi.org/10.3390/jintelligence9040055