DISS. ETH NO. 28791

WORKING MEMORY REQUIREMENTS FOR PROBLEM SOLVING IN ADVANCED SCHOOL MATHEMATICS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

SIMONA DAGUATI

MSc, ETH Zurich

born on *18.01.1994*

accepted on the recommendation of

Prof. Dr. Elsbeth Stern

Prof. Dr. Norbert Hungerbühler

Dr. Claudia von Bastian

2022

Acknowledgements

First and foremost I wish to thank my supervisor Elsbeth Stern for the opportunity to conduct this study and for her valuable advice and support during the different stages of the project. In addition, I would like to thank my co-supervisor Norbert Hungerbühler as well as my third-party reviewer Claudia von Bastian.

I am thankful to the math teachers participating in the TraM-Study for forwarding the information material about the current study to the selected students. I would like to thank Gioia Ehrensperger, Giovanni Ambrosioni, Vanessa Piccolo, Zelim Zanitti, Giulia Docimo, Raffael Schumann and Damian Striegel for their helpful assistance with the data collection process. I am grateful to Stefan Wehrli, Lea Imhof, Giordano Giannoccolo and Joris Stemmle for their support in the computer lab. Many thanks also to Alessandra Naldi Windler for coordinating the reimbursements of the study participants.

I wish to thank Norbert Hungerbühler, Armin Barth, Carina Heiss, Meike Akveld and Lorenz Stäheli for their valuable feedback on the Mathematics Speed Test and the Mathematics Power Test. I would also like to thank Ralph Schumacher. Special thanks to Claudia von Bastian for sharing her implementation of the working memory test battery and her test data as well as to my colleague Thomas Braas for sharing his composition of questionnaire items.

I am grateful to Mojgan Hosseinzadeh Dehaj for her valuable assistance with marking the Mathematics Power Test and reading in the data. Many thanks to my colleagues Peter Edelsbrunner and Christian Thurn for their helpful inputs and particularly for their advice with respect to the data analysis. In addition, I would also like to thank my other colleagues for fruitful discussions: Michal Berkowitz, Thomas Braas, Ursina Markwalder, Sonja Peterandl, Adrian Zwyssig, Daniel Weissmüller.

The research project was funded by the Swiss National Science Foundation, to which I am very thankful.

Finally, I wish to thank my husband Pascal Voser as well as other close family members and friends for their support.

Abstract

The current study focused on the interplay between general reasoning abilities, working memory functions and math achievement in order to investigate the role of working memory in advanced school mathematics. Based on mathematics test scores and scores on a measure of general reasoning ability (three subtests from the *Kognitiver Fähigkeits-Test (KFT)*, Heller & Perleth, 2000), a sample comprising high-achieving, over-achieving and under-achieving students at advanced placement schools in Switzerland (Swiss Gymnasium) was chosen ($N = 120$, $M_{age} = 16.3$ years). All study participants underwent a working memory test battery arranged by von Bastian and Oberauer (2013), which contains nine different tests targeting all facets of working memory described by the facet model (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Oberauer, Süß, Wilhelm, & Wittman, 2003; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). In addition, they completed a questionnaire on mathematical self-concept and interest in mathematics (the majority of the questionnaire items was utilised in *PISA 2000*, the items were taken from Gaspard et al., 2015, 2018; Mang et al., 2018; Trautwein, Lüdtke, Marsh, Köller, & Baumert, 2006). Furthermore, two mathematics tests were administered to the study participants, namely a speed test aiming at the knowledge representations of the students and a power test covering mathematical problem solving tasks. For the data analysis, group comparisons between the different groups of students and within the group of high-achievers were carried out. The findings regarding the group comparison between high-achieving and over-achieving students suggested that different factors, such as knowledge structures, mathematical self-concept and interest in mathematics could play a role in explaining over-achievement in mathematics. The results related to the group comparison between high-achievers and under-achievers identified a poorer mathematical self-concept and a lower performance of the working memory function Storage-Processing as potential root causes of under-achievement in mathematics. Moreover, they showed that in the sample of the current study, under-achievement in mathematics cannot be ascribed to a scarce interest in mathematics. Within the group of high-achievers, the motivational variables of mathematical self-concept and interest in mathematics seemed to be relevant for discriminating between the subgroup of high-achievers with top scores in the Mathematics Power Test and the subgroup of high-achievers with lower scores in the Mathematics Power Test.

Zusammenfassung

Der Fokus dieser Arbeit lag auf dem Zusammenspiel von schlussfolgerndem Denken, Arbeitsgedächtnisfunktionen und Mathematikleistungen, mit dem Ziel, die Rolle des Arbeitsgedächtnisses für Schulmathematik auf gymnasialer Stufe zu untersuchen. Basierend auf den Ergebnissen eines Mathematiktests und eines Tests zum schlussfolgernden Denken (drei Untertests aus dem *Kognitiven Fähigkeits-Test (KFT)*, Heller & Perleth, 2000) wurde eine Stichprobe ausgewählt, die sich aus High-Achievern, Over-Achievern und Under-Achievern an Schweizer Gymnasien zusammensetzt ($N = 120$, $M_{Alter} = 16.3$ Jahre). Allen Studienteilnehmenden wurde eine Testbatterie mit neun verschiedenen Arbeitsgedächtnistests vorgelegt, die von von Bastian and Oberauer (2013) zusammengestellt worden ist, und deren Tests insgesamt alle Arbeitsgedächtnisfacetten des Facettenmodells (Oberauer et al., 2000, 2003; Süß et al., 2002) abdecken. Zudem füllten die Studienteilnehmer:innen einen Fragebogen zu ihrem mathematischen Selbstkonzept sowie zu ihrem Interesse an der Mathematik aus (die meisten Items des Fragebogens wurden für *PISA 2000* verwendet, die Formulierungen der Items wurden von Gaspard et al., 2015, 2018; Mang et al., 2018; Trautwein et al., 2006, übernommen). Schliesslich absolvierten die Studienteilnehmenden zwei Mathematiktests: Einen Speed-Test, welcher auf die Wissensrepräsentationen der Schüler:innen abzielt, und einen Power-Test mit Probemlöse-Aufgaben. Für die Datenanalysen wurden Gruppenvergleiche zwischen den einzelnen Gruppen von Schüler:innen sowie innerhalb der Gruppe der High-Achievers durchgeführt. Die Resultate des Gruppenvergleichs zwischen den High-Achievern und den Over-Achievern gaben Hinweise darauf, dass verschiedene Faktoren, wie beispielsweise Wissensstrukturen, das mathematische Selbstkonzept und das Interesse an der Mathematik, Over-Achievement im Fach Mathematik möglicherweise erklären könnten. Die Ergebnisse in Bezug auf den Gruppenvergleich zwischen den High-Achievern und den Under-Achievern zeigten auf, dass ein geringeres mathematisches Selbstkonzept sowie eine schlechtere Leistung der Arbeitsgedächtnisfunktion Storage-Processing mögliche Ursachen für Under-Achievement im Fach Mathematik sein könnten. Zudem deuteten sie darauf hin, dass in der Stichprobe dieser Studie Under-Achievement im Fach Mathematik nicht auf mangelndes Interesse an der Mathematik zurückgeführt werden kann. Innerhalb der Gruppe der High-Achievers konnte festgestellt werden, dass die motivationalen Aspekte des mathematischen Selbstkonzepts und des Interesses an der Mathematik für die Differenzierung zwischen der Untergruppe mit den besten Leistungen im mathematischen Power-Test und der Untergruppe mit den schlechteren Leistungen im mathematischen Power-Test relevant zu sein scheinen.

Contents

List of Figures

List of Tables

1 Mathematics: Types of Knowledge and Individual Differences in Performance

Mathematics is a demanding subject to learn and challenging to teach. While some students manage to meet the requirements for learning mathematics, others experience difficulties throughout their educational path. The current study deals with the interplay between general reasoning abilities, working memory functions and mathematics performance. As a starting point, this chapter takes a closer look at selected aspects of mathematics learning. The first part of this chapter, Section 1.1, elaborates on different types of knowledge in mathematics. The second part of this chapter, Section 1.2, then focuses on individual differences in math achievement and on factors, which potentially explain these differences. Lastly, Section 1.3 concludes this chapter with some open research questions.

1.1 Different Types of Knowledge in Mathematics

The aim of this section is to discuss different types of knowledge in mathematics, namely conceptual knowledge as well as procedural knowledge. While Subsection 1.1.1 elaborates on conceptual knowledge in mathematics, Subsection 1.1.2 focuses on procedural knowledge in mathematics. Lastly, Subsection 1.1.3 is dedicated to the associations between conceptual knowledge and procedural knowledge in mathematics.

1.1.1 Conceptual Knowledge in Mathematics

According to Hiebert and Lefevre (1986), an important characteristic of conceptual knowledge is its richness of relations between different pieces of information. The authors describe it as a network of knowledge, where the associations between information elements are as relevant as the elements themselves. Hiebert and Lefevre (1986) explained that conceptual knowledge can be built by establishing relationships between information elements, either by linking two pieces of information, which were already memorised, or by connecting an existing knowledge element to a newly learned one. With respect to the first process, the authors emphasised that conceptual knowledge can be augmented by relating previously unassociated knowledge items to each other. They clarified that links can be constructed between smaller pieces of information as well as between larger pieces, which are organised as networks themselves. Regarding the second possibility of growth in conceptual knowledge, namely the creation of associations between present knowledge and new incoming information elements, Hiebert and Lefevre (1986) specified that the core of this process comprises the integration of these new elements into suitable knowledge structures, so that they become segments of existing knowledge networks.

In her review, Stern (2017) discussed how learning processes modify knowledge representations. In terms of conceptual knowledge, she explained that learning enables the restructuring of it, such as chunking different unassociated knowledge units into a few meaningful ones. Furthermore, Stern (2017) explicated two other processes, which can lead to an enhancement of conceptual knowledge, namely concept formation and deductive reasoning. She illustrated the first process with the example that a basic concept of *cow* can be formed by relating the concepts of *animal, produce, milk* to each other. For the second process, she explained that the circumstance that concepts are often embedded hierarchically, such as *animal* being a superordinate term for *cow* or *wombat*, allows for the construction of meaningful knowledge by inferential reasoning: "If the only thing a person knows about a wombat is that it is an animal, she can nonetheless infer that it needs food and oxygen" (Stern, 2017, p. 3). Another aspect emphasised by Stern (2017) is that there can be large individual differences in conceptual representations, as they depend on individual learning histories: She described that a veterinarian's concept of a cow for example is embedded into a broader network of concepts than the basic concept mentioned before, as a veterinarian has another level of expertise. Similarly, she pointed out that while novices in mathematics and experts in mathematics may utilise the same words, they may have totally different conceptual representations associated with them. In particular, Stern (2017) explained that while children predominantly rely on characteristic features when it comes to everyday concepts, educated adults often consider defining features. She illustrated it with the following example: While young children relate the concept of *island* to a warm place for vacation, adults associate it with a piece of land surrounded by water, which is too small to be categorised as a continent. As indicated in Stern (2017), this transformation from characteristic features to defining features is known as "conceptual change" (see also Carey, 2011).

1.1.2 Procedural Knowledge in Mathematics

According to the definition given by Hiebert and Lefevre (1986), procedural knowledge in mathematics consists of two different components: One component relates to the formal language of mathematics (i.e., its symbol representation system), while the other one refers to the algorithms for approaching math tasks. Regarding the first component, the authors explained that it covers a familiarity with mathematical symbols as well as an awareness of the corresponding syntactic rules. Hiebert and Lefevre (1986) illustrated this with the following example: Students, who have this part of procedural knowledge for arithmetic would be able to judge that the expression $6 - 2 =$ is syntactically correct, while the expression $6+ = 2$ is not (regardless of whether they would manage to solve the task or not). The authors pointed out that, in general, knowledge of mathematical symbols and the respective syntax does not automatically imply knowledge of their meanings. Rather, it only implies a familiarity with surface features. With respect to the second component of procedural knowledge in mathematics, Hiebert and Lefevre (1986) described that it consists of algorithms or procedures to tackle math tasks. The authors explained that a key characteristic of these procedures is that their execution follows a prespecified linear sequence and that this sequential nature is one of the most distinctive features distinguishing procedures from other types of knowledge. According to Hiebert and Lefevre (1986), procedural knowledge comprises algorithms for the manipulation of symbols as well as problem solving techniques, which do not directly work with symbols. Regarding the development of procedural knowledge, Stern (2017) explained that repetition and practice play a key role: The repeated practice of procedures enables to build strong connections between the individual steps of the respective procedure, which in turn allows for a more efficient execution of this procedure.

1.1.3 The Relations between Conceptual Knowledge and Procedural Knowledge in Mathematics

Hiebert and Lefevre (1986) emphasised the importance of relations between conceptual knowledge and procedural knowledge in mathematics: Students are not able to acquire full mathematical competence if there are knowledge shortcomings in either of the two types of knowledge or if they remain separate. More specifically, the authors explained that when the connections between concepts and procedures are missing, students may either have a good mathematical intuition without being able to solve math tasks, or they may produce results without understanding the respective steps. Hiebert and Lefevre (1986) described the theoretical advantages of establishing relationships between conceptual knowledge and procedural knowledge in mathematics as follows. On the one hand, if procedural knowledge is linked to conceptual knowledge, then the mathematical symbols obtain meaning and the procedural approaches can be memorised better and applied in a more effective way. On the other hand, the formal language and algorithms provided by procedural knowledge enhance conceptual knowledge and its applicability.

Rittle-Johnson and Alibali (1999) assessed the relationships between pupils' conceptual understanding of mathematical equivalence and the procedures, which they use for approaching equivalence tasks. Their sample consisted of grade four and grade five students, who underwent a test on conceptual and procedural knowledge of mathematical equivalence twice: Once prior to a corresponding math lesson and once after it. The focus of this instruction either lay on the concept of mathematical equivalence or on the procedure to solve equivalence tasks correctly. On the one hand, Rittle-Johnson and Alibali (1999) observed that the conceptual instruction improved students' conceptual understanding and supported the development and transfer of a correct procedural approach. On the other hand, the authors reported that the procedural instruction enhanced students' conceptual understanding and it led to the acquisition of the instructed procedure, while the transfer of it was limited. According to Rittle-Johnson and Alibali (1999), these results illustrate the causal and bidirectional relationships between conceptual knowledge and procedural knowledge in mathematics. In particular, the authors pointed out that the relation between conceptual knowledge and procedural knowledge appears to evolve in an iterative way: Gains in one knowledge type can promote increases in the other knowledge type, which in turn may support further improvements in the first knowledge type. At the same time, Rittle-Johnson and Alibali (1999) explained that the strengths of the mutual influences may not be symmetrical: Their findings indicate that conceptual knowledge may influence procedural knowledge to a greater extent than vice versa. In a further experimental study, Rittle-Johnson, Siegler, and Alibali (2001) addressed the hypothesis that conceptual knowledge and procedural knowledge in mathematics develop iteratively again. Their sample consisted of students attending grade 5 or 6, who were learning about decimal fractions. The authors reported that student's initial conceptual knowledge was a predictor of improvements in procedural knowledge, and these gains in turn predicted enhancements in conceptual knowledge. Rittle-Johnson et al. (2001) therefore concluded that conceptual knowledge and procedural knowledge in mathematics evolve in an iterative manner. Schneider and Stern (2010) also investigated the conceptual knowledge and procedural knowledge regarding decimal fractions in grade five and six students, and they assessed both types of knowledge with four common hypothetical measures each. Their results indicated low convergent validities of the measures under consideration. Furthermore, Schneider and Stern (2010) reported that measures, which were designed to measure the same type of knowledge, appeared to be inhomogeneous.

1.2 Individual Differences in Mathematics Performance

This section deals with individual differences in mathematics performance as well as with potential explanatory factors, to which these differences can be traced back to. While Subsection 1.2.1 explores the associations between general reasoning abilities and math achievement, Subsection 1.2.2 comments on the role of classroom environments for mathematics learning. In addition, Subsection 1.2.3 discusses the influence of motivational and emotional aspects on math performance. Lastly, the focus of Subsection 1.2.4 lies on over-achievement and under-achievement in mathematics.

1.2.1 The Relations between General Reasoning Abilities and Math Performance

As explained in Stern (2017), a characteristic of mathematics is that it involves abstract concepts that are mainly made up of defining features and which are related to other abstract concepts and procedures. Intelligence is therefore expected to play a key role for mathematics learning, as abstract thinking, the comprehension of complex ideas, reasoning and problem solving abilities are, among other abilities, assumed to be core components of intelligence (Gottfredson, 1997). The empirical studies described in the following paragraphs document the relationships between general reasoning abilities and math achievement.

In their longitudinal study, Deary, Strand, Smith, and Fernandes (2007) explored the relations between intelligence, measured at eleven years of age, and scholastic achievement in different subjects at sixteen years of age. The authors found that general intelligence correlated positively with all individual subject scores, and that the corresponding effect sizes were medium to large. For the subjects in the science group, Mathematics showed the highest correlation, while among the arts and humanities subjects, English yielded the highest correlation. More specifically, Deary et al. (2007) reported that general intelligence explained 58.6% of the variance in Mathematics and 48% of the variance in English. Mayes, Calhoun, Bixler, and Zimmerman (2009) studied IQ as well as various neuropsychological factors as predictors of scholastic achievement in primary school pupils. Among other results, their regression analyses revealed that achievement test scores could be predicted by IQ for both, word reading and math computation. In fact, IQ appeared to be the best single predictor of achievement among the explanatory variables included in their study. Mayes et al. (2009) reported that IQ accounted for 35% of the variance in word reading scores and for 22% of the variance in math computation scores.

The descriptions of studies collected in this paragraph illustrate the relations between reasoning abilities and math performance for different age groups. At the primary school level, Männamaa, Kikas, Peets, and Palu (2012) assessed the relations between several cognitive abilities and facets of math skills in third graders using a cross-sectional study design. Among other results, their data analysis in terms of structural equation models indicated a significant association between verbal reasoning and mathematical problem solving. At the high school level, Moenikia and Zahed-Babelan (2010) analysed the relation between IQ (among other explanatory variables) and math performance. Based on correlation and multiple regression analyses, they observed that IQ significantly predicted math achievement. Similarly, Tikhomirova et al. (2016) investigated the association between non-verbal intelligence and math performance in a sample consisting of high school students. Their correlation and regression analyses revealed that math achievement could be predicted by non-verbal intelligence. At the undergraduate level, Berkowitz and Stern (2018) analysed students' performance among students in the fields of mechanical engineering and math-physics during their first year of studies. Based on structural equation models, they found that numerical, verbal and general reasoning abilities were the best predictors for mathematics and physics achievements among the investigated explanatory variables. The authors therefore pointed out that individual differences in reasoning abilities appear to contribute to the prediction of math performance even within groups of highly competent students.

The samples of the following studies comprised various age groups for the investigation of the influence of reasoning abilities on math performance. Among other relations, Floyd, Evans, and McGrew (2003) assessed the predictive power of fluid reasoning on math calculation skills as well as on math reasoning. The representative sample for their study comprised fourteen different age groups with an age range from six to nineteen years. To assess the data, Floyd et al. (2003) carried out multiple regression analyses, which revealed moderate relationships between fluid reasoning and the mathematics measures. Based on structural equation models, Taub, Keith, Floyd, and McGrew (2008) analysed direct as well as indirect effects of general intelligence and other cognitive abilities on math achievement in study participants with an age range from five to nineteen years. The authors reported significant direct effects of fluid reasoning and of crystallised intelligence on math performance, while they observed indirect effects of general intelligence on it. Cormier, Bulut, McGrew, and Singh (2017) explored the relationships between different cognitive abilities and aspects of math performance in a large sample of children and adolescents between six and nineteen years of age. Among other results, the authors found a consistent association between fluid reasoning and math performance across the school years. In particular, they reported that fluid reasoning was the strongest predictor of math calculation skills as well as of math problem solving throughout all age levels in their study.

With respect to longitudinal developments, Primi, Ferrão, and Almeida (2010) studied the relationship between fluid intelligence and inter-individual differences in intra-individual growth patterns in math performance over a time span of two years. Their study participants, who were between eleven and fourteen years old, were tested on a math test at the start and at the end of grade seven and eight respectively. Moreover, they were administered a reasoning test. According to Primi et al. (2010), their results are in line with the finding that fluid intelligence and math performance are strongly related when measured contemporaneously. Furthermore, they observed that compared to students with a lower fluid intelligence, students with a higher fluid intelligence not only showed higher initial math scores, but also a steeper increase in math scores during the two year period. Overall, the authors underlined the importance of fluid intelligence for learning in mathematics. In their longitudinal study, Green, Bunge, Chiongbian, Barrow, and Ferrer (2017) examined the role of fluid reasoning for the acquisition of math skills. Based on structural equation models, the authors found that fluid reasoning significantly predicted future math performance (measured after approximately one and a half years).

To summarise, the studies described above illustrated the central role of general reasoning abilities for math achievement at various age levels. At the same time, there are studies showing that other factors may influence math performance as well - the subsequent subsections focus on some of these studies.

1.2.2 The Influence of Classroom Environments for Mathematics Learning

As demonstrated by Blankson and Blair (2016), the quality of classroom environments plays a role in giving students the possibility to invest their intelligence in mathematics learning: The authors explored the relations between different cognitive variables and math performance as well as the impact of classroom quality on these relationships among preschool pupils. Their results revealed that classroom quality moderated the effects of fluid intelligence and crystallised intelligence on math achievement. Regarding fluid intelligence, Blankson and Blair (2016) reported that its association with math performance became visible in higher quality classroom environments. On the other hand, the authors explained that children with higher levels of fluid intelligence may not be able to unfold their potential if they are placed in classrooms, which are neither well organised nor supportive. Similarly, Blankson and Blair (2016) observed that the benefits of crystallised intelligence were highest in classrooms of higher quality.

The following studies document the effect of perceived classroom environment on math achievement. In her study, LaRocque (2008) investigated potential effects of students' perceptions with respect to their classroom environment on math and reading performance in a sample of elementary school pupils. Based on correlation analyses and multivariate analysis of variance approaches, the author reported that there was a significant association between perceptions of the classroom environment and math as well as reading achievement. Tosto, Asbury, Mazzocco, Petrill, and Kovas (2016) analysed the influence of perceived classroom environment as well as of math self-efficacy, math interest and academic self-concept on math test performance among students of sixteen years of age. The authors observed that the correlations between math achievement and the three variables of math self-efficacy, math interest and academic self-concept were higher than the correlation between math achievement and perceived classroom environment. At the same time, their data analyses suggested that these three variables mediated the relation between perceived classroom environment and math achievement, i.e. there appeared to be an indirect effect of perceived classroom environment on math achievement. According to Tosto et al. (2016), a possible explanation for this finding could be that enhancing the perceptions of the classroom environment may impact math performance, via math self-efficacy, by improving math interest and academic self-concept.

In summary, the research findings from the studies described above showed that classroom environments affect math achievement.

1.2.3 The Impact of Motivational and Emotional Factors on Math Achievement

Several studies assessed the influence of motivational and emotional variables on math performance for different age groups. This paragraph describes studies, which focused on preschool and elementary school children. Dobbs, Doctoroff, Fisher, and Arnold (2006) analysed different socio-emotional correlates of math skills in preschool children using a cross-sectional study design. On the one hand, the authors found that the socio-emotional factors of initiative, self-control and attachment were all positively associated with mathematics skills. On the other hand, they observed that withdrawal, social problems and attention problems were related to low mathematics skills. In their longitudinal study, Mercader, Presentación, Siegenthaler, Molinero, and Miranda (2017) examined the predictive power of motivational factors, measured in preschool, on math performance in grade two. Among other results, the authors observed that student's self-competence was the best explanatory variable for later math achievement. Furthermore, they reported that persistence and attitude toward learning additionally contributed to the prediction of math performance. Among other relations, García, Rodríguez, Betts, Areces, and González-Castro (2016) studied the relationship between affective-motivational factors and math performance in upper elementary school pupils. Based on multiple linear regression analyses, the authors reported a positive association between mathematics enjoyment and math achievement. Furthermore, they observed mean differences with respect to the affective-motivational variables between the group of pupils with high math performance and the group of pupils with low math performance.

At the secondary school level, Singh, Granville, and Dika (2002) studied the influence of motivation, attitude, and scholastic engagement on performance in math and science among students attending grade eight. Based on structural equation models, the authors reported that these explanatory variables accounted for a significant portion of variance in math performance. The strongest impact on math achievement was observed for time spent on scholastic homework. Suárez-Álvarez, Fernández-Alonso, and Muñiz (2014) assessed the associations between conative variables and academic achievement in mathematics and sciences among secondary school students. The authors found that academic self-concept, motivation, and academic expectations were all significantly correlated with academic performance in mathematics and sciences, while academic self-concept emerged as the explanatory variable with the highest predictive power. Lauermann, Meißner, and Steinmayr (2020) analysed the relative contributions of intelligence and ability self-concept to the prediction of scholastic achievement in grade eight students based on multilevel structural equation models. Among other findings, the authors reported that intelligence appeared to be the best unique predictor of standardised math test performance, while students' math grades were best predicted by their ability self-concept. Furthermore, they observed that intelligence accounted for more variance in math achievement than in verbal performance.

Regarding high school students, León, Núñez, and Liew (2015) hypothesised different direct and indirect effects of motivation, autonomy, and self-regulated learning on math performance. Among other results, their structural equation models suggested the following path: If students perceive their schoolwork as being interesting and target-oriented, and if they experience a supportive learning environment, they are likely to be autonomously motivated to take part in self-regulated learning. In turn, self-regulation of effort is reflected in improved math performance. In their longitudinal study, Froiland and Davison (2016) assessed the interplay between expectations and motivational aspects in their relation to math performance among high school students. Their structural equation models indicated that parent expectations, student expectations, and peer interest were predictors of intrinsic math motivation in grade nine, which in turn predicted math performance in grade eleven. In particular, the authors observed that parent expectations predicted intrinsic math motivation and math achievement more strongly than student expectations. Among other factors, Grigg, Perera, McIlveen, and Svetleff (2018) studied the role of math self-efficacy for math achievement among adolescents. Their results indicated that math self-efficacy significantly and positively predicted math performance over and above prior math achievement. The authors reported that the effect of math self-efficacy on math grades was stronger than its effect on standardised test scores. Parhiala et al. (2018) investigated different profiles in terms of scholastic motivation and emotional well-being and assessed their associations with reading and mathematics skills among adolescents of $15 - 16$ years of age. The authors analysed their data by means of latent profile analyses, which showed that low scholastic motivation was related to poor mathematics and reading achievement, while a link between low emotional well-being and low mathematics and reading scores could only be observed together with low scholastic motivation.

The focus of the following studies lay on the contributions of motivational factors and of intelligence to the prediction of performance in adolescents. Steinmayr and Spinath (2009) analysed to which degree motivational variables account for the prediction of scholastic achievement independently from intelligence. Their sample consisted of adolescents attending grade eleven and twelve, who self-reported on different motivational aspects. The authors analysed their data by means of hierarchical regression and relative weights analyses, which revealed that different motivational variables incrementally added to the prediction of scholastic achievement beyond intelligence. With respect to math attainment, Steinmayr and Spinath (2009) even found that ability self-concept accounted for more unique variance than intelligence. Furthermore, the authors reported that when prior achievement was controlled for in the prediction of math performance, motivation added to the prediction, while intelligence did not. Altogether, Steinmayr and Spinath (2009) pointed out that motivation plays an important role for scholastic attainment. Using a similar approach, Kriegbaum, Jansen, and Spinath (2015) focused on the relative contributions of different motivational aspects to the prediction of math performance in students of fifteen years of age. In addition to a standardised math achievement test, their study participants filled in self-reports covering motivational factors. The authors investigated their data using structural equation models, which indicated that all motivational variables included in the study made an incremental contribution in explaining math achievement beyond intelligence. While the comparison with respect to the predictive power of the different variables revealed that intelligence accounted for the largest percentage of variance in math performance, self-efficacy appeared to be the strongest motivational predictor variable. Overall, Kriegbaum et al. (2015) explained that their results support the assumption of motivation being an important factor for the prediction of math achievement. In their study, Lotz, Schneider, and Sparfeldt (2018) examined the differential contributions of intelligence, math self-concept and math interest in predicting math grades and achievement on a math competence test among high school students. They observed that when all three explanatory variables were included in the statistical model, intelligence emerged as the strongest predictor of performance in the math competence test, while math self-concept appeared to be the strongest predictor of math grades. With respect to the individual increments, the authors reported that intelligence explained unique variance beyond the motivational factors in both, the math competence test and math grades. Self-concept accounted for a considerable increment on math grades, but not on the math competence test, while math interest did not exhibit any unique effects. All in all, Lotz et al. (2018) concluded that intelligence plays a more important role for scholastic competence tests, while self-concept seems to be a more important factor for grades.

Summing up, various results from the studies described above illustrated that motivational aspects influence math achievement throughout the educational path. In particular, some research findings suggested that motivational variables incrementally contribute to the prediction of math performance beyond intelligence in adolescents (e.g., Kriegbaum et al., 2015).

1.2.4 Discrepancies between Expected and Actual Math Performance: Over-Achievement and Under-Achievement in Mathematics

As discussed in Subsection 1.2.1, Deary et al. (2007) found that 58.6% of the variance in Mathematics scores could be explained by general intelligence, which is a considerable amount of explained variance. At the same time, the proportion of variance in Mathematics scores, which cannot be explained by general intelligence, indicates that for some students, discrepancies between their expected math performance (as predicted by their general intelligence) and their actual math performance might be present. As explicated in Krouse and Krouse (1981), academic under-achievement has traditionally been viewed as the discrepancy between the predicted performance, which is typically based on psychological tests such as IQ tests, and the actual performance of a student, who consistently scores below the expected level of performance. On the other hand, students exhibiting higher actual achievement than their predicted achievement are viewed as over-achievers (e.g., Castejón, Gilar, Veas, & Miñano, 2016).

Sepie and Keeling (1978) investigated a sample consisting of over-achieving, achieving and under-achieving students in mathematics at the age of eleven and twelve years. Their focus lay on examining the three groups with respect to measures of math anxiety, test anxiety, and general anxiety. For the group comparisons, the authors carried out an analysis of variance, which showed that the group of under-achievers could be distinguished from the groups of achievers and over-achievers more clearly in terms of their math anxiety compared to either their test anxiety or their general anxiety. Sepie and Keeling (1978) therefore concluded that measures of math anxiety appear to capture facets of anxiety related to math under-achievement, which are neither covered by test anxiety measures nor by measures of general anxiety. Castejón et al. (2016) chose a comparable sample composition to study three groups of students, who attended their first or second year at high school: Over-achievers, normal-achievers (i.e., students demonstrating an expected level of performance) and under-achievers. The assignment of the study participants to the three groups was based on an IQ test as a predictor of achievement and a measure of achievement, which was composed of different school grades. Among other findings, Castejón et al. (2016) reported that the higher academic performance in over-achievers could be traced back to an extensive use of learning strategies. In particular, their results indicated that over-achieving students showed a significantly higher utilisation of learning strategies compared to the under-achieving students. Furthermore, Castejón et al. (2016) observed that under-achievers exhibited a lower academic self-concept than normal-achievers and over-achievers.

The following studies investigated the influence of different variables on under-achievement in mathematics. Boehnke (2008) assessed the impact of high peer pressure on math achievement among middle-school students of fourteen years of age. In particular, the assumption underlying his study was that students with high math abilities may under-achieve in mathematics to circumvent social exclusion (for example, to avoid being perceived as nerds by their peers). Boehnke (2008) explained that his results indeed indicated that high fear of social exclusion can lead to a negative relation between abilities and grades among students with high math abilities. In their study, Owens, Stevenson, Norgate, and Hadwin (2008) included emotional as well as cognitive factors for the investigation of scholastic under-achievement in children at the age of $11 - 12$ years. More concretely, their study participants were tested on different working memory tasks, trait anxiety items and tests related to scholastic achievement. Among other results, Owens et al. (2008) found that verbal working memory partially mediated the relation of trait anxiety with math achievement. Overall, the authors highlighted the importance of including cognitive as well as emotional factors when assessing academic under-achievement. Fong and Kremer (2020) studied mathematical under-achievement in grade nine high school students. Among other relations, the authors analysed the influence of motivational factors on under-achievement in mathematics based on generalised structural equation models. The authors reported that their results indicated a significant relation between math motivation and under-achievement in mathematics. More specifically, they found that the extend to which mathematics was important to the identity of the students negatively predicted math under-achievement: The more students view mathematics as being a meaningful part of their identity, the less likely they are to under-achieve in math. Furthermore, Fong and Kremer (2020) observed the following interaction: For students with high confidence in their math skills, the degree to which they intrinsically value mathematics is less influential on math under-achievement. On the other hand, among students showing low confidence in their math skills, students intrinsically valuing mathematics demonstrate lower levels of under-achievement in mathematics than students attributing a low intrinsic value to mathematics.

Regarding women's under-achievement in mathematics, Bonnot and Croizet (2007) addressed the research questions whether the endorsement of the stereotype referring to the math inferiority of women impacts women's math performance and whether this effect can be traced back to a disruption of working memory. In terms of the first question, the authors reported that women, who internalised this stereotype, showed a lower self-evaluation of mathematical ability and lower performance in statistics. Moreover, their analyses revealed that math self-evaluation partially mediated the effect of stereotype internalisation on math achievement. With respect to the second question, Bonnot and Croizet (2007) observed that on difficult items covering basic additions, women with a low self-evaluation of mathematical ability indeed made more calculation mistakes and needed more time to solve the additions compared to women with a high self-evaluation of mathematical ability. The authors explained that this result is in line with the consideration that the observed group difference emerges from temporary disruptions of working memory (e.g., due to doubts or worries about one's own math ability), as the difficult test items were more demanding from a working memory perspective. Hofer and Stern (2016) studied gender-specific under-achievement in another STEM subject, namely in physics. Based on a multiple group latent profile analysis approach, the authors detected different gender-specific student profiles in a sample of students at advanced placement schools in Switzerland (Swiss Gymnasium). Among other profiles, they identified a profile of female under-achievers, who exhibited below-average physics grades, even though they possessed high intellectual potential. Hofer and Stern (2016) reported that, compared to other students, these female physics under-achievers were less interested in physics and appeared to have a low physics self-concept. Furthermore, the authors pointed out that among the investigated profiles, physics under-achievers were only found within the group of female students but not within the group of male students. Overall, the under-achieving females constituted 29% of all girls in their sample.

In his cross-sectional study, Phillipson (2008) investigated the proportions of under-achieving students in mathematics by means of Rasch analyses. More specifically, his sample consisted of grade three and grade five students in primary school as well as of grade one and grade three students in secondary school. His study participants were tested on a standardised math test and on a non-verbal measure of intellectual ability. Phillipson (2008) modelled the data with an optimal achievement model, which indicated that a perceivable number of students were under-achieving in mathematics, while the proportion of under-achieving students was largest in the highest investigated grade level (third grade of secondary school).

In summary, the research findings from the studies described above indicated that there are students, whose actual math achievement deviates from their expected math achievement (as predicted by their general reasoning abilities). With respect to over-achievement, it could be observed that an extensive utilisation of learning strategies appears to help students to over-achieve (Castejón et al., 2016). On the other hand, under-achievement can, among other variables, be influenced by the following factors: Motivational aspects (e.g., Fong & Kremer, 2020), mathematics anxiety (e.g., Sepie & Keeling, 1978), stereotype internalisation (e.g., Bonnot & Croizet, 2007) or a lower academic self-concept (e.g., Castejón et al., 2016).

1.3 Open Questions

Subsection 1.2.4 described various studies, which focused on discrepancies between students' expected math achievement (utilising their intelligence as predictor) and their actual math achievement, i.e. which studied over- and under-achievement in mathematics. In one of these studies, Phillipson (2008) assessed the proportions of under-achievers in mathematics at different grade levels and found that the percentage of under-achievers was largest in the highest grade level under investigation, which was grade three of secondary school. The research goal of the current study is to further investigate over- and under-achievement in mathematics in adolescents, namely in students at advanced placement schools in Switzerland (i.e., students at the Swiss Gymnasium). In terms of over-achievement, Castejón et al. (2016) observed that over-achieving students managed to reach a higher scholastic performance by extensively using learning strategies (see also Subsection 1.2.4). The current study aims at addressing a more specific question: Can over-achievement in advanced school mathematics be partly traced back to successful chunking of knowledge and useful knowledge representations? In addition, the goal of current study is to take motivational factors into account as well: Are there any other factors which seem to help students over-achieve? With respect to under-achievement, the aim of the current study is to further explore the following question: Why do under-achieving students at the Gymnasium not (fully) translate their potential in terms of their general reasoning abilities into math achievement, i.e. why do they not unfold their potential better? As explained in Subsection 1.2.4, several studies showed that motivational and emotional aspects can play a role in explaining under-achievement, such as motivation (e.g., Fong & Kremer, 2020), stereotype internalisation (e.g., Bonnot & Croizet, 2007), mathematics anxiety (e.g., Sepie & Keeling, 1978) or a lower academic self-concept (e.g., Castejón et al., 2016). In addition to emotional factors, Owens et al. (2008) also included cognitive variables in their study of scholastic under-achievement, and they reported that the association of trait anxiety with math achievement was partially mediated by verbal working memory (see also Subsection 1.2.4). Building on these results, the focus of the current study lies on the following question: Do working memory functions play a role in explaining under-achievement in advanced school mathematics in addition to motivational variables? Different aspects of working memory are discussed in the following chapter.

2 The Central Role of Working Memory

This chapter is dedicated to a central construct in the field of cognitive psychology, namely to the construct of working memory. While Section 2.1 gives an introduction to this construct, Section 2.2 deals with the working memory model by A. D. Baddeley and Hitch (1974), and Section 2.3 covers the facet model of working memory capacity by Oberauer et al. (2000, 2003); Süß et al. (2002). Section 2.4 then focuses on the relationship between general reasoning abilities and working memory functions. Furthermore, Sections 2.5 and 2.6 illustrate the important role of working memory for learning in general and for math performance in particular. To conclude this chapter, Section 2.7 comments on some open research questions.

2.1 The Construct of Working Memory

The construct of working memory relates to the capability of maintaining and processing information in the mind for a limited amount of time (e.g., Gathercole & Alloway, 2004). Working memory is often thought of being a flexible mental workspace, in which we can hold relevant information elements while performing complex mental activities (Gathercole & Alloway, 2004). Alloway (2010) described working memory as a mental post-it note.

Gathercole and Alloway (2004) presented the following examples in the domain of mental arithmetic to illustrate our usage of working memory. As a first example, they considered the multiplication of two two-digit numbers (for example, 27 and 48) without any auxiliary means (such as a notepad or a calculator). Gathercole and Alloway (2004) explained that in order to successfully carry out this multiplication, it is not only necessary to hold the two numbers in mind, but also to perform different calculation steps and to store the respective intermediate results. They pointed out that we can only obtain the correct result of the multiplication if both, the storage and the processing requirements of the task are met. Gathercole and Alloway (2004) also emphasised that such mental activities tend to be error-prone. In particular, even minor distractions (e.g., an unrelated thought crossing our mind or being interrupted by another person) may lead to a complete loss of the maintained information, and thereby to an unsuccessful computation attempt. In such a situation, the authors explained, it is not possible for us to recall the lost information again, and therefore we have to start the calculation anew. Furthermore, Gathercole and Alloway (2004) pointed out that our ability to perform mental arithmetic is constrained by the amount of information that has to be maintained and processed. For example, most people could not multiply 142 and 891 without any auxiliary means, even though they would have the mathematical knowledge to carry out multiplications. According to Gathercole and Alloway (2004), this can be traced back to the circumstance that the required storage capacity to solve this kind of task exceeds the working memory capacity. Based on these examples, Gathercole and Alloway (2004) illustrated the following characteristics of working memory:

- Working memory is a very helpful and flexible system, which we use for performing different tasks in our daily life.
- Working memory demands attention. If our focus is turned away from the information elements, which we hold in working memory (for example because we are getting distracted), these information elements may be lost and can therefore not be recovered. This shows how fragile the working memory system is.
- Even though there are individual differences with respect to working memory capacities, the amount of information that can be maintained in working memory is limited for everyone. The exceeding of the individual working memory capacity limit results in loss of information.
- It is possible for us to consciously access information elements in our working memory. We are able to judge whether certain information elements have been successfully stored, or whether they have been lost.

Ricker, AuBuchon, and Cowan (2010) described another situation in everyday life, where we potentially experience the limits of our working memory capacity: Imagine that you ask another person for directions and that you do not have the possibility to write them down. If the given directions do not consist of too many information elements, you may be able to keep them in mind and follow them. However, if the given directions comprise too much information, for example if they contain too many turns or too many different street names, you may not be able to remember all the information elements or even forget all of them. More generally, Ricker et al. (2010) summarised in their review that capacity limits of working memory could be observed in several experimental studies.

Regarding the development of working memory capacity in children, Gathercole, Lamont, and Alloway (2006) made the following observations. First, they illustrated that working memory capacity continuously develops throughout the school years on primary and secondary school levels. They explained that most children get close to the performance of the average adult on working memory assessments by the age of 14 years. Moreover, they pointed out that there are substantial individual differences with respect to working memory abilities. For example, Gathercole et al. (2006) illustrated that in a typical class of 9-year-old pupils, one is likely to find children with working memory capacities varying from the capacity of the average 7-year-old to the capacity of the average 12-year-old.

2.2 The Working Memory Model by Baddeley and Hitch

An influential model of working memory was proposed by A. D. Baddeley and Hitch (1974). In his overview, A. Baddeley (2006) gave a detailed description of this model and reflected on it. According to this model, working memory may be subdivided into the following three components:

- The phonological loop, which is a system responsible for maintaining and processing sound an speech.
- The visuo-spatial sketchpad, which plays a comparable role as the phonological loop with respect to nonverbal information.
- The central executive, which is a system controlling our attention as well as choosing and performing different strategies.

A. Baddeley (2006) described the relation between these components as follows. The central component of the model is the central executive, which represents an attentional control system with confined capacity. The phonological loop and the visuo-spatial sketchpad are two secondary systems supporting the central executive: While the phonological loop is in charge of maintaining sound-based and speech-based information elements, the visuo-spatial sketchpad is responsible for holding visual information elements.

As explained in A. Baddeley (2006), the phonological loop is assumed to comprise two components. The first one temporarily stores phonological information elements either until these elements fade or until the second component refreshes them. According to A. Baddeley (2006), the assumption behind these processes is that inner speech is utilised to rehearse the information elements, which involves successive retrieval of the elements from the store component and feeding them back by articulating them. A. Baddeley (2006) illustrated this with the following example. If the task is to maintain a sequence of three digits, then less than a second is needed to rehearse this sequence, which means that the first digit will not have faded once the last digit has been articulated. On the other hand, if the sequence of digits becomes longer and longer, at a certain point it will not be possible any more to rehearse all the digits contained in the sequence quickly enough to avoid losing some of the information elements. As pointed out by A. Baddeley (2006), this circumstance shows why our digit span is limited.

According to A. Baddeley (2006), for visuo-spatial information elements, the visuo-spatial sketchpad has a comparable function to the one of the phonological loop for phonological information elements: The visuo-spatial sketchpad appears to be important for building visual and spatial world knowledge, such as understanding how a bicycle works or how to find the way to a certain location. However, A. Baddeley (2006) clarified that due to a lack of well-developed and standardised measures of visuo-spatial world knowledge, the visuo-spatial sketchpad could not be studied as easily as the phonological loop.

Lastly, A. Baddeley (2006) explained that while the central executive is assumed to be the most important component of working memory, it is also the most complex and least understood component. He described that the central executive was viewed as a source of general processing capacity, which could carry out different functions exceeding the limits of the phonological loop and the visuo-spatial sketchpad. According to A. Baddeley (2006), the construct of the central executive should be regarded as a starting point for investigating further questions related to different functional aspects of the central executive.

A. Baddeley (2000) proposed to add a fourth component to the original three-component model of working memory by A. D. Baddeley and Hitch (1974) discussed above, namely the so-called episodic buffer: The episodic buffer is thought of being a temporary storage system with limited capacity, which is able to integrate information elements from different sources. As described in A. Baddeley (2000), it is assumed that the episodic buffer is controlled by the central executive. Furthermore, it is suggested that the episodic buffer takes a constructive role in building new cognitive representations and forming chunks by binding information elements together (A. Baddeley, 2000, 2006). A. Baddeley (2000) explained that the revised four-component model of working memory focuses more on the processes of information integration compared to the original three-component model by A. D. Baddeley and Hitch (1974), which aimed at isolating the different subsystems.

2.3 The Facet Model of Working Memory Capacity

For the current study, the facet model of working memory capacity by Oberauer et al. (2000, 2003); Süß et al. (2002) was chosen as a theoretical framework for the concept of working memory. The authors hypothesised that different aspects of working memory can be categorised along two facets: One facet is associated with the content-related side of working memory, while the other facet describes the functional categories of working memory. The content-related facet distinguishes between three different domains: Working memory can be classified either as verbal working memory, numerical working memory, or spatial-figural working memory. On the functional facet, three functional categories of working memory are comprised in the model (see also von Bastian & Oberauer, 2013): Storage-Processing, Relational Integration and Supervision. The functional category of Storage-Processing (or simultaneous storage and processing) relates to the function of concurrently maintaining and manipulating information. Furthermore, the process of coordinating isolated pieces of information into new compositions is described by the working memory function of Relational Integration. Lastly, the functional category of Supervision is associated with the selective activation of relevant information elements on the one hand, and with the inhibition of irrelevant information elements on the other hand. Süß et al. (2002) illustrated the facet model of working memory graphically: Figure 1 shows an adapted version of this graphical representation.

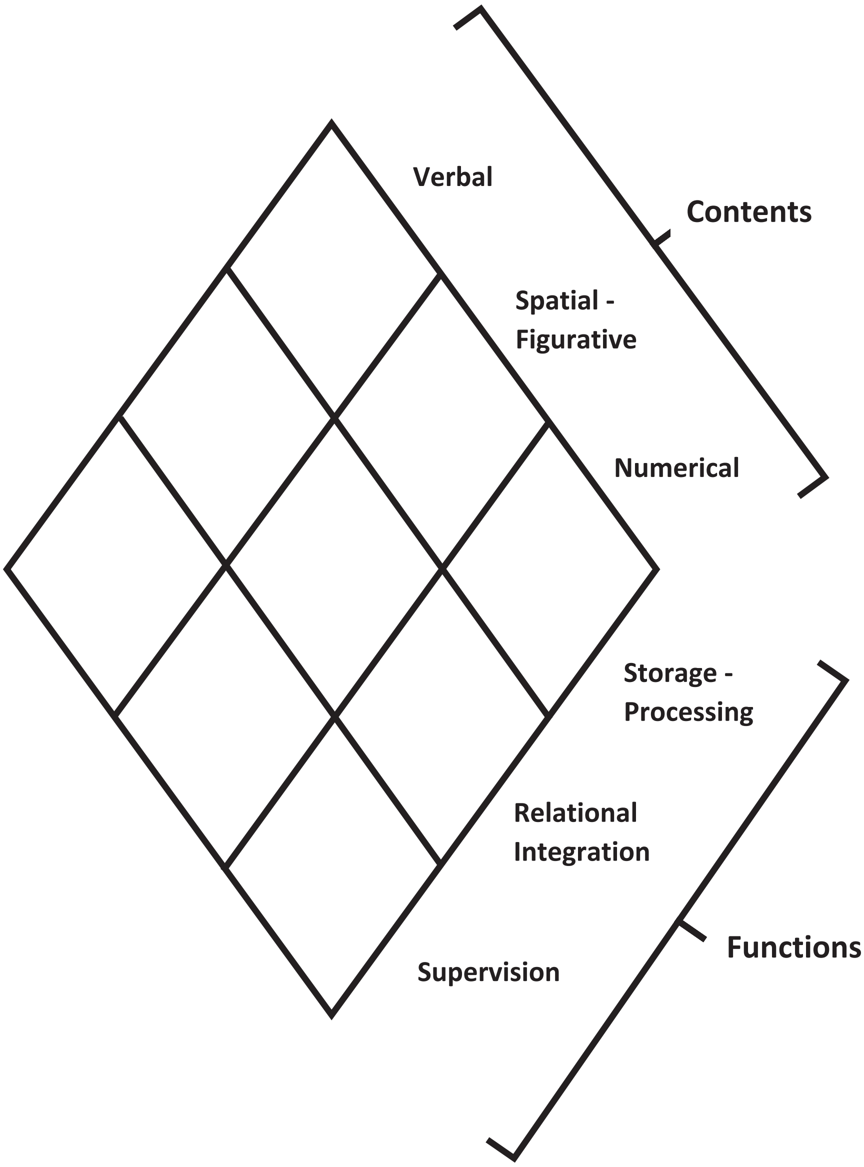


Figure 1: Graphical representation of the facet model of working memory, adapted from Süß et al. (2002).

In their study, Oberauer et al. (2000) investigated the proposed facet structure by means of various working memory tasks targeting different facets. The authors carried out exploratory and confirmatory factor analyses to asses the data from the working memory tasks and found the following. With respect to the content-related facet, they observed that spatial working memory could be clearly separated from the other two content categories, while an evident distinction between verbal and numerical working memory did not emerge. As explained by Oberauer et al. (2000), this finding is in line with the working memory model suggested by A. D. Baddeley and Hitch (1974) (see Section 2.2), which distinguishes between the phonological loop, a system for phonological information (such as words and numbers), and the visuo-spatial sketchpad, a system for visual and spatial information. Regarding the functional facet, Oberauer et al. (2000) reported that the functional categories of Storage-Processing and of Relational Integration were not separable. On the other hand, there appeared to be a clear differentiation between the working memory function Supervision and the other two functions. For the interpretation of these results, Oberauer et al. (2000) pointed out that there was a considerable overlap between the functional categories of Storage-Processing and of Relational Integration in the corresponding working memory tasks selected from the literature, so that they could not exclude the possibility of the two categories being distinguishable with more pure measures of working memory. In fact, when Oberauer et al. (2003) constructed a variety of working memory tasks to operationalise the different facets of the facet model, structural equation models of their test data showed that all three working memory functions could be separated from each other. At the same, Oberauer et al. (2003) reported that their data's support for a content-related differentiation between verbal–numerical working memory and spatial working memory was rather weak.

2.4 The Relation between General Reasoning Abilities and Working Memory Functions

Various findings indicated that there is a strong relation between working memory and intelligence (e.g., Colom, Flores-Mendoza, & Rebollo, 2003; Fry & Hale, 2000; Süß et al., 2002). Several scientists debated on the magnitude of this association: While Ackerman, Beier, and Boyle (2005) reported an average correlation of $r = .479$ between estimates of working memory and g in their meta-analysis, the reanalysis conducted by Kane, Hambrick, and Conway (2005) yielded a median correlation of $r = .72$ between working memory and reasoning factors. Moreover, Oberauer, Schulze, Wilhelm, and Süß (2005) also carried out a reanalysis and found an estimated correlation of $r = .85$ between working memory and g. In terms of working memory functions, Oberauer, Süβ, Wilhelm, and Wittmann (2008) found that not only the functional category of Storage-Processing, but also the one of Relational

Integration predicted reasoning ability well. The results from Wongupparaj, Kumari, and Morris (2015) revealed that both, the processing as well as the storage component of working memory made a contribution to the relation with intelligence.

Given the strong association between working memory and intelligence described above, one could ask the question whether these two constructs are actually isomorphic. In their reviews, Conway, Kane, and Engle (2003) and Chooi (2012) summarised that the two constructs did not appear to be isomorphic, even though a close relationship could be observed between them. de Abreu, Conway, and Gathercole (2010) assessed the link between working memory and fluid intelligence in young children. A confirmatory factor analysis of their data revealed that the two constructs were strongly associated but separable (de Abreu et al., 2010). The results from Alloway and Alloway (2010) showed that while there was a considerable amount of shared variance between IQ and working memory skills with respect to learning outcomes, both constructs were unique predictors of attainments in literacy and numeracy. Given the unique links between working memory and academic achievement, the authors concluded that "working memory is not a proxy for IQ but rather represents a dissociable cognitive skill" (Alloway & Alloway, 2010, p. 20). In particular, they found that at the beginning of the educational path, working memory predicts successive academic performance more strongly than IQ. Giofrè, Mammarella, and Cornoldi (2013) investigated the relation between working memory and intelligence in 4th and 5th grade pupils. Their data analysis based on structural equation models revealed that a considerable amount (66%) of the variance in g could be predicted by working memory. The authors explained that this result is in line with the conceptualisation of working memory and intelligence as being two closely associated but separable constructs.

To summarise the results from the studies described above, two points should be noted: First, there are several findings showing that working memory and intelligence are strongly related (e.g., Colom et al., 2003; Fry & Hale, 2000; Süß et al., 2002). Second, despite their close association, the constructs of working memory and intelligence seem to be distinguishable (e.g., Chooi, 2012; Conway et al., 2003).

2.5 The Importance of Working Memory for Learning

With respect to learning processes, Alloway and Copello (2013) explicated that working memory measures lead to better predictions of learning than IQ measures, because working memory is associated with our potential to learn rather than with the knowledge we already built up. In particular, the authors pointed out that several aspects of IQ tests target the knowledge that students already acquired. Alloway and Copello (2013) illustrated this difference between working memory tests and IQ tests on the basis of the following examples. One example of a working memory task is to present a sequence of numbers to the students and then ask them to recall the sequence in reversed order. The authors explained that if students are experiencing difficulties with this kind of task, it is often not due to a lack of understanding of number magnitude or counting, but it can be traced back to their working memory capacity, which is not sufficiently large for remembering three or four numbers. On the other hand, a common subscale in IQ tests consists of a vocabulary test. If students are familiar with the definitions of the words covered in the vocabulary test, this is likely to be reflected in high scores on this test, while students not knowing the definitions are likely to obtain low scores (Alloway & Copello, 2013). Overall, Alloway and Copello (2013) emphasised that the role of working memory for learning is crucial throughout the educational path, from preschool (Alloway et al., 2005) to tertiary education (Alloway & Gregory, 2013). Furthermore, they pointed out that while working memory appears to be a strong predictor of reading success, there are associations between working memory and math outcomes as well (see also Section 2.6). In particular, Alloway and Alloway (2010) found that a child's working memory skills at the age of five years significantly predict language and math performances six years later.

On the other hand, there are several learning problems, which can be traced back to poor working memory skills (Gathercole et al., 2006). For example, Gathercole et al. (2006) referred to the research findings showing that students with poor working memory skills obtained low scores on national curriculum assessments in England at seven, eleven and fourteen years of age (Gathercole & Pickering, 2000; Gathercole, Pickering, Knight, & Stegmann, 2004; Jarvis & Gathercole, 2003). Siegel and Ryan (1989) investigated working memory functions in normally achieving and learning disabled students. More specifically, they carried out group comparisons with respect to two working memory tasks, one with verbal and one with numerical content. Their results indicated that the reading disabled children scored significantly lower than the normally achieving children on both working memory tasks. In addition, compared to the normal achievers, the performance of arithmetic disabled students was significantly lower only on the numerical working memory task, which involved counting. Siegel and Ryan (1989) therefore drew the following conclusions: While a generalised deficit in working memory seems to be a characteristic of a reading disability, students with an arithmetic disability appear to have a specific working memory deficit associated with the processing of numerical content.

Gathercole et al. (2006) addressed the question on how the underlying mechanisms contributing to the relationship between working memory and learning could be described. They summarised that there are two main types of suggested models. The assumption behind the first model is that poor working memory skills arise from difficulties in a certain processing field. As explained by Gathercole et al. (2006), according to this first model, low scores on a verbal working memory task for example would be traced back to a language processing problem rather than to a low working memory capacity per se. However, the authors clarified that the evidence in favour of this model is not convincing, as several studies showed that the associations between working memory measures and measures of learning outcomes in reading and mathematics cannot be explained purely by the processing elements of the respective working memory tasks (for example, Gathercole et al., 2006 referred to the review by Engle, Tuholski, Laughlin, & Conway, 1999 to strengthen their argument). Gathercole et al. (2006) outlined that it seems that the general working memory capacity plays a crucial role, rather than a specific skill within a processing domain. The second model presented in Gathercole et al. (2006) suggests that the ability to acquire complex skills and to build knowledge is directly confined by working memory capacity. However, as pointed out by Gathercole et al. (2006), the details of the role, which working memory plays in the acquisition of new skills and knowledge are not fully understood. The authors explicated that statistical relations between working memory measures and measures of learning outcomes do not contribute to the understanding of this role, as they provide information on the learning outcomes rather than on the learning process.

As explained in Section 2.1, working memory capacity is limited (see for example Ricker et al., 2010). A cognitive process, which is assumed to be helpful for circumventing this capacity constraint to a certain extent, is the one of chunking (Thalmann, Souza, & Oberauer, 2019). Miller (1956) described the process of chunking as grouping or organising bits of information into familiar units or chunks of information. He emphasised that the formation of these chunks is an integral part of learning. Similarly, Cowan (2001) defined a chunk as a group of concepts, which are strongly associated to each other and weakly related to other chunks. In their study, Thalmann et al. (2019) addressed the question whether chunking helps lowering the load on working memory. More specifically, they assessed the use of chunks in different working memory tasks. Their results revealed that chunking is not only beneficial for recalling the chunked information, but it is also advantageous for retrieving not-chunked information that is contemporaneously held in working memory. The authors therefore concluded that the load on working memory can be reduced by chunking, which in turn improves memory for the simultaneous maintenance of other information. The study conducted by Portrat, Guida, Phénix, and Lemaire (2016) illustrated the benefits of chunking on the basis of the following working memory task: Young adults were given the task to memorise a series of seven letters while concurrently performing location judgement tasks. Afterwards, they were asked to recall the series of letters. The series of letters, which were presented to the study participants, either comprised a random string of letters or a string including a three-letter acronym in the first, third or fifth position in the series. The results indicated that the recall performance of the study participants was higher for series containing acronyms compared to series without acronyms. Based on this observation, Portrat et al. (2016) pointed out that chunking seems to boost working memory performance.

In summary, research findings indicated that working memory plays an important role for learning across all grade levels (e.g., Alloway & Copello, 2013; Alloway et al., 2005; Alloway & Gregory, 2013). As such, poor working memory skills can be the source of different learning problems (e.g., Gathercole et al., 2006). Another aspect discussed above is the learning process of chunking: Thalmann et al. (2019) reported that chunking can help reducing the load on working memory, which in turn frees up memory resources for maintaining other information elements contemporaneously.

2.6 The Relation between Working Memory Functions and Math Performance

Working memory functions assumably play an important role in mathematical problem solving, as math tasks often require maintaining partial information and processing new information (Raghubar, Barnes, & Hecht, 2010). Multiple studies found an association between working memory and math achievement (e.g., Campos, Almeida, Ferreira, Martinez, & Ramalho, 2013; Caviola, Colling, Mammarella, & Szűcs, 2020; Giofrè, Donolato, & Mammarella, 2018; Holmes & Adams, 2006; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Musso, Boekaerts, Segers, & Cascallar, 2019; St Clair-Thompson & Gathercole, 2006; Swanson & Kim, 2007; Zheng, Swanson, & Marcoulides, 2011). In their review, Raghubar et al. (2010) observed that working memory and math achievement appear to be related. At the same time, the authors pointed out that the associations between working memory and mathematics are complex, as they may be influenced by different factors, such as age or skill level of the study participants, type of math skill and type of working memory measure. The focus of the meta-analysis conducted by Friso-Van den Bos, Van der Ven, Kroesbergen, and Van Luit (2013) lay on the investigation of the relationship between working memory and math performance in primary school pupils. The authors reported that all working memory components under consideration were positively and significantly related to math achievement. With respect to different math measures, Friso-Van den Bos et al. (2013) found that general math tests (such as composite measures or national curriculum tests) were more strongly correlated with working memory measures compared to math tests targeting a specific mathematical skill. In their meta-analysis, Peng, Namkung, Barnes, and Sun (2016) also investigated the relationship between working memory and mathematics. Based on 110 studies with 829 effect sizes, the authors reported a significant medium correlation between working memory and mathematics ($r = 0.35$). Furthermore, Peng et al. (2016) carried out moderation analyses, which suggested that mathematics was comparably related to verbal working memory, numerical working memory, and visual-spatial working memory. The authors therefore concluded that it seems to be the domain-general central executive component of working memory, which shapes the relationship between working memory and mathematics, rather than domain-specific aspects of working memory. With respect to different math skills, Peng et al. (2016) observed that while all types of math skills included in the meta-analysis were significantly associated with working memory, word problem solving and whole number calculations demonstrated the closest associations ($r = 0.37$ and $r = 0.35$ respectively) and geometry showed the weakest association ($r = 0.23$).

The studies presented in this paragraph investigated the role of working memory for early math skills. Passolunghi, Vercelloni, and Schadee (2007) assessed the relationships between different cognitive abilities and math performance in a longitudinal study design. To this end, their study participants were tested twice: Once at the start and once at the end of grade one in primary school. The authors modelled the data by means of a structural linear model. Among other findings, this model suggested that working memory significantly predicts mathematics learning at the start of primary school and that it can therefore be viewed as a direct precursor of early math performance. In a further study, Passolunghi and Lanfranchi (2012) investigated working memory (among other cognitive abilities), measured in preschool children, as a predictor for math performance later on. Working memory measures (and other measures of different cognitive abilities) were administered to the study participants at the start of their last year of preschool. In a second and third step, the study participants underwent a test on numerical competence at the end of their preschool, and a test of math performance at the end of their first year in school. Based on path analysis models, Passolunghi and Lanfranchi (2012) presented the following results with respect to working memory measures. On the one hand, they observed a direct influence of working memory for the prediction of numerical competence at the end of preschool. On the other hand, they found an indirect effect of working memory on math performance at the end of grade one, which was mediated by numerical competence. Among other relations, Purpura and

Ganley (2014) assessed the relationships between working memory and different aspects of early math skills in preschool children. They observed that working memory was significantly associated with only a few, but critical, early math skills, namely with cardinality, subitizing, set comparison, and number order. Purpura and Ganley (2014) explained that a characteristic of the math tasks targeting these concepts was that they were a bit more complex than some of the other math tasks. In particular, study participants needed to carry out multiple steps to obtain the correct result. Purpura and Ganley (2014) illustrated this with the following example: For the set comparison task, children were required to first enumerate each of the given sets, and then to maintain these set sizes in their mind, while judging which of the sets was the largest one.

The following studies explored the predictive value of working memory for math performance. In their longitudinal study, De Smedt et al. (2009) assessed the relation between different measures of working memory and individual differences in math performance based on a longitudinal correlational design. There were three measurement points in total: At the beginning of first grade, the study participants performed different working memory measures associated with the working memory components of the visuo-spatial sketchpad, the phonological loop, and the central executive. Measures of math performance were collected four months (halfway through the first school year) and one year (at the beginning of the second school year) after the working memory assessment respectively. To investigate the data, De Smedt et al. (2009) carried out correlational analyses, which revealed that all working memory components were predictively associated with subsequent math performance in grade one and two. Furthermore, the authors explained that their regression analyses indicated the following pattern: With respect to math performance in first grade, the central executive as well as the visuo-spatial sketchpad appeared to be unique predictors. Regarding math achievement in second grade, unique variance was predicted by the central executive and the phonological loop. Thus, De Smedt et al. (2009) summarised that working memory seems to be a precursor of subsequent math achievement and that the respective predictive value of working memory can already be observed at the beginning of formal education. Allen, Giofrè, Higgins, and Adams (2021) conducted a follow-up study building on an earlier study, in which the study participants were tested on different working memory measures as well as on a standardised math measure (Allen, Giofrè, Higgins, & Adams, 2020b). The aim of the follow-up study was to investigate the relation between working memory and later math achievement. To this end, the former study participants were administered a math test again after a time period of two years. Taken together, the results from Allen et al. (2020b) and Allen et al. (2021) showed that there is an association between working memory and mathematics at both measurement points. In particular, Allen et al. (2021) pointed out that it seems to be possible to predict math performance based on working memory data, which was collected two years before.

There are mixed research findings regarding the specific roles of verbal working memory and visual-spatial working memory for math performance. For example, among other relations, Stevenson, Bergwerff, Heiser, and Resing (2014) investigated the association between working memory and math achievement in young school children during a school year. More specifically, their study comprised two measurement points: In the middle of the school year, the study participants underwent verbal and visual-spatial working memory tasks as well as a math performance test. Six months later, math achievement was assessed again. As described by Stevenson et al. (2014), a multilevel mixed-effects model revealed that verbal working memory emerged as a unique predictor of both, concurrent as well as later math performance, while this could not be observed for visual-spatial working memory. On the other hand, Bresgi, Alexander, and Seabi (2017) also assessed the relations between verbal working memory and math achievement as well as between visual-spatial working memory and math achievement. Their study participants (grade two pupils) underwent different working memory tasks and completed a mathematics test. While a predictive relationship between verbal working memory and math performance did not emerge from the regression analyses, Bresgi et al. (2017) pointed out that their results indicated a significant association between visual-spatial working memory and math achievement: Higher scores on visual-spatial working memory tasks were reflected in higher scores on the mathematics test. Berkowitz, Edelsbrunner, and Stern (2022) investigated the relationship between working memory and math achievement among undergraduate students of mathematics, physics and mechanical engineering at the beginning of their math-intensive study programs. In particular, they focused on the question whether verbal working memory and visual-spatial working memory differ in their contributions to math performance. Based on a latent correlational analysis, the authors found that verbal working memory as well as visual-spatial working memory were both significantly associated with math achievement, and that the strengths of these associations were comparable.

Also in terms of a potential age-dependence of the influence of verbal working memory and of visual-spatial working memory on math achievement different research findings can be observed. On the one hand, Van de Weijer-Bergsma, Kroesbergen, and Van Luit (2015) assessed the relationships between verbal and visual-spatial working memory and math performance at different time points during primary school. The study participants were selected among pupils attending grades 2,3,4,5 and 6 respectively. The authors modelled the test data by means of multilevel multigroup latent growth models in order to study the predictive role of working memory for each grade. They observed the following pattern: While the predictive power of visual-spatial working memory as a predictor of individual differences with respect to math achievement decreased with increasing school grade level, the predictive value of verbal working memory grew. On the other hand, Allen, Giofrè, Higgins, and Adams (2020a) focused on the question how verbal working memory and visual-spatial working memory contribute to math achievement in primary school pupils as well. Different working memory tasks and a math test were administered to the study participants. Allen et al. (2020a) reported that their correlation analyses showed that all working memory tasks were correlated with math performance. Furthermore, the authors investigated whether the different contributions of working memory to math achievement varied with the age of the study participants. They observed that, while verbal working memory and visual-spatial working memory both uniquely contribute to math performance, the influence of visual-spatial working memory was higher in older pupils.

Several studies analysed the specific role of visual-spatial working memory for math achievement. In their review, Allen, Higgins, and Adams (2019) systematically studied the literature regarding the association between visual-spatial working memory and math performance in school-aged children with respect to the reported effect sizes. Based on 35 independent studies, which specified effect sizes ranging from $r = 0.040$ to $r = 0.690$, the authors found that the overall effect size is positive, which indicates that visual-spatial working memory and math achievement are positively related. At the preschool and primary school level, Bull, Espy, and Wiebe (2008) analysed the role of working memory (among other cognitive measures) for subsequent scholastic achievement in a longitudinal study. More concretely, their study participants underwent a test battery of cognitive measures in preschool as well as math tests and reading tests at the beginning of primary school and at the end of grade one and three respectively. With respect to math performance, Bull et al. (2008) reported that visual-spatial working memory was a predictor of math skills at the end of grade three in primary school. According to Bull et al. (2008), this finding illustrates that the ability of processing visual-spatial information in working memory plays an important role for math performance. van der Ven, van der Maas, Straatemeier, and Jansen (2013) assessed the role of visual-spatial working memory for math performance in primary school. Their regression analyses revealed a significant association between visual-spatial working memory and math achievement. Fanari, Meloni, and Massidda (2019) studied the influence of visual-spatial working memory on math performance in young primary school pupils. Based on a longitudinal study design, their study participants were administered different measures of visual-spatial working memory as well as math tasks at three different points in time: At the start and at the end of the first school year, as well as at the end of the second school year. Based on correlation analyses, Fanari et al. (2019) observed that while a strong association between both visual and spatial working memory and math achievement emerged at the start of grade one, only visual working memory appeared to be associated with math performance at the end of grade one. However, at the end of grade two, both working memory components (visual and spatial) were related to math achievement again.

Li and Geary (2013) investigated gains in visual-spatial working memory from first grade to fifth grade and their importance for mathematics learning. Among other variables, the authors measured the study participant's math performance from preschool to fifth grade and their working memory capacity in grade one and five respectively. Li and Geary (2013) reported that pupils with the largest first-to-fifth grade gains in visual-spatial working memory showed higher math achievement at the end of fifth grade compared to their peers with smaller gains. To extend these findings, Li and Geary (2017) addressed the question whether the first-to-fifth grade gains in visual-spatial working memory are predictive of gains in math performance later on as well. The authors indeed observed that these gains also make a contribution to individual differences with respect to math performance in grade six to grade nine. Reuhkala (2001) studied the relation between visual-spatial working memory and mathematical skills in grade nine students. Her findings indicated that both, the storage as well as the processing component of visual-spatial working memory appear to be associated with mathematical skills. Kyttälä and Lehto (2008) analysed the association between visual-spatial working memory and math achievement in adolescents. The authors reported that intelligence measures as well as measures of visual-spatial working memory explained scores on different mathematical subscales, which covered geometry, word problems, and mental arithmetic. In particular, Kyttälä and Lehto (2008) observed that students, who obtained low scores on the mathematics measures, had a poorer visual-spatial working memory capacity compared to their peers with average or high scores on these measures.

The numerous studies described in the foregoing paragraphs documented several aspects of the relationship between working memory and math achievement. While most of these studies were carried out with primary school pupils, there are also research findings indicating that working memory still plays a role for math performance in tertiary education as well: As explained above, Berkowitz et al. (2022) reported that verbal working memory and visual-spatial working memory both showed significant relations to math achievement in undergraduate students in STEM fields, i.e. in a sample with high math competence.

2.7 Open Questions

Section 2.6 described various studies, which assessed the association between working memory functions and math performance. If one takes a closer look at the age of the study participants in these studies, it can be noted that the majority of them were conducted with primary school children. Alloway, Banner, and Smith (2010) also pointed out that while there are several research findings documenting the important role of working memory for learning at the primary school level, only a couple of studies investigated this role at the secondary school level. The authors therefore examined a sample consisting of 13-year-old students and found that working memory predicted learning outcomes in mathematics in secondary school students as well. The research aim of the current study is to make a contribution to the investigation of the relation between working memory functions and math performance in adolescents. More specifically, the current study focuses on a sample of students at the Swiss Gymnasium in order to approach the following research questions: Which role do working memory functions play for advanced school mathematics? Which conclusions can be drawn about the interplay between working memory functions, general reasoning abilities and math performance in students at the Swiss Gymnasium? Based on these research questions and the ones discussed in the previous chapter (see Section 1.3), the subsequent chapter elaborates on the rationale for the current study and it formulates the research goals of the current study in a more concrete manner.

3 The Current Study

The previous chapters discussed individual differences in math achievement as well as the role of working memory for math performance. The overarching research goal of the current study is to further investigate the interplay between general reasoning abilities, working memory functions and math achievement in students at advanced placement schools in Switzerland (i.e., at the Swiss Gymnasium). In particular, the current study aims at assessing over- and under-achievement in advanced school mathematics. The focus of this chapter lies on giving a detailed description of the core of the current study: While Section 3.1 is dedicated to the rationale for the current study, Section 3.2 elaborates on the specific research goals of the current study.

3.1 Rationale for the Current Study

The aim of this section is to explain and illustrate how the current study approaches its overarching research goal. While the subsequent subsection focuses on the considerations regarding the study participants of the current study, the remaining subsections describe the ideas behind different measures for the current study.

3.1.1 Study Participants

The review by Raghubar et al. (2010) illustrated that most studies assessing the relationship between working memory and math achievement were carried out with primary school children or with children experiencing math difficulties (see also Sections 2.6 and 2.7). In particular, the number of studies that focused on this relationship in secondary school students seems to be comparatively small (a few examples of studies, which analysed the association between working memory and math performance in adolescents, are Gathercole et al., 2004, Kyttälä & Lehto, 2008 and Reuhkala, 2001). Given these considerations, one aim of the current study is to contribute to the understanding of the relationship between working memory functions and math achievement in adolescents. Instead of further investigating students with learning difficulties, the current study focuses on students in advanced placement schools in Switzerland (Swiss Gymnasium). More specifically, the sample of the current study should consist of high-achieving, over-achieving and under-achieving students, as this composition enables the assessment of the interplay between general reasoning abilities, working memory functions and math performance by carrying out group comparisons between the different groups of students. It is important to mention that for the sample selection, the current study has access to the data base of the TraM-study, which is a large scale study examining the learning transfer within mathematics and from mathematics to physics at the Swiss Gymnasium. A detailed description of the TraM-Study, of the sample selection process as well as of the sample for the current study is provided in Section 4.1. More details on the research goals behind the different group comparisons are given in Section 3.2.

3.1.2 Working Memory Test Battery

For the assessment of working memory, Lewandowsky, Oberauer, Yang, and Ecker (2010) pointed out the following observation: Even though Oberauer et al. (2000) reported that there are various types of tasks, which load on the general working memory factor, in practice working memory is often measured by only one type of tasks, namely by so called complex span tasks (for a review on working memory span tasks, see Conway et al., 2005). As explained in Lewandowsky et al. (2010), complex span tasks comprise two components: A memory task (e.g., memorising a list of items) followed by a processing task (e.g., judging a statement). Figures 2 and 3 visualise an example of a complex span task with verbal content, which was designed by von Bastian and Oberauer (2013). The first part of the task is to memorise a list of words (as indicated in Figure 2), while the processing task in the second part (given in Figure 3) is to match the term shown centrally on the screen to one of the four options displayed in the four corners of the screen in such a way that the central term represents an umbrella term for the chosen option. For the example presented in Figure 3, *autumn* would be the correct choice.

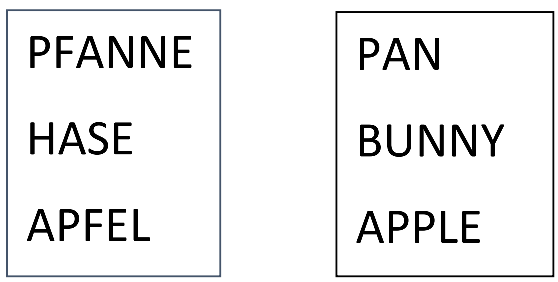


Figure 2: First part of a verbal complex span task designed by von Bastian and Oberauer (2013). The German version is the original version of the task.

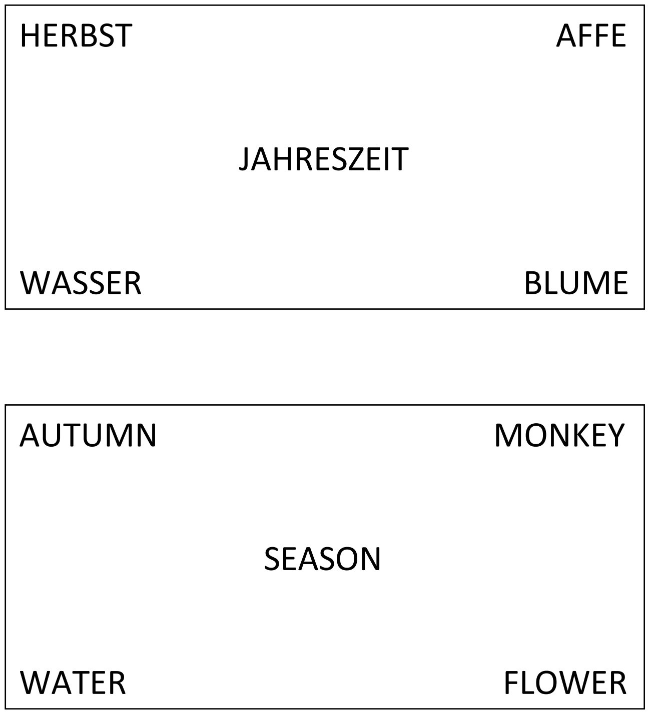


Figure 3: Second part of a verbal complex span task designed by von Bastian and Oberauer (2013). The German version is the original version of the task.

Given that complex span tasks are an established type of task to assess working memory (a review on working memory span tasks is provided in Conway et al., 2005), the current study aims at including these tasks as cognitive measures. However, Lewandowsky et al. (2010) emphasised that although complex span tasks may be particularly handy to implement, measuring working memory by only one type of tasks comes at a price: The authors argued that this practice results in contaminated measurements of working memory, as they include variance, which is not associated with general working memory, but rather with the characteristics of the respective task paradigm. Lewandowsky et al. (2010) therefore recommended to use a test battery of heterogeneous indicators for the assessment of working memory, so that such task-specific variance can be reduced. Following this recommendation, the aim was to find an extensive working memory test battery containing different types of tasks, which could be implemented for the current study. After examining a couple of working memory test batteries, the ideal test battery for the current study seemed to be the facet test battery proposed by von Bastian and Oberauer (2013). In addition to the established complex span tasks (see Conway et al., 2005, for an overview on working memory span tasks), it also comprises other types of working memory tasks. Furthermore, the facet test battery does not only fulfil the criterion to include different types of working memory tasks, but it also has the advantage of covering all three functional categories as well as all three content-related categories of working memory as suggested by the facet model (Oberauer et al., 2000, 2003; Süß et al., 2002, see Section 2.3 for a short description of the facet model) - the heterogeneous test battery provided by Lewandowsky et al. (2010) for example only applies to two of the three functional categories of working memory. More concretely, the facet test battery comprises nine tasks in total: For each of the working memory functions Storage-Processing, Relational Integration and Supervision, there is a figural, a numerical, and a verbal task respectively. A graphical representation of the facet test battery is given in Figure 4. Based on von Bastian and Oberauer (2013), the different tasks of the facet test battery and their implementation are described in more detail in Subsection 4.2.2.

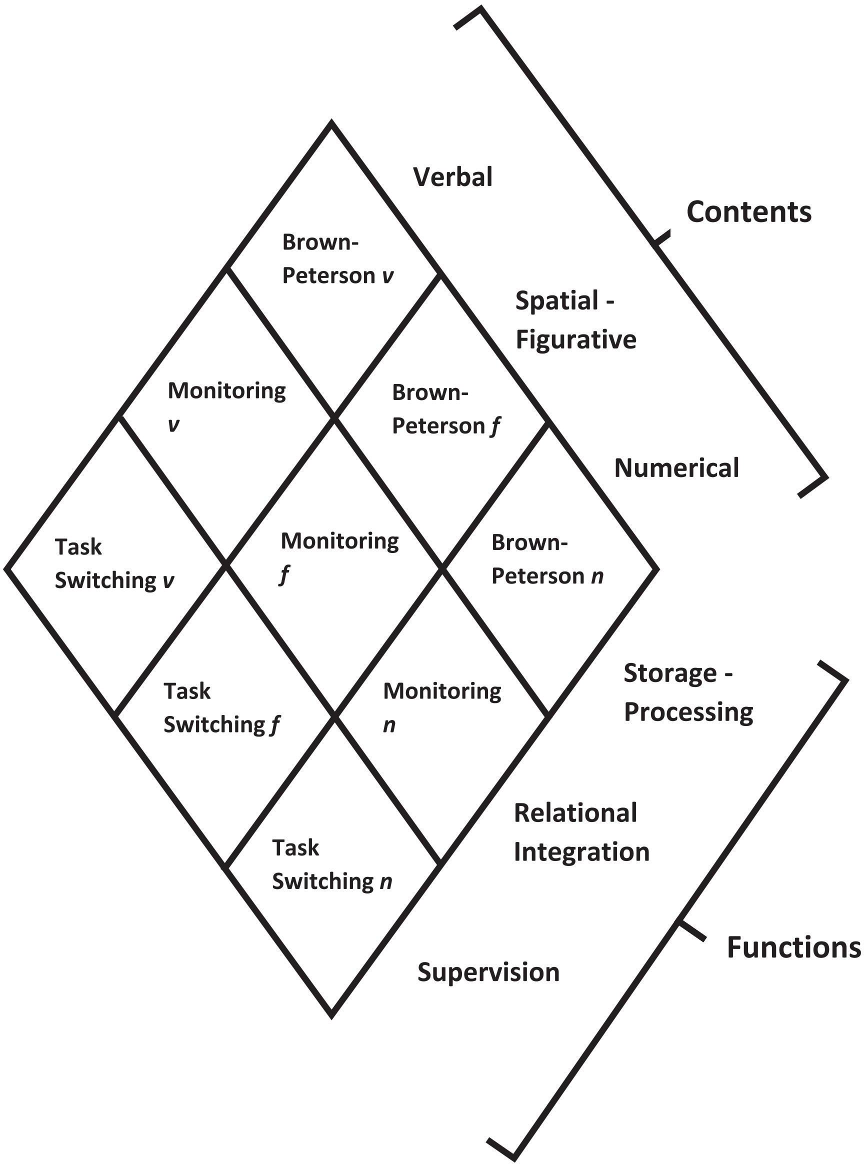


Figure 4: Graphical representation of the facet test battery of working memory suggested by von Bastian and Oberauer (2013). This graphical representation is an adapted version of the graphical representation in Süß et al. (2002).

3.1.3 Mathematics Measures

As listed in the review by Raghubar et al. (2010), a considerable number of studies investigating the relation between working memory and math performance utilises arithmetic tasks as a measure of math performance. In order to gain additional insights into the associations between working memory functions and other facets of math achievement (such as conceptual understanding and procedural knowledge in advanced school mathematics), the current study aims at devising two different mathematics measures. As explained in the previous chapter, successful chunking of knowledge is expected to unburden working memory to a certain degree. One goal of the current study is to assess this interplay by a speed test with mathematical content, which targets the knowledge representations of the students. On the other hand, as anticipated in the foregoing chapter, working memory functions are assumed to be important for mathematical problem solving. In addition to the speed test, the aim of the current study is therefore to design a mathematical power test, which contains different problem solving tasks. The ideas behind the two mathematics tests for the current study are illustrated in more detail in the paragraphs below.

Mathematics Speed Test Regarding the Mathematics Speed Test, the focus lies on developing a test consisting of various mathematical statements, which should be judged by the study participants as quickly and at the same time as correctly as possible without any auxiliary means. The goal behind this test design is that the answer patterns of the students should reveal how well they organised and represented their knowledge. More concretely, study participants, who managed to successfully chunk their knowledge, are expected to be able to solve the tasks in a quick and correct way. On the other hand, students with insufficiently organised knowledge representations may struggle to solve the test items correctly within a given time frame and without any auxiliary means.

To illustrate the considerations above in a more concrete manner, Figure 5 shows three examples of possible test items. The idea behind the first statement is the following: If the first example of a linear function, which comes to students' minds when they think about linear functions, is one of a monotonically increasing linear function - for example because they were introduced to monotonically increasing linear functions before they learned about monotonically decreasing linear functions, or because they worked with more examples of monotonically increasing linear functions than of monotonically decreasing linear functions -, then they may be tempted to think that the statement is true if they only have a limited time frame to think about it. In contrast, if they manage to quickly recall examples of monotonically decreasing linear functions, they are able to judge the statement correctly within a short time interval. The second statement covers the concepts of the origin and of a root of a function. If study participants managed to represent the two concepts in their knowledge structures in such a way that they can quickly combine them, they should be able to draw the right conclusions in order to solve the task. Contrariwise, if they fail to relate the two concepts to each other, they may not be sure about their answer to the test item. Lastly, the third test item focuses on the meanings of slopes and intercepts of linear functions: In a first step, the students should recognise that the slopes of the functions f and g are identical, and therefore that the graphs of the two functions are parallel. In a second step, they need to recall the meaning of the intercept of a linear function in order to be able to decide which of the two graphs lies above the other one. Again, if students built helpful knowledge structures on the concepts of slopes and intercepts of linear functions, then it can be assumed that they manage to quickly perform the line of reasoning explained above. Otherwise, they may experience difficulties when approaching the test item without any auxiliary means, as they do not have the possibility to sketch the two graphs for example.

A linear function is always monotonically increasing.

- True
- False

A function whose graph passes through the origin always has at least one root.

- True
- False

The graph of the function $f(x) = \frac{3}{2}x - 3$ lies above the graph of the function $g(x) = \frac{3}{2}x - 1$ in every point.

- True
- False

Figure 5: Examples of test items for a mathematical speed test.

A detailed presentation of the Mathematics Speed Test and its implementation is given in Subsection 4.2.3.

Mathematics Power Test The main idea behind the Mathematics Power Test is to construct a test, which comprises different mathematical problem solving tasks. In particular, its tasks should draw on both, the conceptual an the procedural knowledge of the study participants.

To illustrate why the working memory function Storage-Processing is assumed to play an important role for mathematical problem solving, Figure 6 displays a concrete example of a mathematical task. As explained in Section 2.3, the functional category of Storage-Processing relates to the function of simultaneously maintaining and manipulating information (Oberauer et al., 2000, 2003; Süß et al., 2002). These cognitive processes are needed in order to successfully work on the task given in Figure 6, as it demands the holding of information about the original linear function, while manipulating this linear function in such a way that it meets the requirements for the adjusted linear function. More specifically, subtask a) focuses on changing the slope of the original linear function f, subtask b) asks for the adaptation of the intercept of the original linear function g, and subtask c) aims at shifting the graph of the original linear function h.

Consider the linear functions $f(x) = 2x + 2$, $g(x) = -3x + 1$ and $h(x) = -x - 8$.

a) How should the slope of f be adjusted in order for the graph of f to pass through the point $A(4, 14)$ (assuming that the y-intercept of f remained the same)?
b) How should the y-intercept of g be changed in order for g to have the root $x = 3$ (assuming that the slope of g remained the same)?
c) By how many units in y-direction should the graph of the function h be shifted, in order for h to have the root $x = 5$?

Figure 6: Example of a mathematical task.

The Mathematics Power Test and its implementation are presented in more detail in Subsection 4.2.4.

3.1.4 Questionnaire on Mathematical Self-Concept and Interest in Mathematics

In addition to the different cognitive measures discussed in the foregoing subsections, the goal of the current study is to include motivational variables as well. More specifically, the idea is to assess the mathematical self-concept of the study participants as well as their interest in mathematics. Thomas Braas, a colleague at the chair for Research on Learning & Instruction, assembled a set of established questionnaire items in order to investigate these constructs for his own research project. He kindly shared his selection of questionnaire items for the current study. More information on the questionnaire items on mathematical self-concept and interest in mathematics is provided in Subsection 4.2.1.

3.2 Research Goals

As illustrated in the previous section, the objective of the current study is to further assess the role of working memory in advanced school mathematics. More concretely, the current study aims at investigating a sample consisting of high-achieving, over-achieving and under-achieving adolescents in terms of multiple facets of working memory and of math performance, as well as with respect to their mathematical self-concept and interest in mathematics. To this end, the research goal of the current study is to conduct group comparisons between the different groups of students as well as within the group of high-achievers. The subsequent subsections specify the research questions behind the various group comparisons in more detail and explain some thoughts associated with them.

3.2.1 Group Comparison between High-Achievers and Over-Achievers

The main research aim underlying the group comparison between high-achievers and over-achievers is to gain a deeper understanding of over-achievement in mathematics and its limits. As explained in Subsection 1.2.4, Castejón et al. (2016) investigated possible characteristics of under- and over-achievement and found that learning strategies seem to play an important role. More concretely, their results demonstrated that higher scholastic achievement in over-achieving students can be explained by a major utilisation of learning strategies. In the current study, the goal is to assess the following more specific consideration. A potential explanatory factor for over-achievement in mathematics could be that over-achievers manage to partly compensate the difference in general reasoning abilities by successfully chunking their knowledge and representing it in an advantageous way. Therefore, a special interest lies in the comparison of the two groups with respect to their performances in the Mathematics Speed Test, as the Mathematics Speed Test aims at the knowledge representations of the study participants. Moreover, these results should be contrasted with the respective group comparison related to the Mathematics Power Test, as the Mathematics Power Test covers additional facets of math performance. Furthermore, the aim is to study the group comparison between high-achievers and over-achievers with respect to the various working memory tests as well, in order to gain insights on the potential role of working memory for over-achievement in mathematics. Lastly, to complement the picture, the two groups should also be compared in terms of their mathematical self-concept and interest in mathematics. The different results associated with the group comparison between high-achievers and over-achievers are collected in Section 5.2.

3.2.2 Group Comparison between High-Achievers and Under-Achievers

The research goal behind the group comparison between high-achievers and under-achievers is to further investigate why under-achievers do not exploit their potential in terms of general reasoning abilities, i.e. why their potential is not (fully) translated into math performance. As illustrated in Subsection 1.2.4, different factors were assessed in relation to under-achievement in mathematics, such as mathematics anxiety (e.g., Sepie & Keeling, 1978), motivation (e.g., Fong & Kremer, 2020), stereotype internalisation (e.g., Bonnot & Croizet, 2007) or academic self-concept (e.g., Castejón et al., 2016). Other considerations were also discussed in Subsection 1.2.4. The current study focuses on analysing whether working memory functions could be an additional explanatory variable for under-achievement as well, as working memory and general reasoning ability are distinguishable constructs (see Section 2.4). Moreover, to obtain a more exhaustive picture, the two groups should also be compared with respect to the questionnaire on mathematical self-concept and interest in mathematics. Section 5.3 presents the various results related to the group comparison between high-achievers and under-achievers.

3.2.3 Subgroup Comparison within High-Achievers

The leading research question underlying a subgroup comparison within the group of high-achievers is to assess how top scoring high-achievers with respect to the Mathematics Power Test may differ from the subgroup of lower scoring high-achievers in the Mathematics Power Test. More specifically, the aim is to compare the two subgroups in terms of the various working memory tests, of the Mathematics Speed Test and of the questionnaire on mathematical self-concept and interest in mathematics. The respective results are presented in Section 5.4.

4 Sample and Method

The goal of this chapter is to give an overview on the various procedures of the current study. While Section 4.1 focuses on the sample selection process and the sample itself, Section 4.2 describes the different measures of the current study. Section 4.3 explains the data collection procedure and Section 4.4 briefly comments on the data analysis.

4.1 Sample Selection and Sample

As all students in the sample of the current study formerly participated in the TraM-Study, which is a large scale study on learning mathematics, Subsection 4.1.1 gives a brief introduction to the TraM-Study. In addition, Subsection 4.1.2 describes some of the measures of the TraM-Study, which constitute the basis for the sample selection process of the current study. Subsection 4.1.3 focuses on the sample selection process, while Subsection 4.1.4 comments on the recruitment of the participants for the current study. Finally, Subsection 4.1.5 describes the sample of the current study in more detail.

4.1.1 The TraM-Study

The TraM-Study ("TraM" abbreviates "Transfer in Mathematics") is a large scale study conducted at the chairs for Research on Learning & Instruction and for Mathematics & Education at ETH Zurich. It investigates the learning transfer within mathematics and from mathematics to physics by cognitively activating means of instruction. These means comprise activities such as inventing with contrasting cases (Schalk, Schumacher, Barth, & Stern, 2018; Schwartz, Chase, Oppezzo, & Chin, 2011), comparing and contrasting (Ziegler & Stern, 2014, 2016), self-explanation prompts (Rittle-Johnson, Loehr, & Durkin, 2017; Schworm & Renkl, 2006, 2007) as well as meta-cognitive questions (Zepeda, Richey, Ronevich, & Nokes-Malach, 2015). Among other things, these activities aim at improving students' understanding of mathematical concepts by taking students' prior knowledge up and raising students' awareness about their gaps and misconceptions in current knowledge.

Within the TraM-Study, four separate projects are evaluating the efficacy of providing cognitively activating instruction on students' learning outcomes. For the assessment, all of these projects follow a pre- and posttest design, comparing classes in the experimental group to classes in the control group. While classes in the experimental group work with teaching materials consistently incorporating cognitively activating types of instruction, classes in the control group work with "best practices" materials. Two of the projects focus on students' learning outcomes with respect to the topics of mathematical functions and differential calculus respectively. The goal of the other two projects is to assess transfer effects, either from mathematical functions to kinematics or from mathematical functions to differential calculus. Additional measures of the TraM-Study include a general reasoning ability test and a questionnaire on the socio-economic status of the students.

Participants of the TraM-Study are students in advanced placement schools in Switzerland, i.e. students at the Swiss Gymnasium. The investigated topics of mathematical functions and differential calculus are both core topics of the mathematics syllabus at the Swiss Gymnasium. Within the TraM-Study, teachers of classes in the control group are asked to cover the same mathematical contents as the teachers of classes in the experimental group. Moreover, implementation fidelity measures such as classroom observations are in place in order to monitor the mathematical contents covered in class.

The study participants for the current study formerly participated in the project of the TraM-Study related to mathematical functions. During their participation in the TraM-Study, they covered a complete introduction into the concept of mathematical functions in class, including a detailed discussion of linear functions as well as of quadratic functions. In particular, they were introduced to the notion of a real-valued function, to graphs of functions and to basic properties of functions and their graphs. Lastly, they touched on the topic of inverse functions.

4.1.2 Measures of the TraM-Study

This subsection focuses on two measures of the TraM-Study (see Subsection 4.1.1), namely on the Mathematics Posttest on mathematical functions and on the General Reasoning Ability Test. The sample selection process of the current study, which will be explained in Subsection 4.1.3, was based on these two measures.

Mathematics Posttest on Mathematical Functions The goal of the Mathematics Posttest is to evaluate the learning outcomes regarding the syllabus on mathematical functions covered during the participation in the TraM-Study. To this end, it comprises the following mathematical concepts: Conceptual understanding of functions, domains of definition, image sets, pre-images, graphs of functions and their interpretation, properties of functions and their graphs, linear functions, quadratic functions, roots of functions, angular points of quadratic functions, bijective functions as well as inverse functions and their graphs. The Mathematics Posttest consists of 15 multiple choice tasks (with a maximum point score of 66 points in total), and students are given 40 minutes to complete it. The tasks were developed by Armin P. Barth, a colleague from the MINT Learning Center at ETH Zurich.

General Reasoning Ability Test The General Reasoning Ability Test conducted in the TraM-Study comprises three subtests from the Cognitive Ability Test (Kognitiver Fähigkeits-Test (KFT), Heller & Perleth, 2000), namely a verbal, a quantitative (numerical) and a non-verbal (figural) subtest (V-Test 2, Q-Test 2 and N-Test 1 respectively). While the verbal and the non-verbal subtests both contain 25 test items, the quantitative subtest consists of 20 test items. All test items were designed as single choice tasks with five answer options each, and every test item, which is correctly answered, is graded with one point. Correspondingly, the maximum possible number of points in the verbal and the non-verbal subtests is 25 points respectively, and 20 points in the quantitative subtest. Therefore, the total maximum point score in the General Reasoning Ability Test amounts to 70 points. For each subtest, the time limit is set to nine minutes.

4.1.3 Sample Selection Process

As anticipated in Subsection 4.1.1, the selected sample for the current study exclusively consists of students, who formerly participated in the TraM-Study, namely in the project related to mathematical functions. These students were recruited from former experimental group classes or control group classes. However, as explained in Subsection 4.1.1, they all covered the same syllabus on mathematical functions during their participation in the TraM-Study. It can therefore be assumed that all participants for the current study were familiar with the mathematical contents of the mathematics measures implemented in the current study.

The basis for the sample selection process was given by the two measures of the TraM-Study introduced in the previous subsection: While the Mathematics Posttest on mathematical functions served as a first measure of math performance, the General Reasoning Ability Test was utilised as an indicative measure of general reasoning abilities (see Subsection 4.1.2). To enable the analysis of the interplay between general reasoning abilities, working memory functions and math performance, students belonging to one of the groups listed below were selected for the sample of the current study:

- high-achievers
- over-achievers
- under-achievers

Prior to describing these groups and explaining the corresponding terminology, it is important to remark the following. As mentioned earlier, all participants in the TraM-Study were students at advanced placement schools in Switzerland (Swiss Gymnasium). In the subsequent description of the groups specified above, the term *average* relates to the average score of participants in the TraM-Study, not to the average score of a sample, which would be representative for the respective age group in the general population. In particular, the average score on the General Reasoning Ability Test is higher in the sample of the TraM-Study than in a sample, which would be representative for the respective age group in the general population. For the current study, students with above-average scores in the General Reasoning Ability Test as well as in the Mathematics Posttest are referred to as high-achievers. Students with below-average scores in the General Reasoning Ability Test and above-average scores in the Mathematics Posttest are described as over-achievers, while students with above-average scores in the General Reasoning Ability Test and below-average scores in the Mathematics Posttest are characterised as under-achievers.

Figure 7 depicts the data of 391 students from 25 different classes that were enrolled in the project of the TraM-Study related to mathematical functions when the sample selection for the current study was performed. The mean age in this sample was 16.4 years, and it consisted of 222 female students (57%), 164 male students (42%) and 5 students, who did not specify their gender (1%). More specifically, the scatterplot in Figure 7 visualises the scores in the General Reasoning Ability Test (KFT) against the scores in the Mathematics Posttest (Post). The vertical line and the horizontal line indicate the average score on the General Reasoning Ability Test (51.36) and the average score on the Mathematics Posttest (37.69) respectively.

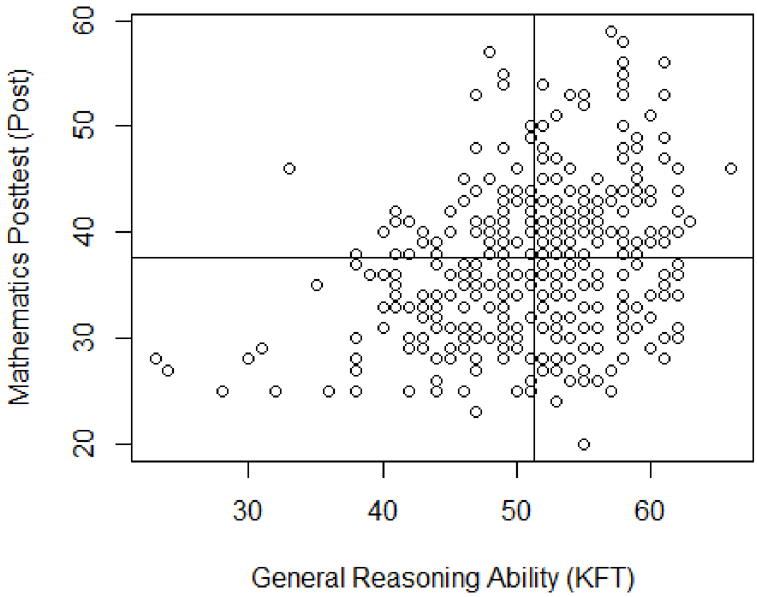


Figure 7: Scatterplot of the data of the TraM-study with respect to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post). The vertical line indicates the average score on the General Reasoning Ability Test (51.36) and the horizontal line indicates the average score on the Mathematics Posttest (37.69).

Following the terminology introduced above, data points in the first (upper right) quadrant represent high-achievers, while data points in the second (upper left) and fourth (bottom right) quadrant refer to over-achievers and under-achievers respectively. In total, 128 high-achievers (62% female, 37% male), 68 over-achievers (50% female, 49% male) and 84 under-achievers (56% female, 43% male) were invited to participate in the current study. The argument for not including data points in the third (bottom left) quadrant, i.e. students with below-average scores in the General Reasoning Ability Test and below-average scores in the Mathematics Posttest, is the following. As explained in more detail in Section 4.2, the difficulty level of one of the mathematics measures implemented in the current study, namely of the Mathematics Power Test, was designed to be rather sophisticated in order to avoid possible ceiling effects in the upper range of performance. On the other hand, the assumption was that noticeable floor effects related to the performance in the Mathematics Power Test would emerge if students with below-average scores in the General Reasoning Ability Test and below-average scores in the Mathematics Posttest were included in the current study. More specifically, the Mathematics Power Test could not be expected to be an adequate mathematics measure for these students.

A detailed description of the sample of the current study as well as the corresponding scatterplot are given in Subsection 4.1.5.

4.1.4 Recruitment of Participants

As explained earlier in this section, the participants for the current study were recruited from the pool of students, who formerly participated in the TraM-Study (see Subsection 4.1.1 for a description of the TraM-Study). For the TraM-Study, several measures of data security were implemented, one of which being that participants of the TraM-Study are not identified by their full names, but by a personal code which remains constant over time. For all participants of the TraM-Study, these personal codes consist of the following elements: The first two letters of their first name, the first two letters of their last name, the first two letters of their mother's or their (female) legal guardian's first name, the first two digits of their date of birth (i.e., the day) and the postal code of their school. Therefore, in the initial phase of the recruitment process it was not possible to contact potential participants of the current study directly. Instead, the mathematics teachers of potential participants were contacted via email and they were asked to forward the information material about the participation in the current study to the selected students (the teachers were able to relate the personal codes to their students, as they know their students' names). While the teachers received information material about the current study as well, they did not obtain any information on their students' cognitive abilities at any time.

In order to take part in the current study, all participants had to sign a respective consent form. In addition, for students under the age of 18 the written consent from their parents or legal guardians was required as well. As a final remark, it should be mentioned that all study participants were rewarded with a monetary compensation for their participation.

4.1.5 Description of the Sample

As explained in Subsection 4.1.3, students belonging to one of the following groups were recruited to participate in the current study: High-achievers, over-achievers or under-achievers. After excluding five students with missing data and three extreme outliers, the final sample for the current study consists of 120 participants. The scatterplot in Figure 8 visualises this sample.

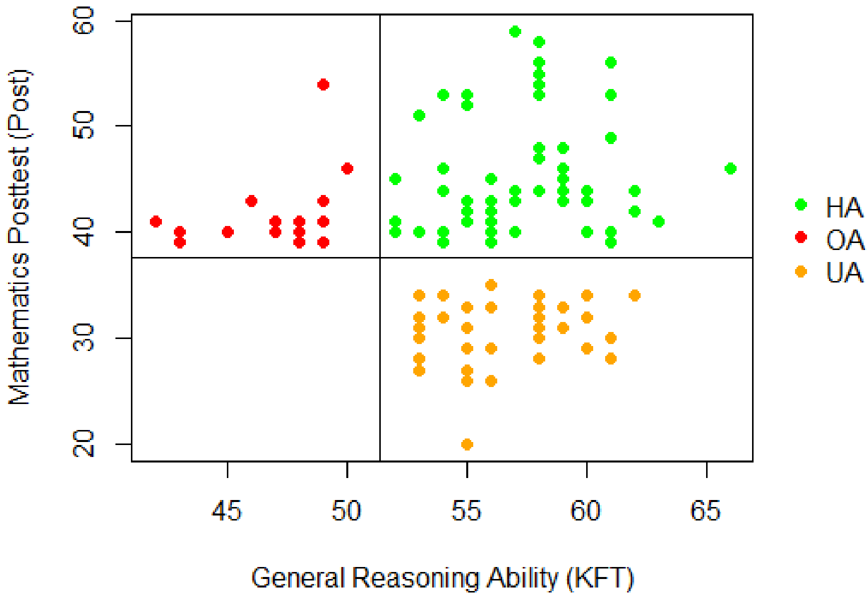


Figure 8: Scatterplot of the data of the whole sample of the current study with respect to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post). The high-achievers are marked with a green dot, the over-achievers with a red dot and the under-achievers with an orange dot. The vertical line indicates the average score for the data of the TraM-study on the General Reasoning Ability Test (51.36) and the horizontal line indicates the average score for the data of the TraM-study on the Mathematics Posttest (37.69).

There are 67 high-achievers, 19 over-achievers and 34 under-achievers in the sample of the current study, while the mean age is 16.3 years. 71 students in the sample of the current study identify as female (59%), 48 students as male (40%), and one student did not specify the gender (1%). The gender ratios within the individual groups are as follows: The group of high-achievers consists of 44 female students (66%), 22 male students (33%) and one student, who did not indicate the gender (1%). 10 over-achieving students (53%) and 17 under-achieving students (50%) are female, while 9 over-achieving students (47%) and 17 under-achieving students (50%) are male. Even though there are more female students than male students in the group of high-achievers, the gender ratios seem to be sufficiently balanced in all three groups. Furthermore, they appear to approximately reflect the respective gender ratios in the groups of invited high-achievers, over-achievers and under-achievers (see Subsection 4.1.3).

4.2 Measures of the Current Study

The aim of this section is to give an overview on the various measures of the current study. While Subsection 4.2.1 focuses on the questionnaire on mathematical self-concept and interest in mathematics, Subsection 4.2.2 explains the different tasks of the working memory test battery and their implementations. A detailed description of the Mathematics Speed Test and of the Mathematics Power Test and their implementations is given in Subsections 4.2.3 and 4.2.4 respectively. To conclude the section, Subsection 4.2.5 comments on the development of the Mathematics Speed Test and of the Mathematics Power Test as well as on the pilot phase of the different tests.

4.2.1 Questionnaire on Mathematical Self-Concept and Interest in Mathematics

To contribute to the coherence of projects embedded within the TraM-Study (a short description of the TraM-Study is given in Subsection 4.1.1), a set of questionnaire items, which was assembled by Thomas Braas (a colleague at the chair for Research on Learning & Instruction), was implemented in the current study. Most of the questionnaire items correspond to items used in *PISA 2000*, as listed in Trautwein et al. (2006). The German translation of these items was utilised in this form before (Kunter et al., 2002). More concretely, the individual items were taken from the following sources:

- Items 1) – 3): Trautwein et al. (2006)
- Item 4): Mang et al. (2018)
- Items 5) – 7): Trautwein et al. (2006)
- Item 8): Gaspard et al. (2015, 2018)

All questionnaire items are structured in the same way: Each item covers a different statement, and the study participants are asked to indicate how strongly they agree or disagree with this statement. More specifically, for each questionnaire item they can choose between the following four options: "disagree", "disagree somewhat", "agree somewhat", "agree". Figure 9 presents two of the questionnaire items: While the upper item refers to the mathematical self-concept of the students, the lower item relates to their interest in mathematics.

I have always done well in mathematics.

- disagree
- disagree somewhat
- agree somewhat
- agree

Mathematics is important to me personally.

- disagree
- disagree somewhat
- agree somewhat
- agree

Figure 9: Examples of questionnaire items in the questionnaire on mathematical self-concept and interest in mathematics. The items were taken from Trautwein et al. (2006).

The German version as well as the English version of the entire questionnaire are collected in the appendix (see Sections A.6 and A.7 respectively). The study participants of the current study filled in the German version of the questionnaire.

For the evaluation of the questionnaire data, the answers of the study participants to all questionnaire items except to the fourth item were recoded numerically as follows: "disagree" → 1, "disagree somewhat" → 2, "agree somewhat" → 3, "agree" → 4. For the fourth item, the numerical recoding was inverted in the following way, as the statement in the fourth item is negatively formulated: "disagree" → 4, "disagree somewhat" → 3, "agree somewhat" → 2, "agree" → 1.

4.2.2 Working Memory Test Battery

As motivated in Subsection 3.1.2, a working memory test battery arranged by von Bastian and Oberauer (2013) was chosen for the current study, as its tasks cover all functional categories of working memory suggested by the facet model (Oberauer et al., 2000, 2003; Süß et al., 2002, a brief overview on the facet model is given in Section 2.3). Dr. Claudia von Bastian is a Senior Lecturer in Psychology at the University of Sheffield, her research interests include cognitive individual differences and the question how cognitive abilities can change through experience (*Cognitive Ability & Plasticity Lab*, n.d.). Dr. Claudia von Bastian kindly supported the current study by sharing her templates of the working memory test battery. Closely following von Bastian and Oberauer (2013), this subsection aims at describing and explaining the various tasks of this working memory test battery as well as their implementation.

As a measure for the functional category of Storage-Processing, von Bastian and Oberauer (2013) developed dual tasks based on the Brown-Peterson paradigm (Brown, 1958). These tasks consist of two elements: A simple span task and a distracting decision task. More concretely, von Bastian and Oberauer (2013) explained the procedure as follows: The starting point for each dual task is the presentation of a list of items, which should be memorised within a limited time interval. During the second phase of the dual task, whose duration is fixed to five seconds, participants perform multiple trials of the distracting task. For each trial of the distracting task, a stimulus (shown centrally on the screen) as well as four alternatives (simultaneously displayed in the four corners of the screen) are presented on the screen. Participants should then decide as quickly and at the same time as correctly as possible which of the four alternatives belongs to the stimulus with respect to a certain rule. After completing the trials of the distracting task, participants are asked to recall the items, which were originally displayed on the list, in correct order. For the dual tasks with numerical content, von Bastian and Oberauer (2013) used two-digit numbers as stimuli, while they utilised words as stimuli for the verbal dual tasks (see Figures 2 and 3 in the foregoing chapter for a concrete example of such a verbal dual task). Lastly, they employed cells of 3 x 3 partially filled matrices as stimuli for the figural dual tasks. von Bastian and Oberauer (2013) structured the respective sequences of the trials as follows: First, participants are given the chance to perform three practice trials. Afterwards, they complete three trials per set size of items on the memory list (varying from two to eight items to be recalled) in three pseudo-randomised blocks. For the evaluation of the performance on dual tasks, the accuracy on each task is measured by the proportion of correctly recalled items.

To measure the functional category of Relational Integration, von Bastian and Oberauer (2013) included monitoring tasks developed by Oberauer et al. (2003) in the test battery. As described in von Bastian and Oberauer (2013), participants are assigned with the following task: They observe certain objects on the screen, which change independently from each other, and they are supposed to press a given key whenever a critical relation between these objects occurs. In the figural task for example, participants monitor dots in a 10 x 10 matrix, and they should press the key whenever "four dots form a square" (von Bastian & Oberauer, 2013, Table 4). Figure 10 visualises two examples of such matrices: While there are four dots forming a square in the matrix on the left hand side, no such square can be detected in the matrix on the right hand side. Each monitoring task comprises 16 runs, while it takes $2 - 8$ changes per run for the critical relation to occur. For the numerical monitoring task, three-digit numbers serve as objects, which are arranged in a 3 x 3 matrix, and the critical condition refers to the occurrence where "three numbers in a row, in a column, or along a diagonal have identical last digit" (von Bastian & Oberauer, 2013, Table 4). Similarly, the monitoring task with verbal content uses words in a 3 x 3 matrix as objects, while the critical condition is fulfilled whenever "three words in a row, in a column, or along a diagonal are rhyming" (von Bastian & Oberauer, 2013, Table 4). To analyse the performance on a monitoring task, the respective detection performance is computed.

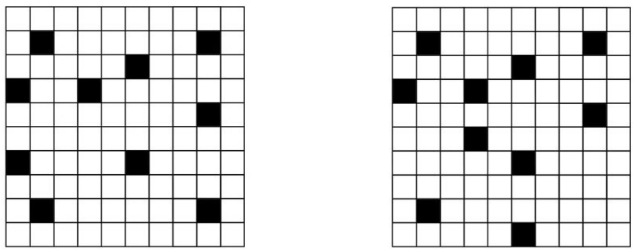


Figure 10: Figural monitoring task described in von Bastian and Oberauer (2013).

As explained in von Bastian and Oberauer (2013), the functional category of Supervision is measured by task switching tasks based on the task switching paradigm (Monsell, 2003). The core of these tasks is that participants should categorise bivalent stimuli as quickly and at the same time as correctly as possible. More specifically, von Bastian and Oberauer (2013) chose the following implementation of task switching tasks: After completing eight practice trials, the study participants perform 64 trials, which are ordered in a pseudo-randomised way, while the classification rule for the categorisation switches in alternating runs of two. For the numerical task switching task, the two classification rules are associated with categorising a given number either as odd or even, or as smaller or larger than 500 respectively. Figures 11 and 12 illustrate the classification rules for the numerical task switching task. For the task switching task with verbal content, participants should either classify a given word as *animal* or *plant*, or decide whether the word contains one or two syllables. For the figural task switching task, participants are either asked to categorise a given pattern as symmetrical or asymmetrical, or to decide whether the pattern consists of one or two parts. As a measure of performance on a task switching task, the corresponding switch costs are calculated.

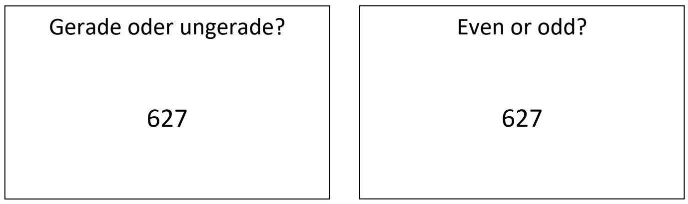


Figure 11: Numerical task switching task described in von Bastian and Oberauer (2013) for the categorisation of a number as being even or odd. The German version is the original version of the task.

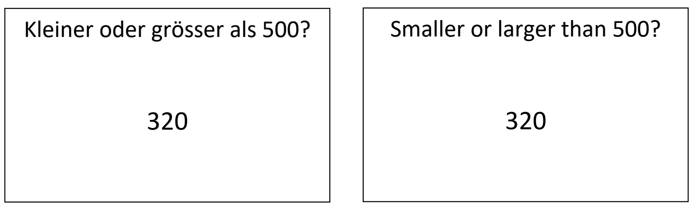


Figure 12: Numerical task switching task described in von Bastian and Oberauer (2013) for the categorisation of a number as being smaller or larger than 500. The German version is the original version of the task.

To conduct the working memory test battery proposed by von Bastian and Oberauer (2013) in the current study, a computer implementation of the tasks described above was used (based on Dr. Claudia von Bastian's templates). For the testing procedure in the current study, the working memory test battery was divided into three test batches, where each batch comprises three subtests. More specifically, each batch represents a functional category of working memory as described by the facet model (Oberauer et al., 2000, 2003; Süß et al., 2002, a short presentation of the facet model is provided in Section 2.3), and it contains the respective figural, numerical and verbal subtest. The following enumeration shows the sequence of the tests.

- **Relational Integration**
  1. Monitoring Task – figural
  2. Monitoring Task – numerical
  3. Monitoring Task – verbal
- **Supervision**
  4. Task Switching Task – numerical
  5. Task Switching Task – verbal
  6. Task Switching Task – figural
- **Storage-Processing**
  7. Dual Task – verbal
  8. Dual Task – figural
  9. Dual Task – numerical

During the testing procedure in the current study, there was a 10- minutes break for the participants between two different test batches. Moreover, there were 2-minutes breaks between two different subtests within the same test batch.

4.2.3 Mathematics Speed Test

The mathematical contents for the Mathematics Speed Test were taken from the syllabus on mathematical functions covered by the TraM-Study, as the participants for the current study formerly participated in the respective project of the TraM-Study (see Subsection 4.1.1). The Mathematics Speed Test aims at the knowledge representations of the students. More specifically, the main idea behind the Mathematics Speed Test is that students, who successfully chunked their knowledge on mathematical functions, are able to solve the tasks correctly within the given time limit.

The Mathematics Speed Test comprises 45 test items in total. Each test item consists of a statement related to mathematical functions, which is either true or false. Accordingly, for each statement the task is to decide as quickly and at the same time as correctly as possible whether it is true or false. However, participants were instructed to prioritise correctness over quickness. In addition, participants were informed that, unless stated otherwise, they should judge all given statements based on the consideration of real numbers (and not of complex numbers). Figure 13 shows some examples of test items in the Mathematics Speed Test. The test items in Figure 5, which were discussed in the previous chapter, were also taken from the Mathematics Speed Test.

It is possible that the graph of a linear function and the graph of a quadratic function do not intersect.

- True
- False

If the graph of a linear function passes through the points $A(2,4)$ and $B(4,8)$ then the point $C(3,6)$ also lies on this graph.

- True
- False

The graphs of the linear functions $f(x) = 2x - 1$ and $g(x) = \frac{1}{2}x + 2$ are orthogonal.

- True
- False

All values of the function $g(x) = -2(x+2)^2 + 2$ are less than or equal to 2.

- True
- False

Figure 13: Examples of test items in the Mathematics Speed Test.

The study participants were not allowed to use any auxiliary means to work on the Mathematics Speed Test. In particular, they were not permitted to use a separate notepad or another tool to make any additional computation steps. All test items were designed in such a manner that they can be solved without additional computation steps. However, this requires well chunked and properly represented knowledge on mathematical functions, which is in line with the objective of the Mathematics Speed Test.

For the Mathematics Speed Test, a computer-based implementation was chosen. A time limit of 45 seconds per task was set, as the pilot testing of the Mathematics Speed Test showed that this time limit seemed to be adequate (more details on the pilot testing can be found in Subsection 4.2.5). Prior to taking the Mathematics Speed Test, participants were informed about this time limit on task-level and they were told that the respective time interval should be sufficiently long to solve the task.

Every correctly answered test item is graded with one point, which adds up to a total maximum point score of 45 points in the Mathematics Speed Test.

The final version of the Mathematics Speed Test and its English translation are collected in the appendix (see Sections A.8 and A.9 respectively). The study participants of the current study were given the German version of the Mathematics Speed Test.

4.2.4 Mathematics Power Test

As explained in Subsection 4.1.1, the participants for the current study covered a syllabus on mathematical functions during their participation in the TraM-Study. The mathematical contents for the Mathematics Power Test were therefore taken from this syllabus. The aim of the Mathematics Power Test is to measure the mathematical problem solving skills of the students, and its tasks draw on the conceptual as well as on the procedural knowledge of the students.

The Mathematics Power Test consists of six more complex mathematical tasks with three subtasks each. Some of these tasks were designed to be rather sophisticated with respect to their difficulty level. The idea behind this setting was to develop tasks, which differentiate between individual performances in the upper range of performance as well, as the group of high-achievers was one of the investigated groups of students for the current study (see Subsection 4.1.3 for a description of the different groups). Two of the tasks in the Mathematics Power Test are collected in Figure 14. The mathematical task presented in Figure 6 in the previous chapter is also included in the Mathematics Power Test.

Consider the functions $f(x) = 2x^2 - 16x + 24$, $h(x) = -\frac{1}{2}x - 6$ and $k(x) = x + 8$.

a) What is the functional equation of the function g, whose graph is the reflection of the graph of f along the x-axis?
b) By how many units in the x-direction should the graph of f be shifted in order for f to have the angular point $S(8, -8)$? What is the functional equation of the shifted graph?
c) How should the y-intercept of k be adjusted in order for the graphs of the functions h and k to intersect at the point $B(-4, -4)$ (assuming that the slope of k remained the same)?

The graphs of the linear functions f and g are orthogonal to each other and intersect at exactly one point, namely at the point $P(1, 3)$. The function f is strictly monotonically increasing, while the function g is strictly monotonically decreasing. The function f has a root at $x = \frac{1}{4}$ and is positive over the interval $[1,10]$. The inverse function of the function g has the y-intercept 13. The graph of the linear function k is parallel to the graph of the function f and intersects the x-axis at $x = -1$. The function k has a positive y-intercept and, together with the x-axis and the y-axis, its graph encloses a triangle of area 2.

a) The graphs of the functions f and g enclose a triangle together with the x-axis. Determine the area of this triangle.
b) Determine the functional equation of k.
c) Determine the intersection point between the graph of the function f and the graph of the inverse function of f.

Figure 14: Examples of mathematical tasks in the Mathematics Power Test.

In contrast to the computer-based implementation of the Mathematics Speed Test, a paper-pencil implementation was chosen for the execution of the Mathematics Power Test. The main argument for this choice is the following: The complexity of the tasks in the Mathematics Power Test demands additional notes, computation steps, and possibly geometrical sketches on a separate notepad for the mathematical problem solving. Accordingly, participants were allowed to use the provided coordinate systems, a separate notepad, a set square and/or a ruler as auxiliary means for solving the Mathematics Power Test. On the other hand, they were not permitted to utilise a calculator or a formulary. However, they were given the advice to work with fractions rather than decimal numbers in their calculations.

There was a time limit of two hours for the Mathematics Power Test as a whole. This time limit was set based on the empirical values determined during the piloting phase of the Mathematics Power Test (see Subsection

4.2.5). It was specified in such a way that the time pressure was rather low, and participants were instructed to prioritise correctness over quickness. In particular, they were asked to try to work on all tasks of the Mathematics Power Test, and to accurately document the respective solution approaches. In addition, they were told that they had to justify all solutions algebraically, and that geometrical considerations were not sufficient for the justification of a solution. Participants could freely choose the order in which they dealt with the different tasks.

The maximum point score for each task in the Mathematics Power Test is 3 points (1 point for every subtask, which is correctly solved). The maximum possible number of points in the Mathematics Power Test is therefore 18 points.

For the sample of the current study, the Mathematics Power Test yielded a standardised Cronbach's alpha of 0.88, indicating a good reliability. Moreover, the item whole correlations for all six items of the Mathematics Power Test were above 0.58, while the ones for the items 1, 4 and 6 were above 0.73.

The final version of the Mathematics Power Test and its English translation are given in the appendix (see Sections A.10 and A.11 respectively). The German version of the Mathematics Power Test was administered to the study participants of the current study.

4.2.5 Pilots of the Tests

The aim of this subsection is to comment on the development of the Mathematics Speed Test and of the Mathematics Power Test and to give an overview on the pilot phase of the different tests.

Development of the Mathematics Speed Test A primary paper-pencil version of the Mathematics Speed Test contained 46 test items. This version was piloted with 19 students at the Gymnasium in order to obtain first estimates of the response times of the students and to check whether students understood the tasks correctly. Based on these estimates of the response times, the goal was to find an adequate fixed time interval per item in such a way that the large majority of the participants in the current study would not exceed this time limit (to avoid a possible negative effect of such a time limit on the validity of the Mathematics Speed Test). Moreover, the pilot testing revealed that one of the test items seemed to be ambiguous to some students. Therefore, this item was subsequently removed from the Mathematics Speed Test. On the other hand, the pilot testing showed that the majority of the students understood the remaining 45 test items correctly. In addition to the pilot testing at the Gymnasium, the Mathematics Speed Test was reviewed by two graduate students in mathematics, a mathematics teacher at the Gymnasium and a professor in mathematics. These experts confirmed that none of the remaining 45 test items was formulated in an ambiguous way, and that for each test item it was ensured that the solution was unique. The final test runs related to the Mathematics Speed Test, which aimed at testing the computer-based implementation of the Mathematics Speed Test and validating the chosen time limit on task-level, will be described in the last paragraph of this subsection.

Development of the Mathematics Power Test The original version of the Mathematics Power Test (comprising nine tasks with three subtasks each) was piloted with 13 students at the Gymnasium. As expected, these students perceived the tasks in the Mathematics Power Test as being rather sophisticated with respect to their difficulty level. However, the formulation of all tasks appeared to be comprehensible for these students, and they were familiar with the mathematical contents on mathematical functions covered by the tasks. In addition, two graduate students in mathematics, two mathematics teachers at the Gymnasium and a professor in mathematics reviewed the Mathematics Power Test and assessed its understandability. They agreed that all tasks in the Mathematics Power Test were clearly formulated and they also judged the difficulty level as being rather high for students at the Gymnasium. Based on their feedback and the insights from the pilot testing at the Gymnasium, the Mathematics Power Test was revised and some of the features were adapted in order to increase the explanatory power of it. A description of the final pilot phase of the Mathematics Power Test will be included in the next paragraph.

Pilot Study In addition to the preliminary pilot testings described in the previous paragraphs, a more extensive pilot study was conducted in the final stage of the pilot phase. The aim of this study was to thoroughly pilot the testing procedure of the current study as a whole. To this end, 25 first year students at ETH Zurich, who were enrolled in the bachelor studies of mechanical engineering, were recruited to participate in the pilot study. The registrations for participation in the pilot study were considered on a first-come-first-serve basis. Five women and twenty men participated in the pilot study, while their mean age was 20 years. All participants of the pilot study were rewarded with a monetary compensation for their participation.

The data collection for the pilot study took place in two different sessions, which were one week apart from each other. The first session was fully dedicated to the Mathematics Power Test: The participants could work on the Mathematics Power Test during a time interval of three hours. The Mathematics Speed Test as well as the working memory test battery were conducted during the second session of the pilot study, which took place in the computer lab. As a first task of the second session, the pilot study participants completed the computer-implemented version of the Mathematics Speed Test. Afterwards, they took a 20-minutes break before proceeding with the various working memory tests.

For the working memory test battery, the schedule described in Subsection 4.2.2 was followed (including the respective breaks). The most important results from the data analysis regarding the pilot study as well as some of the conclusions drawn from it are discussed in the following. For the working memory test battery, the only distinctive feature that could be observed was a slight ceiling effect associated with the figural dual task. The data related to all the other working memory tests showed a good amount of variance. In addition, Claudia von Bastian (see Subsection 4.2.2) kindly shared her data, so that it could be used as benchmark data for the data analysis of the pilot study. Except for the figural dual task, the distributions and descriptive statistics of her data were comparable to the ones from the pilot study data. Given these results, it seemed reasonable to use the same implementation of the working memory test battery in the current study as in the pilot study.

With respect to the Mathematics Speed Test, the data of the pilot study indicated a good amount of variance. The distribution was slightly left-skewed, but there was not a major ceiling effect. As an additional check of the quality of the test items, a data analysis in terms of the item response theory was performed. This analysis yielded a few 'conspicuous' test items. However, all these results could be explained, so that it was not necessary to exclude any of the items. The second part of the evaluation of the Mathematics Speed Test data concerned the imposed time limit on task-level. As explained in the first paragraph of this subsection, the aim was that the large majority of the participants should not exceed this time limit. The data of the pilot study showed that none of the pilot study participants had exceeded the time limit of 45 seconds for any of the test items. The majority of the participants had also not exceeded the time interval of 40 seconds for any of the test items, and some participants had exceeded the time interval of 30 seconds for a very small number of test items. Based on these results, the following decisions were taken with respect to the implementation of the Mathematics Speed Test in the current study: None of the test items were adapted or excluded from the Mathematics Speed Test, and the time limit on task-level was fixed to 45 seconds.

Regarding the Mathematics Power Test, a first important observation was that for most pilot study participants, the time pressure during the execution of the Mathematics Power Test was not high (as intended). For the analysis of the data of the pilot study related to the Mathematics Power Test, the main focus lay on assessing how the Mathematics Power Test could be shortened in such a way that it could be executed within two hours (without a high time pressure), while maintaining a good explanatory power. To this end, reliability measures as well as item whole correlations were computed as a basis for decision-making. For the final version, the Mathematics Power Test was shortened by a third (content- and timewise). More concretely, the final version contained six tasks with three subtasks each (instead of the original nine tasks with three subtasks each), and participants in the current study were given two hours to work on the Mathematics Power Test (instead of the time interval of three hours in the pilot study).

As a concluding remark it should be added that based on the experience in the pilot study, the instructions related to the various tests were revised and complemented for the data collection procedure of the current study.

4.3 Procedure

All study participants attended a single test day, and they had the opportunity to choose from a couple of proposed dates. All test days took place either on Saturdays or during school holidays, so that participants did not miss any classes at school.

Prior to the chosen test day, each participant received detailed information about the test day. In addition to organisational information, the information material also included the following preparatory task for the participants. They received a list of important mathematical terms related to the topic of mathematical functions and were asked to revise these terms with respect to their definitions and meanings. The main goal of this preparatory task was to ensure that all participants understood the formulation of the tasks in the Mathematics Speed Test (presented in Subsection 4.2.3) as well as in the Mathematics Power Test (described in Subsection 4.2.4). The list of mathematical terms for the preparatory task and its English translation can be found in the appendix (see Sections A.12 and A.13 respectively).

As a first task on the test day itself, participants were asked to complete the questionnaire on mathematical self-concept and interest in mathematics (see Subsection 4.2.1). In this way, their answers on the questionnaire would not be influenced by their test experiences during the remaining time on the test day. In addition to the questionnaire, the test day comprised the execution of the working memory test battery (see Subsection 4.2.2), of the Mathematics

Speed Test (see Subsection 4.2.3) and of the Mathematics Power Test (see Subsection 4.2.4).

It was guaranteed that there was a sufficient number of breaks (including a lunch break) during each test day, so that participants could rest in between different tests. Moreover, it was assured that participants received detailed and extensive instructions for the questionnaire as well as for all the tests, and that they always had the opportunity to ask questions if something was unclear to them. At the same time, participants were supervised during all test procedures, to ensure that they worked on the given tasks in the intended way.

As the data collection took place during the Covid-19 pandemic, the following final remark should be added. During all test days, several safety and hygiene measures were in place in order to protect the health of the study participants.

4.4 Data Analysis

All analyses, which will be presented and discussed in the following chapter, have been carried out using the statistical software R, R version 4.0.3 (R Core Team, 2020; RStudio Team, 2021). In particular, in addition to several standard R commands, the following R packages have been applied (the R packages are given in alphabetical order): dplyr (Wickham, François, Henry, & Müller, 2021), effsize (Torchiano, 2020), GGally (Schloerke et al., 2021), ggplot2 (Wickham, 2016), gridExtra (Auguie, 2017), MatchIt (Ho, Imai, King, & Stuart, 2011), optmatch (Hansen & Klopfer, 2006), psych (Revelle, 2020).

5 Results

This chapter is structured as follows: Section 5.1 aims at giving a first presentation of the data by depicting several scatterplots, discussing the corresponding correlations, and illustrating the distributions of the data. The focus of Section 5.2 lies on the group comparison between high-achievers and over-achievers, while Section 5.3 discusses the group comparison between high-achievers and under-achievers. Lastly, Section 5.4 presents a subgroup comparison within the group of high-achieving students.

5.1 Descriptive Statistics of the Sample of the Current Study

The overarching goal of this section is to give a first presentation of the data. To this end, various scatterplots are displayed together with the respective correlations. Moreover, the distributions of the data are visualised. As an initial remark for this section, Subsection 5.1.1 tries to illustrate the influence of the highly selective sample selection process (described in Subsection 4.1.3) on correlations. Subsection 5.1.2 then focuses on the presentation of the data associated with the working memory tests. In addition, the relations between the different working memory tests and the General Reasoning Ability Test are visualised in Subsection 5.1.3. Lastly, Subsection 5.1.4 relates the data on the Self-Concept Questionnaire Items and on the Interest Questionnaire Items to the data on the Mathematics Power Test and on the Mathematics Speed Test.

5.1.1 Consequences of a Highly Selective Sample Selection Process

Figure 15 depicts the scatterplot of the TraM-study data on the General Reasoning Ability Test (KFT) and on the Mathematics Posttest (Post) once again, which was discussed in Subsection 4.1.3. As already explained in Subsection 4.1.1, the sample of the TraM-study is selective itself, as it exclusively consists of students at advanced placement schools (Swiss Gymnasium). Subsection 4.1.3 then described the highly selective sample selection process for the current study based on the sample of the TraM-study shown in Figure 15. For a better overview, the scatterplot of the sample for the current study is also displayed once again in Figure 16.

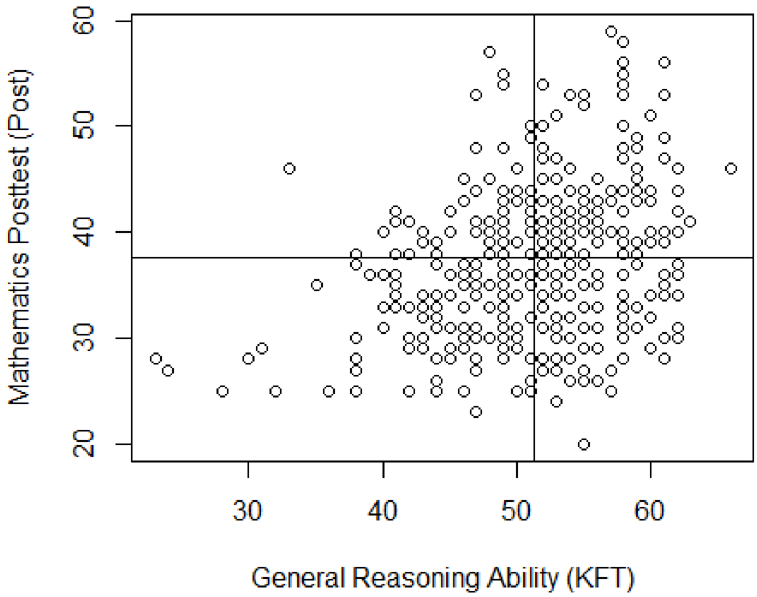


Figure 15: Scatterplot of the data of the TraM-study with respect to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post). The vertical line indicates the average score on the General Reasoning Ability Test (51.36) and the horizontal line indicates the average score on the Mathematics Posttest (37.69). Note that this figure is identical to Figure 7 in Chapter 4.

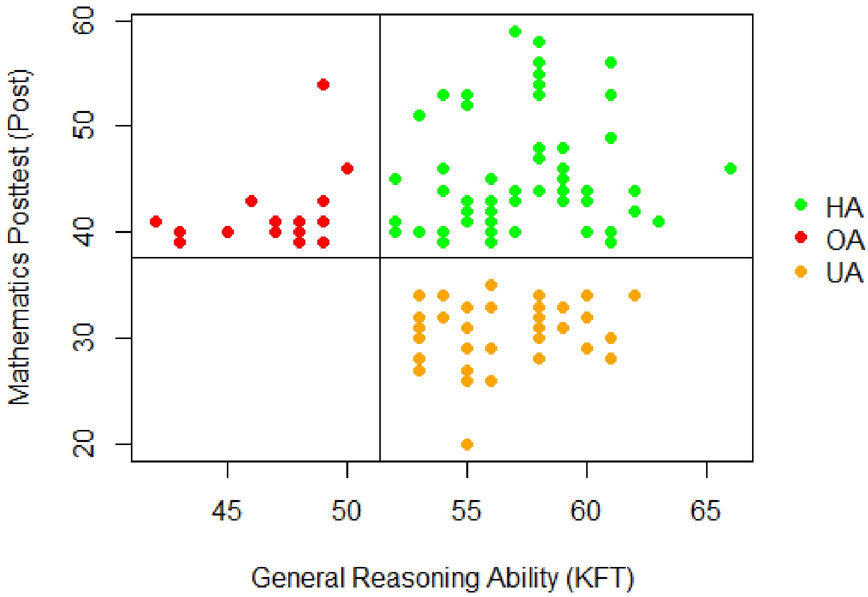
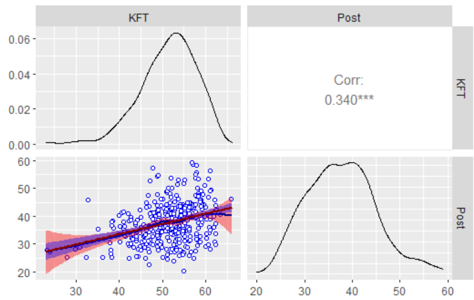
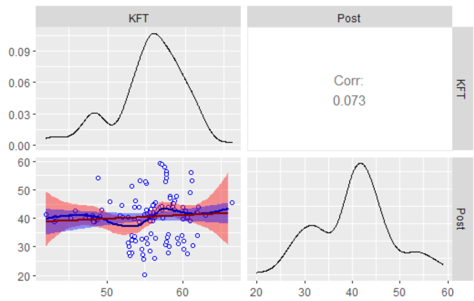


Figure 16: Scatterplot of the data of the whole sample of the current study with respect to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post). The high-achievers are marked with a green dot, the over-achievers with a red dot and the under-achievers with an orange dot. The vertical line indicates the average score for the data of the TraM-study on the General Reasoning Ability Test (51.36) and the horizontal line indicates the average score for the data of the TraM-study on the Mathematics Posttest (37.69). Note that this figure is identical to Figure 8 in Chapter 4.

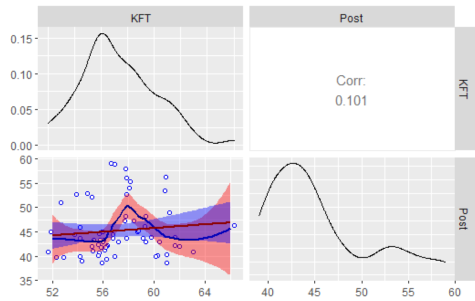
In the following, the goal is to compare and contrast the correlations between the General Reasoning Ability Test and the Mathematics Posttest for the sample of the TraM-study on the one hand, and for the sample of the current study as well as for the individual groups on the other hand. Figure 17a depicts the scatterplot together with the respective correlation as well as the distributions of the data on the General Reasoning Ability Test (KFT) and on the Mathematics Posttest (Post) for the sample of the TraM-study, while Figure 17b shows the corresponding plot for the whole sample of the current study. The analogous plots for the high-achievers, the over-achievers and the under-achievers are collected in Figures 17c, 17d and 17e respectively.



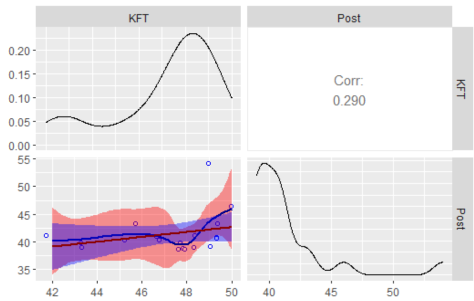
(a) Data of the **TraM-study**



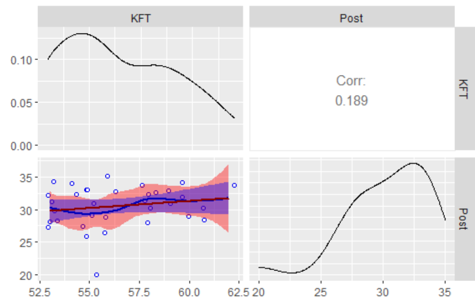
(b) Data of the **whole sample**



(c) Data of the **high-achievers**



(d) Data of the **over-achievers**



(e) Data of the **under-achievers**

Figure 17: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the **General Reasoning Ability Test (KFT)** and the **Mathematics Posttest (Post)**.

A first important observation is that while the positive correlation between the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post) is statistically significant for the sample of the TraM-study, it seems to disappear for the whole sample of the current study. This finding illustrates the effect of the highly selective sample selection process (described in Subsection 4.1.3) on the correlation between the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post). Similarly, the respective correlations for the high-achievers, the over-achievers and the under-achievers are not statistically significant either. Regarding the correlations between the individual subtests of the General Reasoning Ability Test (KFT_V, KFT_Q and KFT_N) among each other as well as the correlations between these subtests and the Mathematics Posttest (Post), a comparable pattern emerges: While all these correlations are positive and statistically significant for the sample of the TraM-study (ranging from 0.186 to 0.313), the respective correlations seem to vanish for the whole sample of the current study (ranging from -0.065 to 0.136, none of them being statistically significant). Again, this observation visualises the consequences of the highly selective sample selection process (described in Subsection 4.1.3) on correlations.

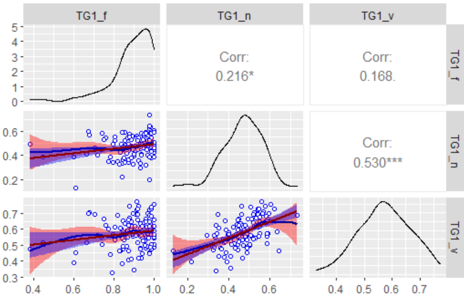
Concluding Remarks To conclude Subsection 5.1.1, the focus of this paragraph lies on the following remark. As illustrated and explained in the previous paragraph, the correlations for the whole sample of the current study as well as for the individual groups of high-achievers, over-achievers and under-achievers are influenced by the very selective sample selection process. It is therefore of utmost importance to keep these observations in mind when looking at the correlations presented in the following subsections. In particular, these correlations should be interpreted with caution.

Lastly, the following remark regarding the terminology of the data sets should be added: From the subsequent subsection on, the term "(whole) sample" will always refer to the whole sample of the current study and not to the sample of the TraM-study. As explained in Sections 4.2 and 4.3, the participants of the current study underwent additional tests and questionnaire items, which were not part of the TraM-study. The scatterplots together with the respective correlations as well as the distributions on these measures are provided and discussed in the following subsections.

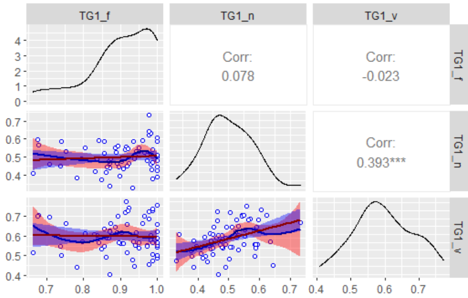
5.1.2 Working Memory Tests

The first part of this subsection focuses on the working memory function Storage-Processing, while the second part investigates the functional category of Supervision. Finally, the data on the tests associated with the working memory function Relational Integration is visualised in the third part of this subsection.

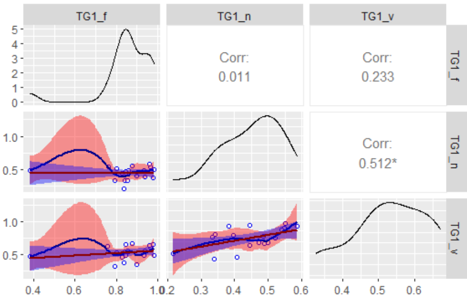
Functional Category of Storage-Processing The data on the working memory tests associated with the functional category of Storage-Processing (TG1) is presented in the following figures. Figure 18a shows the scatterplots together with the respective correlations as well as the distributions of the data on the figural (TG1_f), numerical (TG1_n) and verbal (TG1_v) subtest for the whole sample. Figure 18b visualises the corresponding plots for the high-achievers, while Figure 18c focuses on the over-achievers. Lastly, the data referring to the under-achievers is displayed in Figure 18d.



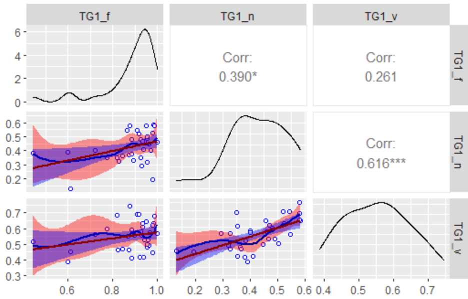
(a) Data of the **whole sample**



(b) Data of the **high-achievers**



(c) Data of the **over-achievers**

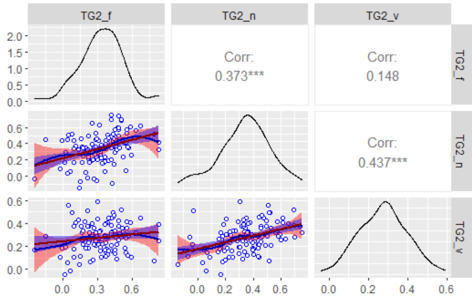


(d) Data of the **under-achievers**

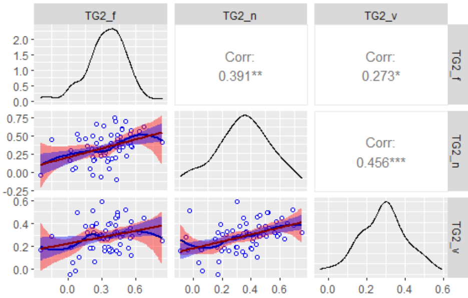
Figure 18: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the figural (TG1_f), numerical (TG1_n) and verbal (TG1_v) subtest associated with the working memory function **Storage-Processing**.

The most outstanding finding is that for all four data sets, the positive correlation between the numerical (TG1_n) and the verbal (TG1_v) subtest is statistically significant. Moreover, the positive correlation between the figural (TG1_f) and the numerical (TG1_n) subtest appears to be statistically significant for the whole sample as well as for the under-achievers. In terms of the distributions of the data, slight ceiling effects can be observed for the figural subtest (TG1_f). On the other hand, the distributions of the scores on the numerical subtest (TG1_n) and on the verbal subtest (TG1_v) seem to be approximately symmetrical for the whole sample as well as for the high-achievers.

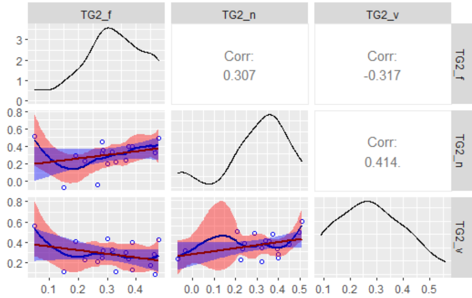
Functional Category of Supervision The following figures visualise the data on the tests related to the working memory function Supervision (TG2). The scatterplots together with the respective correlations as well as the distributions of the data on the figural (TG2_f), numerical (TG2_n) and verbal (TG2_v) subtest for the whole sample are displayed in Figure 19a. While Figure 19b collects the corresponding plots for the high-achievers, Figure 19c presents the data associated with the over-achievers. Finally, Figure 19d visualises the data related to the under-achievers.



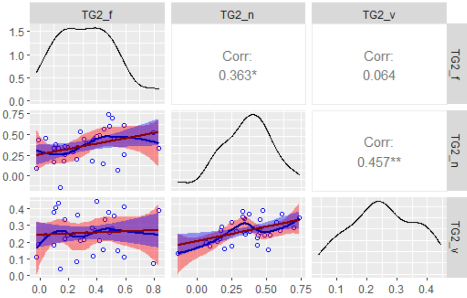
(a) Data of the **whole sample**



(b) Data of the **high-achievers**



(c) Data of the **over-achievers**



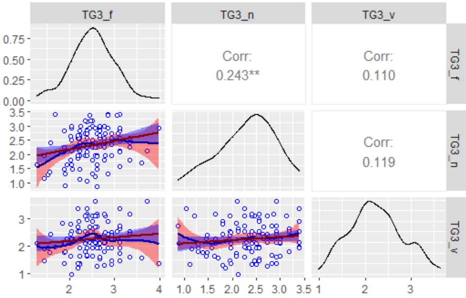
(d) Data of the **under-achievers**

Figure 19: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the figural (TG2_f), numerical (TG2_n) and verbal (TG2_v) subtest associated with the working memory function **Supervision**.

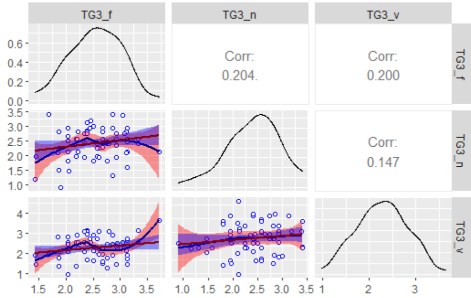
The positive correlations between the figural (TG2_f) and the numerical (TG2_n) subtest as well as the ones between the numerical (TG2_n) and the verbal (TG2_v) subtest appear to be statistically significant for all data sets except for the one of the over-achievers. In addition, it can be observed that the positive correlation between the figural (TG2_f) and the verbal (TG2_v) subtest is statistically significant for the high-achievers. Lastly, it can be noted that the distributions of the scores on all three subtests (TG2_f, TG2_n and TG2_v) seem to be roughly symmetrical for the whole sample as well as for the high-achievers.

Functional Category of Relational Integration The presentation of the data regarding the functional category of Relational Integration (TG3) is given in the following figures. Figure 20a depicts the scatterplots together with the respective correlations as well as the distributions of the data on the figural (TG3_f), numerical (TG3_n) and verbal (TG3_v) subtest for the whole sample. The corresponding plots for the high-achievers are shown in

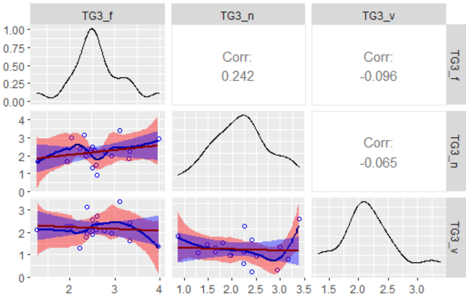
Figure 20b, while Figure 20c displays the data related to the over-achievers. Lastly, the data associated with the under-achievers is presented in Figure 20d.



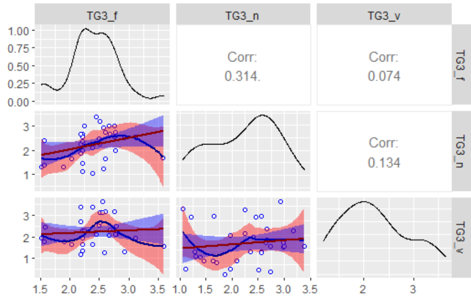
(a) Data of the **whole sample**



(b) Data of the **high-achievers**



(c) Data of the **over-achievers**



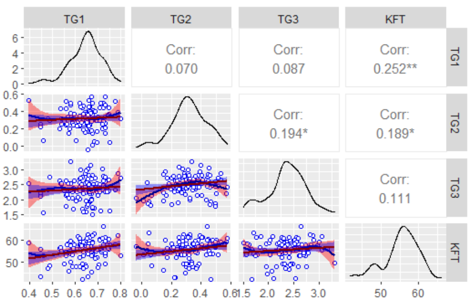
(d) Data of the **under-achievers**

Figure 20: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the figural (TG3_f), numerical (TG3_n) and verbal (TG3_v) subtest associated with the working memory function **Relational Integration**.

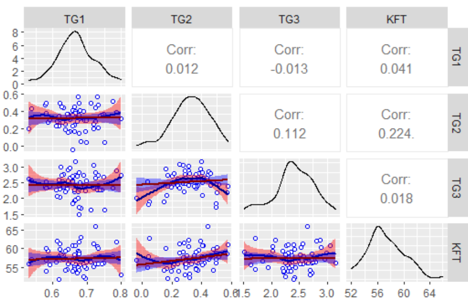
As a first observation, it can be noted that the positive correlation between the figural (TG3_f) and the numerical (TG3_n) subtest is statistically significant for the whole sample, while the corresponding correlations could be relevant for the high-achievers and the under-achievers as well. For the under-achievers, the distribution of the scores on the numerical subtest (TG3_n) seems to be slightly left-skewed, while the one associated with the verbal subtest (TG3_v) appears to be slightly right-skewed.

5.1.3 Working Memory Tests and General Reasoning Ability Test

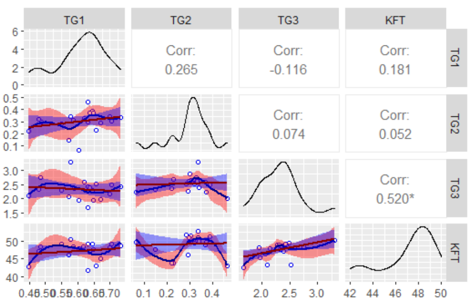
This subsection aims at exploring the correlations between the mean scores of the working memory tests related to the functional categories of Storage-Processing, Supervision and Relational Integration as well as the sum scores of the General Reasoning Ability Test. Figure 21a collects the scatterplots together with the respective correlations and the distributions of the data of the whole sample for the following tests: The working memory tests related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) as well as the General Reasoning Ability Test (KFT). Figure 21b depicts the corresponding plots for the high-achievers, while the ones for the over-achievers are presented in Figure 21c. Finally, Figure 21d displays the data related to the under-achievers.



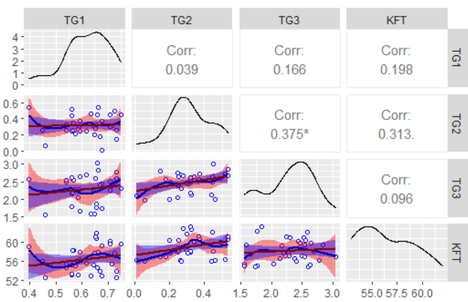
(a) Data of the **whole sample**



(b) Data of the **high-achievers**



(c) Data of the **over-achievers**



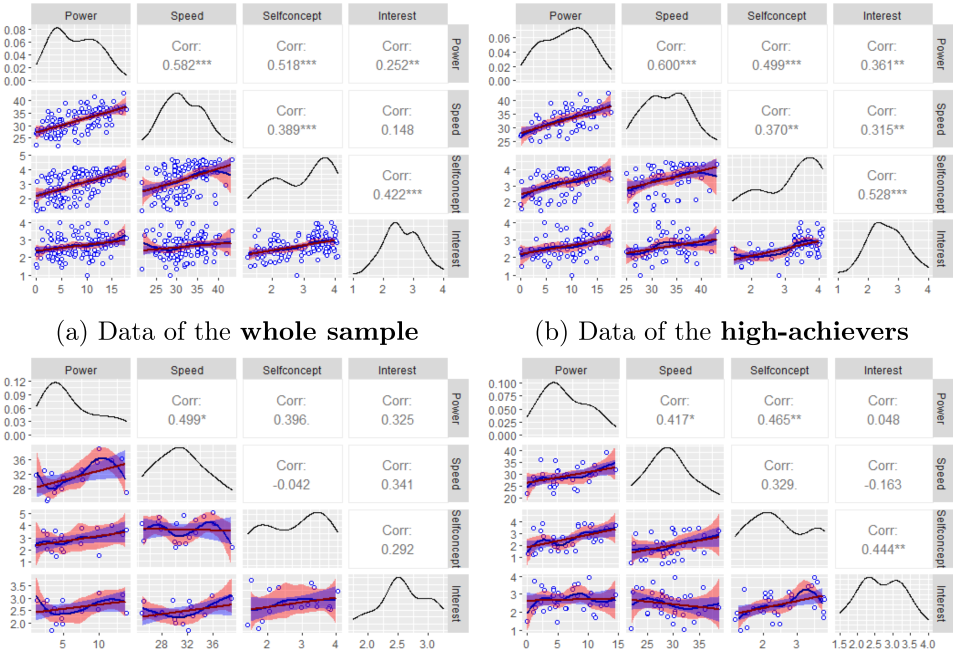
(d) Data of the **under-achievers**

Figure 21: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** as well as the **General Reasoning Ability Test (KFT)**.

With respect to the working memory tests, it can be observed that the positive correlation between the tests related to the functional category of Supervision (TG2) and the ones associated with the functional category of Relational Integration (TG3) is statistically significant for the whole sample as well as for the under-achievers. Moreover, for the whole sample, the positive correlation between the working memory tests related to the functional category of Storage-Processing (TG1) and the General Reasoning Ability Test (KFT) appears to be statistically significant. The positive correlation between the working memory tests related to the functional category of Supervision (TG2) and the General Reasoning Ability Test (KFT) is statistically significant for the whole sample, while it could be relevant for the high-achievers and the under-achievers as well. Finally, for the over-achievers, the positive correlation between the working memory tests related to the functional category of Relational Integration (TG3) and the General Reasoning Ability Test (KFT) appears to be statistically significant. While the distribution of the scores on the General Reasoning Ability Test (KFT) seems to be slightly left-skewed for the over-achievers, it seems to be slightly right-skewed for the under-achievers.

5.1.4 Mathematics Power Test, Mathematics Speed Test, Mathematical Self-Concept and Interest in Mathematics

This subsection aims at presenting the correlations between the data on the Self-Concept Questionnaire Items, on the Interest Questionnaire Items, on the Mathematics Power Test and on the Mathematics Speed Test. Figure 22a displays the scatterplots together with the respective correlations as well as the distributions of the data related to the Mathematics Power Test (Power), the Mathematics Speed Test (Speed), the Self-Concept Questionnaire Items (Selfconcept) and the Interest Questionnaire Items (Interest) for the whole sample. For the high-achievers, the corresponding plots are collected in Figure 22b, while Figure 22c focuses on the over-achievers. Lastly, the data of the under-achievers is visualised in Figure 22d.



(a) Data of the **whole sample** (b) Data of the **high-achievers**

(c) Data of the **over-achievers** (d) Data of the **under-achievers**

Figure 22: Scatterplots together with the respective correlations and the distributions for the different data sets with respect to the **Mathematics Power Test (Power)**, the **Mathematics Speed Test (Speed)**, the **Self-Concept Questionnaire Items (Selfconcept)** and the **Interest Questionnaire Items (Interest)**.

One of the most outstanding results is that the positive correlation between the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed) is statistically significant for all four data sets. In particular, the correlation of 0.582 for the whole sample indicates that the two measures of math performance are related but not identical. The distribution of the scores on the Mathematics Power Test (Power) seems to be slightly right-skewed for both, the over-achievers and the under-achievers. While the positive correlation between the Mathematics Power Test (Power) and the Self-Concept Questionnaire Items (Selfconcept) appears to be statistically significant for all data sets except for the one of the over-achievers, it could be relevant for the over-achievers as well. In addition, the positive correlation between the Mathematics Speed Test (Speed) and the Self-Concept Questionnaire Items (Selfconcept) is statistically significant for the whole sample as well as for the high-achievers, while it could also be relevant for the under-achievers. On the other hand, it can be observed that the positive correlation between the Mathematics Power Test (Power) and the Interest Questionnaire Items (Interest) is statistically significant for the whole sample as well as for the high-achievers. For the high-achievers, the positive correlation between the Mathematics Speed Test (Speed) and the Interest Questionnaire Items (Interest) also appears to be statistically significant. Lastly, the positive correlation between the Self-Concept Questionnaire Items (Selfconcept) and the Interest Questionnaire Items (Interest) is statistically significant for all data sets except for the one of the over-achievers. Regarding the distributions related to the Self-Concept Questionnaire Items (Selfconcept), it can be noted that they are slightly bimodal for all four data sets.

5.2 Group Comparison between High-Achievers and Over-Achievers

The approach for the comparison between high-achievers and over-achievers is based on a comparison between a subgroup of 19 high-achievers (12 female students (63%) and 7 male students (37%)) and the group of 19 over-achievers (10 female students (53%) and 9 male students (47%)). The subgroup of high-achievers has been determined by a matching algorithm: A nearest neighbor matching algorithm has been carried out to match the subgroup of high-achievers to the group of over-achievers in such a way that their performances in the Mathematics Posttest differ as little as possible. The selected high-achievers (marked with a green dot) as well as the over-achievers (marked with a red dot) are shown in Figure 23.

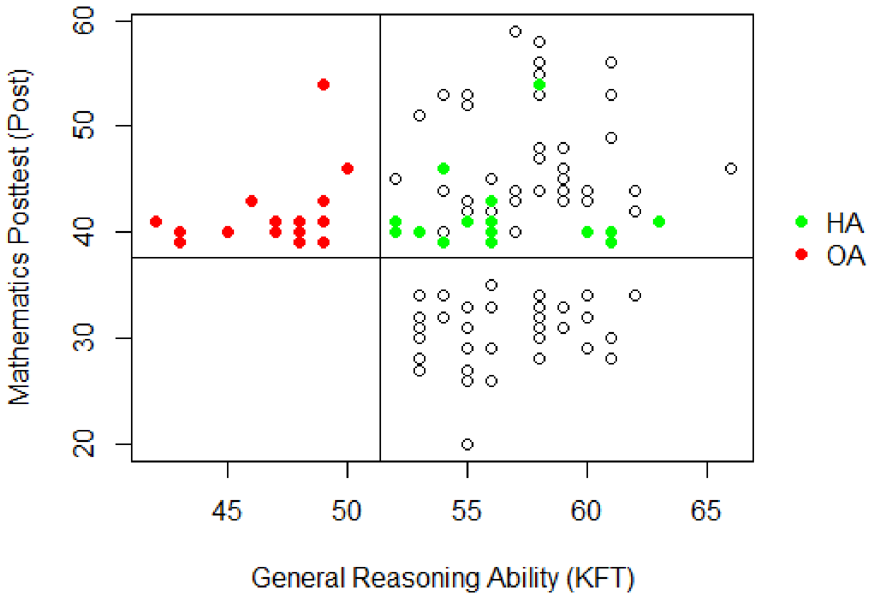


Figure 23: Selected high-achievers (marked with a green dot) and over-achievers (marked with a red dot).

5.2.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

The focus of this subsection lies on a comparison between the selected high-achievers and the over-achievers shown in Figure 23 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores Figures 24, 25 and 26 depict the histograms of the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the various tests. More specifically, Figure 24 displays the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), while Figure 25 shows the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed). Moreover, Figure 26 collects the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3). In addition to the different histograms, Table 1 contains the descriptive statistics of the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the second and third column respectively. The t-values and p-values of a two sample t-test comparing the two groups are indicated in the fourth and fifth column of Table 1. Finally, the last column of Table 1 shows the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals.

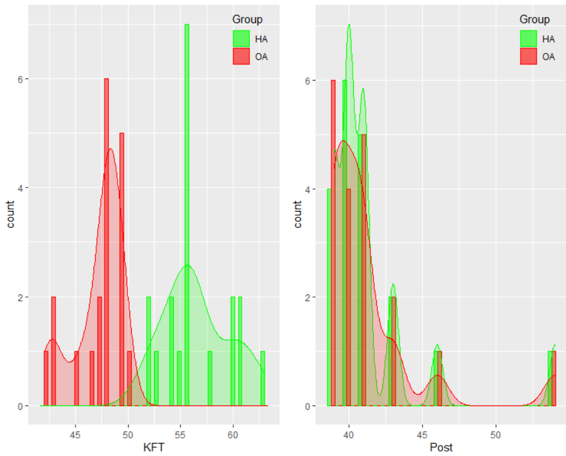


Figure 24: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

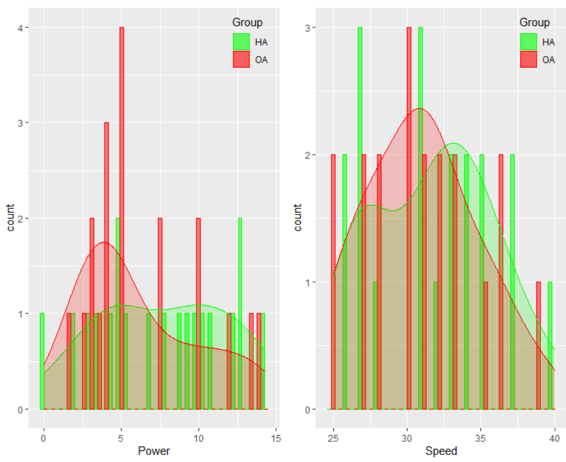


Figure 25: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

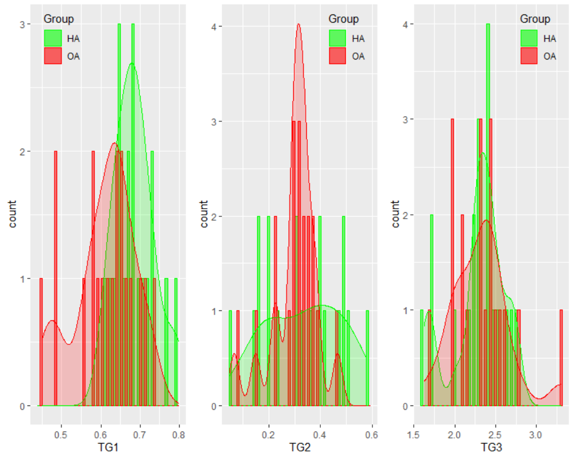


Figure 26: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.58 (3.15)	47.16 (2.32)	10.50	0.000	3.41 (2.38, 4.44)
Mathematics Posttest (Post)	41.42 (3.50)	41.32 (3.56)	0.09	0.927	0.03 (−0.63, 0.69)
Mathematics Power Test (Power)	7.64 (4.16)	6.13 (3.84)	1.17	0.252	0.38 (−0.29, 1.04)
Mathematics Speed Test (Speed)	31.79 (4.13)	30.95 (3.84)	0.65	0.519	0.21 (−0.45, 0.87)
WM Storage-Processing (TG1)	0.69 (0.05)	0.61 (0.08)	3.65	0.001	1.18 (0.47, 1.90)
WM Supervision (TG2)	0.33 (0.15)	0.30 (0.09)	0.77	0.446	0.25 (−0.41, 0.91)
WM Relational Integration (TG3)	2.29 (0.32)	2.32 (0.35)	−0.25	0.803	−0.08 (−0.74, 0.58)

Table 1: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 2), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 2), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 2).

A first important observation is the proper assignment of the two groups: While there are clear group differences in performance on the General Reasoning Ability Test, the two groups do not seem to differ in performance on the Mathematics Posttest.

While substantial group differences cannot be observed with respect to the Mathematics Power Test or the Mathematics Speed Test, the point estimate for Cohen's d related to the Mathematics Power Test is larger than the one associated with the Mathematics Speed Test. In terms of working memory functions, the two groups differ in performance on the working memory test related to the functional category of Storage-Processing. In addition, the point estimate for Cohen's d regarding the working memory function of Supervision indicates a small positive effect, while effect sizes ranging from -0.41, a small negative effect, to 0.91, a large positive effect, would also be supported by the data. The results associated with the functional category of Relational Integration suggest a negligible effect. However, effect sizes ranging from -0.74, a medium negative effect, to 0.58, a medium positive effect, are reasonably compatible with the data as well.

Individual Subtests or Tasks Table 2 shows the subtests or tasks of the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests for a more detailed group comparison between the selected high-achievers and the over-achievers. The non-verbal (N), quantitative (Q) and verbal (V) subscales of the General Reasoning Ability Test (KFT) are denoted by KFT_N, KFT_Q and KFT_V, while Power_T1 - Power_T6 describe the scores in Task 1 - Task 6 in the Mathematics Power Test (Power). In addition, the figural (f), numerical (n) and verbal (v) subtests for each functional category of working memory (TG1 - TG3) are given. In an analogous way as Table 1, Table 2 presents the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals. While KFT and Power in Table 1 refer to the sum scores of the related subtests or tasks displayed in Table 2, TG1 - TG3 in Table 1 describe the mean scores of the associated subtests presented in Table 2.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT_N	22.95 (1.81)	18.16 (2.48)	6.80	0.000	2.21 (1.37, 3.04)
KFT_Q	17.53 (2.17)	15.89 (2.90)	1.96	0.058	0.64 (−0.04, 1.31)
KFT_V	16.11 (2.08)	13.11 (3.28)	3.37	0.002	1.09 (0.39, 1.80)
Power_T1	1.63 (1.09)	1.05 (0.85)	1.83	0.077	0.59 (−0.08, 1.26)
Power_T2	2.00 (1.04)	1.80 (0.96)	0.61	0.547	0.20 (−0.46, 0.86)
Power_T3	1.33 (0.93)	1.04 (0.74)	1.06	0.296	0.34 (−0.32, 1.01)
Power_T4	1.36 (0.82)	1.20 (0.85)	0.58	0.563	0.19 (−0.47, 0.85)
Power_T5	0.49 (0.68)	0.43 (0.76)	0.22	0.824	0.07 (−0.59, 0.73)
Power_T6	0.84 (0.81)	0.61 (0.84)	0.88	0.384	0.29 (−0.38, 0.95)
TG1_f	0.91 (0.08)	0.85 (0.13)	1.48	0.148	0.48 (−0.19, 1.15)
TG1_n	0.52 (0.08)	0.45 (0.09)	2.57	0.015	0.83 (0.15, 1.52)
TG1_v	0.64 (0.07)	0.53 (0.10)	3.66	0.001	1.19 (0.47, 1.90)
TG2_f	0.37 (0.21)	0.32 (0.11)	1.00	0.328	0.32 (−0.34, 0.99)
TG2_n	0.35 (0.20)	0.31 (0.15)	0.78	0.443	0.25 (−0.41, 0.91)
TG2_v	0.27 (0.13)	0.28 (0.13)	−0.18	0.856	−0.06 (−0.72, 0.60)
TG3_f	2.51 (0.40)	2.60 (0.59)	−0.52	0.609	−0.17 (−0.83, 0.49)
TG3_n	2.26 (0.61)	2.18 (0.65)	0.36	0.720	0.12 (−0.54, 0.78)
TG3_v	2.11 (0.48)	2.18 (0.53)	−0.45	0.656	−0.15 (−0.80, 0.51)

Table 2: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

Group differences in performance can be observed for all three subtests of the General Reasoning Ability Test, which is a further indication of the proper assignment of the two groups.

While the two groups may differ in performance on the first task (Power_T1) of the Mathematics Power Test, they do not appear to differ in performance on the fifth task (Power_T5). The point estimates for Cohen's d related to the remaining tasks (Power_T2, Power_T3, Power_T4 and Power_T6) all indicate a small positive effect, with effect sizes ranging from small negative effects to large positive effects being compatible with the data as well. Regarding the working memory subtests, the results suggest group differences in performance on the numerical and verbal subtest of the functional category of Storage-Processing (TG1_n and TG1_v). For the figural subtest (TG1_f), the point estimate for Cohen's d specifies a small positive effect. However, effect sizes ranging from negligible effects to large positive effects would also be in line with the data. The results associated with the figural and numerical subtest of the working memory function Supervision (TG2_f and TG2_n) show a small positive effect, while effect sizes ranging from small negative effects to large positive effects would be supported by the data as well. Moreover, the two groups do not seem to differ in performance on the verbal subtest (TG2_v). Finally, the point estimates for Cohen's d related to the subtests of the working memory function Relational Integration (TG3_f, TG3_n and TG3_v) all suggest a negligible effect. Nonetheless, effect sizes ranging from large or medium negative effects to small or medium positive effects would also be compatible with the data.

5.2.2 Mathematical Self-Concept and Interest in Mathematics

To further compare the selected high-achievers and the over-achievers shown in Figure 23, a comparison of the two groups with respect to their mathematical self-concept as well as their interest in mathematics is given in this subsection.

Sum Scores and Mean Scores The mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the over-achievers are visualised in Figure 27. Additionally, Table 3 shows the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals. Moreover, to facilitate an overview on the different results, Table 3 also lists the results related to the mathematics tests and the General Reasoning Ability Test already given in Table 1. The structures of Table 3 and Table 1 are analogous.

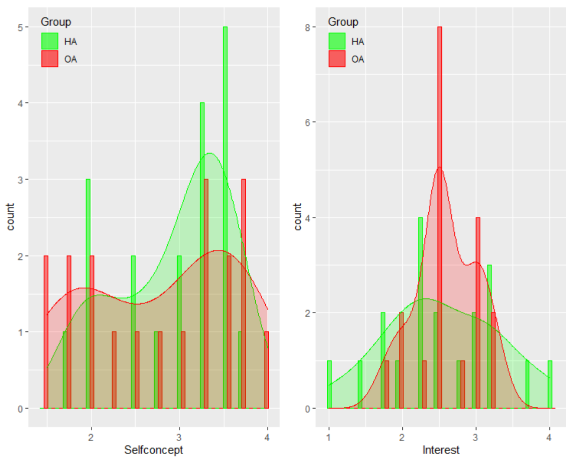


Figure 27: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the over-achievers (OA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.58 (3.15)	47.16 (2.32)	10.50	0.000	3.41 (2.38, 4.44)
Mathematics Posttest (Post)	41.42 (3.50)	41.32 (3.56)	0.09	0.927	0.03 (−0.63, 0.69)
Mathematics Power Test (Power)	7.64 (4.16)	6.13 (3.84)	1.17	0.252	0.38 (−0.29, 1.04)
Mathematics Speed Test (Speed)	31.79 (4.13)	30.95 (3.84)	0.65	0.519	0.21 (−0.45, 0.87)
Self-Concept Questionnaire Items (Selfconcept)	2.93 (0.63)	2.79 (0.85)	0.60	0.555	0.19 (−0.47, 0.85)
Interest Questionnaire Items (Interest)	2.54 (0.78)	2.59 (0.42)	−0.26	0.797	−0.08 (−0.74, 0.57)

Table 3: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 2), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 2), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 4) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 4) respectively.

The point estimates for Cohen's d associated with the mean scores of the Self-Concept Questionnaire Items as well as with the mean scores of the Interest Questionnaire Items both indicate a negligible effect. Nonetheless, effect sizes ranging from −0.47, a small negative effect, to 0.85, a large positive effect, would also be in line with the Self-Concept data, while effect sizes ranging from −0.74, a medium negative effect, to 0.57, a medium positive effect, would be compatible with the Interest data as well.

Individual Questionnaire Items Table 4 lists the individual Questionnaire Items to enable a more detailed group comparison between the selected high-achievers and the over-achievers with respect to the Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the Interest Questionnaire Items (Interest_QI5 - Interest_QI8). The structure of Table 4 regarding the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals is analogous to the one of Table 3. It is important to mention that Selfconcept and Interest in Table 3 refer to the mean scores of Selfconcept_QI1 - Selfconcept_QI4 and of Interest_QI5 - Interest_QI8 respectively.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
Selfconcept_QI1	2.84 (0.96)	2.53 (0.90)	1.04	0.303	0.34 (−0.32, 1.00)
Selfconcept_QI2	3.05 (0.62)	2.95 (1.03)	0.38	0.705	0.12 (−0.53, 0.78)
Selfconcept_QI3	2.53 (0.77)	2.42 (1.02)	0.36	0.722	0.12 (−0.54, 0.78)
Selfconcept_QI4	3.32 (0.75)	3.26 (0.93)	0.19	0.849	0.06 (−0.60, 0.72)
Interest_QI5	2.47 (1.02)	2.32 (0.89)	0.51	0.614	0.17 (−0.49, 0.82)
Interest_QI6	2.21 (0.98)	2.21 (0.63)	0.00	1.000	0.00 (−0.66, 0.66)
Interest_QI7	2.53 (0.96)	2.37 (0.68)	0.58	0.565	0.19 (−0.47, 0.85)
Interest_QI8	2.95 (0.85)	3.47 (0.61)	−2.19	0.035	−0.71 (−1.39, −0.03)

Table 4: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

With respect to the Self-Concept Questionnaire Items, the point estimate for Cohen's d related to the first item (Selfconcept_QI1) shows a small positive effect. Nonetheless, effect sizes ranging from −0.32, a small negative effect, to 1.00, a large positive effect, would also be compatible with the data. For the second and third item (Selfconcept_QI2 and Selfconcept_QI3), the results both suggest a negligible effect, while effect sizes ranging from medium negative effects to medium positive effects would be supported by the data as well. In addition, the two groups do not seem to differ on the fourth item (Selfconcept_QI4). Regarding the Interest Questionnaire Items, the point estimates for Cohen's d associated with the fifth and seventh item (Interest_QI5 and Interest_QI7) both indicate a negligible effect. However, effect sizes ranging from small negative effects to large positive effects would also be in line with the data. Finally, there do not appear to be any group differences with respect to the sixth item (Interest_QI6), while group differences can be observed on the eighth item (Interest_QI8).

5.2.3 Summary of the Results

In Subsections 5.2.1 and 5.2.2, several group comparisons between high-achieving and over-achieving students were carried out. The results presented in those subsections are listed in Table 5 in descending order of the respective effect sizes in terms of Cohen's d. The goal of this subsection is to summarise them.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT	56.58 (3.15)	47.16 (2.32)	10.50	0.000	3.41 (2.38, 4.44)
KFT_N	22.95 (1.81)	18.16 (2.48)	6.80	0.000	2.21 (1.37, 3.04)
TG1_v	0.64 (0.07)	0.53 (0.10)	3.66	0.001	1.19 (0.47, 1.90)
TG1	0.69 (0.05)	0.61 (0.08)	3.65	0.001	1.18 (0.47, 1.90)
KFT_V	16.11 (2.08)	13.11 (3.28)	3.37	0.002	1.09 (0.39, 1.80)
TG1_n	0.52 (0.08)	0.45 (0.09)	2.57	0.015	0.83 (0.15, 1.52)
KFT_Q	17.53 (2.17)	15.89 (2.90)	1.96	0.058	0.64 (−0.04, 1.31)
Power_T1	1.63 (1.09)	1.05 (0.85)	1.83	0.077	0.59 (−0.08, 1.26)
TG1_f	0.91 (0.08)	0.85 (0.13)	1.48	0.148	0.48 (−0.19, 1.15)
Power	7.64 (4.16)	6.13 (3.84)	1.17	0.252	0.38 (−0.29, 1.04)
Power_T3	1.33 (0.93)	1.04 (0.74)	1.06	0.296	0.34 (−0.32, 1.01)
Selfconcept_QI1	2.84 (0.96)	2.53 (0.90)	1.04	0.303	0.34 (−0.32, 1.00)
TG2_f	0.37 (0.21)	0.32 (0.11)	1.00	0.328	0.32 (−0.34, 0.99)
Power_T6	0.84 (0.81)	0.61 (0.84)	0.88	0.384	0.29 (−0.38, 0.95)
TG2_n	0.35 (0.20)	0.31 (0.15)	0.78	0.443	0.25 (−0.41, 0.91)
TG2	0.33 (0.15)	0.30 (0.09)	0.77	0.446	0.25 (−0.41, 0.91)
Speed	31.79 (4.13)	30.95 (3.84)	0.65	0.519	0.21 (−0.45, 0.87)
Power_T2	2.00 (1.04)	1.80 (0.96)	0.61	0.547	0.20 (−0.46, 0.86)
Selfconcept	2.93 (0.63)	2.79 (0.85)	0.60	0.555	0.19 (−0.47, 0.85)
Power_T4	1.36 (0.82)	1.20 (0.85)	0.58	0.563	0.19 (−0.47, 0.85)
Interest_QI7	2.53 (0.96)	2.37 (0.68)	0.58	0.565	0.19 (−0.47, 0.85)
Interest_QI5	2.47 (1.02)	2.32 (0.89)	0.51	0.614	0.17 (−0.49, 0.82)
Selfconcept_QI2	3.05 (0.62)	2.95 (1.03)	0.38	0.705	0.12 (−0.53, 0.78)
TG3_n	2.26 (0.61)	2.18 (0.65)	0.36	0.720	0.12 (−0.54, 0.78)
Selfconcept_QI3	2.53 (0.77)	2.42 (1.02)	0.36	0.722	0.12 (−0.54, 0.78)
Power_T5	0.49 (0.68)	0.43 (0.76)	0.22	0.824	0.07 (−0.59, 0.73)
Selfconcept_QI4	3.32 (0.75)	3.26 (0.93)	0.19	0.849	0.06 (−0.60, 0.72)
Post	41.42 (3.50)	41.32 (3.56)	0.09	0.927	0.03 (−0.63, 0.69)
Interest_QI6	2.21 (0.98)	2.21 (0.63)	0.00	1.000	0.00 (−0.66, 0.66)
TG2_v	0.27 (0.13)	0.28 (0.13)	−0.18	0.856	−0.06 (−0.72, 0.60)
TG3	2.29 (0.32)	2.32 (0.35)	−0.25	0.803	−0.08 (−0.74, 0.58)
Interest	2.54 (0.78)	2.59 (0.42)	−0.26	0.797	−0.08 (−0.74, 0.57)
TG3_v	2.11 (0.48)	2.18 (0.53)	−0.45	0.656	−0.15 (−0.80, 0.51)
TG3_f	2.51 (0.40)	2.60 (0.59)	−0.52	0.609	−0.17 (−0.83, 0.49)
Interest_QI8	2.95 (0.85)	3.47 (0.61)	−2.19	0.035	−0.71 (−1.39, −0.03)

Table 5: Summary of all group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

The substantial group differences in performance on the General Reasoning Ability Test (KFT) and on its subscales (KFT_N, KFT_V, KFT_Q) respectively are an indicator of the proper assignment of the two groups. Apart from these group differences in performance, the largest effect sizes in terms of Cohen's d appear for the working memory tests associated with the functional category of Storage-Processing (TG1 and TG1_v, TG1_n respectively). Regarding the questionnaire items, the absolute values of the point estimates for Cohen's d show the largest effect on the eighth item (Interest_QI8), while the ones associated with the Mathematics Power Test specify the largest effect on the first task (Power_T1). Lastly, it can be observed that the point estimate for Cohen's d related to the Mathematics Power Test (Power) is larger than the one associated with the Mathematics Speed Test (Speed), as already mentioned in Subsection 5.2.1.

5.2.4 Concluding Remark

The following note should be added to conclude Section 5.2. In order to provide further evidence for the group differences presented in this section, two additional group comparisons between other subgroups of high-achieving students and over-achieving students were carried out. These group comparisons are collected in Sections A.1 and A.2 in the appendix. It is important to note that all three group comparisons led to similar main findings. More concretely, the main results discussed in this section also emerge for other approaches of selecting a subgroup of high-achievers for the comparison against the group of over-achievers.

5.3 Group Comparison between High-Achievers and Under-Achievers

The following approach for the comparison between high-achievers and under-achievers compares a subgroup of 34 high-achievers (23 female students (68%) and 11 male students (32%)) to the group of 34 under-achievers (17 female students (50%) and 17 male students (50%)). The subgroup of high-achievers has been determined by a matching algorithm: An optimal matching algorithm has been performed in order to match the subgroup of high-achievers to the group of under-achievers in such a way that their performances in the General Reasoning Ability Test differ as little as possible. The selected high-achievers (marked with a green dot) as well as the under-achievers (marked with an orange dot) are depicted in Figure 28.

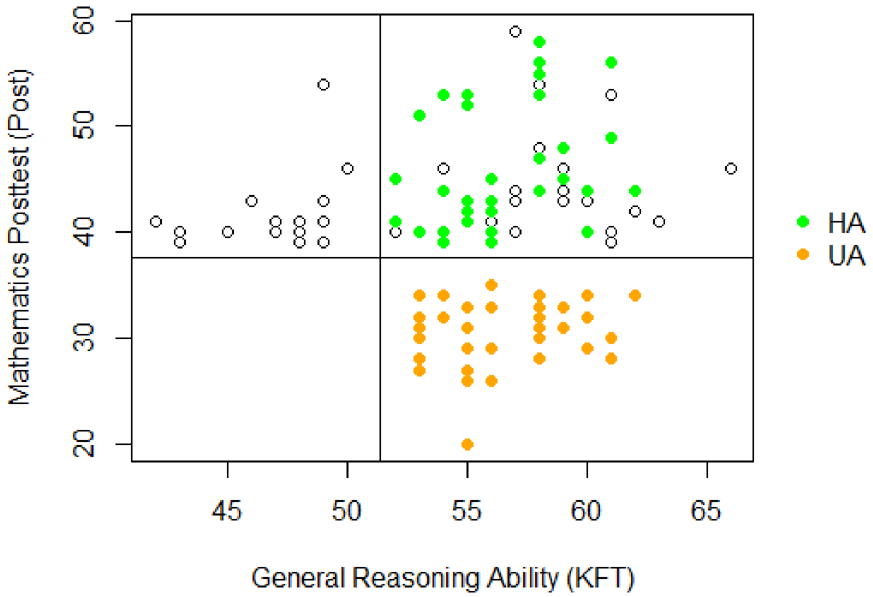


Figure 28: Selected high-achievers (marked with a green dot) and under-achievers (marked with an orange dot).

5.3.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

This subsection discusses a comparison between the selected high-achievers and the under-achievers shown in Figure 28 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms of the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the various tests are shown in Figures 29,30 and 31. While Figure 29 presents the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), Figure 30 depicts the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed). Lastly, Figure 31 arranges the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3). To complement the different histograms, Table 6 contains the descriptive statistics of the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the second and third column respectively. The fourth and fifth column of Table 6 indicate the t-values and p-values of a two sample t-test comparing the two groups. In addition, the last column of Table 6 lists the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals.

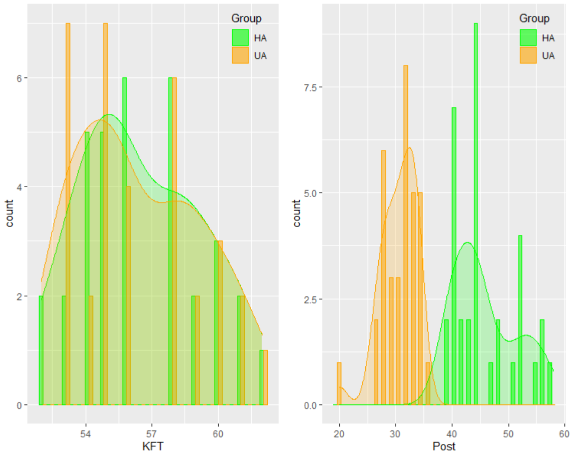


Figure 29: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

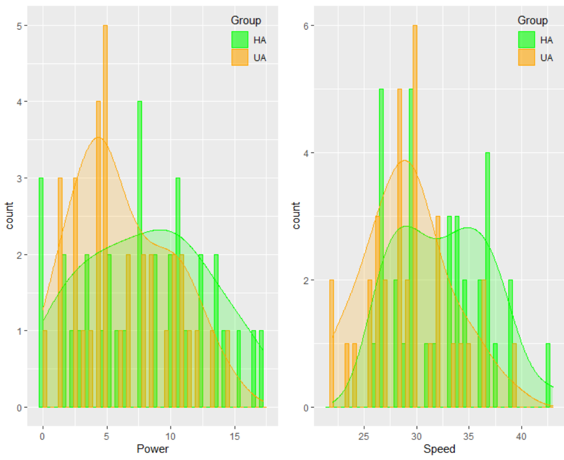


Figure 30: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

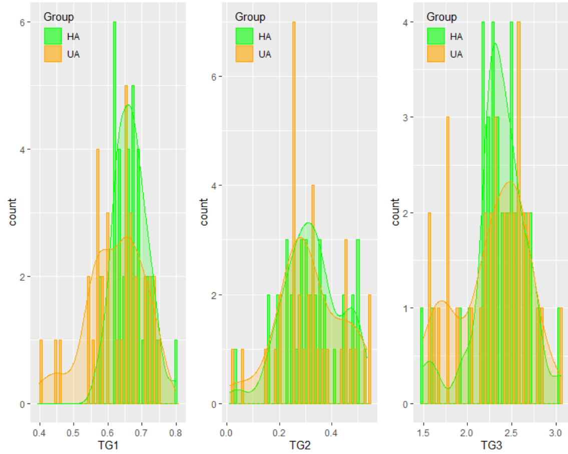


Figure 31: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.50 (2.71)	56.41 (2.74)	0.13	0.894	0.03 (−0.45, 0.52)
Mathematics Posttest (Post)	45.91 (5.64)	30.50 (3.16)	13.90	0.000	3.37 (2.62, 4.12)
Mathematics Power Test (Power)	8.03 (4.91)	6.17 (3.80)	1.75	0.085	0.42 (−0.07, 0.91)
Mathematics Speed Test (Speed)	32.62 (4.37)	29.15 (4.05)	3.40	0.001	0.82 (0.32, 1.33)
WM Storage-Processing (TG1)	0.67 (0.05)	0.62 (0.08)	2.90	0.005	0.70 (0.21, 1.20)
WM Supervision (TG2)	0.32 (0.11)	0.31 (0.13)	0.33	0.742	0.08 (−0.40, 0.56)
WM Relational Integration (TG3)	2.37 (0.31)	2.28 (0.40)	1.04	0.303	0.25 (−0.23, 0.74)

Table 6: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 7), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 7), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 7).

The appropriate assignment of the two groups could be confirmed: While substantial group differences in performance cannot be detected with respect to the General Reasoning Ability Test, the two groups clearly differ in performance on the Mathematics Posttest.

The point estimate for Cohen's d associated with the Mathematics Power Test shows a small positive effect, while effect sizes ranging from -0.07, a negligible effect, to 0.91, a large positive effect, would be compatible with the data as well. In addition, the results of a one-sided two sample t-test with respect to the Mathematics Power Test should also be considered, as the initial hypothesis was that high-achievers would outperform under-achievers in the Mathematics Power Test. The one-sided t-test yields a p-value of $p = 0.043$ for the t-value of $t = 1.75$, suggesting group differences in performance in favour of the high-achieving students. Furthermore, the results related to the Mathematics Speed Test indicate group differences in performance. Regarding working memory functions, it can be observed that the two groups differ in performance on the working memory test measuring the functional category of Storage-Processing. In contrast, the point estimate for Cohen's d associated with the functional category of Supervision specifies a negligible effect, while effect sizes ranging from -0.40, a small negative effect, to 0.56, a medium positive effect, would be supported by the data as well. Lastly, the results related to the working memory function of Relational Integration show a small positive effect. Nonetheless, effect sizes ranging from -0.23, a small negative effect, to 0.74, a medium positive effect, would also be in line with the data.

Individual Subtests or Tasks Table 7 arranges the different subtests or tasks of the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests for a more detailed group comparison between the selected high-achievers and the under-achievers. More concretely, Table 7 presents the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subscales of the General Reasoning Ability Test as well as the scores in Task 1 - Task 6 in the Mathematics Power Test (Power_T1 - Power_T6). Additionally, it collects the figural (f), numerical (n) and verbal (v) subtests for each functional category of working memory (TG1 - TG3). The structure of Table 7 regarding the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals is analogous to the one of Table 6. While KFT and Power in Table 6 relate to the sum scores of the corresponding subtests or tasks given in Table 7, TG1 - TG3 in Table 6 describe the mean scores of the associated subtests displayed in Table 7.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
KFT_N	22.79 (1.82)	22.21 (2.11)	1.23	0.224	0.30 (−0.19, 0.78)
KFT_Q	18.03 (2.18)	17.85 (2.08)	0.34	0.734	0.08 (−0.40, 0.57)
KFT_V	15.68 (2.03)	16.35 (1.76)	−1.47	0.146	−0.36 (−0.84, 0.13)
Power_T1	1.62 (1.09)	1.37 (0.89)	1.04	0.304	0.25 (−0.23, 0.74)
Power_T2	1.80 (1.02)	1.71 (1.00)	0.36	0.720	0.09 (−0.40, 0.57)
Power_T3	1.24 (0.74)	0.82 (0.64)	2.53	0.014	0.61 (0.12, 1.11)
Power_T4	1.46 (0.95)	1.21 (0.99)	1.10	0.277	0.27 (−0.22, 0.75)
Power_T5	0.84 (1.07)	0.49 (0.82)	1.49	0.140	0.36 (−0.13, 0.85)
Power_T6	1.07 (1.06)	0.57 (0.76)	2.20	0.032	0.53 (0.04, 1.03)
TG1_f	0.90 (0.08)	0.88 (0.12)	0.91	0.367	0.22 (−0.27, 0.71)
TG1_n	0.51 (0.07)	0.42 (0.11)	3.61	0.001	0.87 (0.37, 1.38)
TG1_v	0.59 (0.08)	0.55 (0.09)	1.89	0.064	0.46 (−0.03, 0.95)
TG2_f	0.37 (0.16)	0.33 (0.21)	0.93	0.358	0.22 (−0.26, 0.71)
TG2_n	0.31 (0.19)	0.36 (0.19)	−1.14	0.257	−0.28 (−0.76, 0.21)
TG2_v	0.29 (0.12)	0.25 (0.11)	1.42	0.160	0.34 (−0.14, 0.83)
TG3_f	2.44 (0.44)	2.41 (0.43)	0.30	0.768	0.07 (−0.41, 0.56)
TG3_n	2.36 (0.60)	2.22 (0.65)	0.93	0.355	0.23 (−0.26, 0.71)
TG3_v	2.31 (0.39)	2.22 (0.68)	0.70	0.487	0.17 (−0.32, 0.66)

Table 7: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

The point estimates for Cohen's d related to the non-verbal (KFT_N) and verbal (KFT_V) subtest of the General Reasoning Ability Test suggest a small positive effect and a small negative effect respectively, while there do not seem to be any group differences in performance on the quantitative subtest (KFT_Q).

Substantial group differences in performance can be observed for the third and sixth task of the Mathematics Power Test (Power_T3 and Power_T6), while the two groups do not appear to differ in performance on the second task (Power_T2). For the remaining tasks (Power_T1, Power_T4 and Power_T5), the point estimates for Cohen's d all indicate a small positive effect, while effect sizes ranging from small negative or negligible effects to medium or large positive effects would also be in line with the data. With respect to the subtests of the working memory function Storage-Processing, the two groups clearly differ in performance on the numerical subtest (TG1_n), while there seem to be group differences in performance on the verbal subtest (TG1_v) as well. The results associated with the figural subtest (TG1_f) specify a small positive effect. Nonetheless, effect sizes ranging from -0.27, a small negative effect, to 0.71, a medium positive effect, would be compatible with the data as well. For the figural and verbal subtest of the functional category of Supervision (TG2_f and TG2_v), the point estimates for Cohen's d both show a small positive effect, with effect sizes ranging from small negative or negligible effects to medium or large positive effects being also supported by the data. On the other hand, the results related to the numerical subtest (TG2_n) suggest a small negative effect. However, effect sizes ranging from -0.76, a medium negative effect, to 0.21, a small positive effect, would be in line with the data as well. Finally, substantial group differences in performance cannot be observed for the figural subtest of the working memory function Relational Integration (TG3_f). Moreover, the point estimates for Cohen's d indicate a small positive effect with respect to the numerical subtest (TG3_n) and a negligible effect regarding the verbal subtest (TG3_v). Nonetheless, effect sizes ranging from small negative effects to medium positive effects would also be supported by the data in both cases.

5.3.2 Mathematical Self-Concept and Interest in Mathematics

As a further comparison between the selected high-achievers and the under-achievers shown in Figure 28, the two groups have been compared with respect to their mathematical self-concept and their interest in mathematics as well. The corresponding results are discussed in this subsection.

Sum Scores and Mean Scores Figure 32 depicts the histograms related to the mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the under-achievers respectively. To complement this graphical representation, the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are presented in Table 8. In order to facilitate the comparison between the different results, Table 8 additionally lists the results associated with the mathematics tests and the General Reasoning Ability Test already given in Table 6. The structures of Table 8 and Table 6 are analogous.

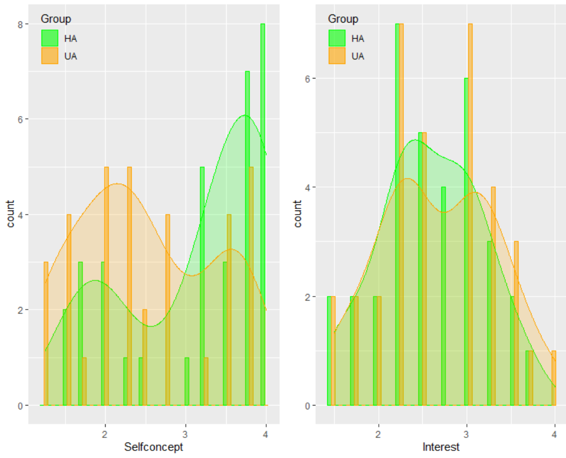


Figure 32: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the under-achievers (UA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.50 (2.71)	56.41 (2.74)	0.13	0.894	0.03 (−0.45, 0.52)
Mathematics Posttest (Post)	45.91 (5.64)	30.50 (3.16)	13.90	0.000	3.37 (2.62, 4.12)
Mathematics Power Test (Power)	8.03 (4.91)	6.17 (3.80)	1.75	0.085	0.42 (−0.07, 0.91)
Mathematics Speed Test (Speed)	32.62 (4.37)	29.15 (4.05)	3.40	0.001	0.82 (0.32, 1.33)
Self-Concept Questionnaire Items (Selfconcept)	3.15 (0.87)	2.49 (0.84)	3.14	0.003	0.76 (0.26, 1.26)
Interest Questionnaire Items (Interest)	2.60 (0.57)	2.68 (0.65)	−0.55	0.587	−0.13 (−0.62, 0.35)

Table 8: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 7), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 7), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 9) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 9) respectively.

The results suggest that there are group differences with respect to the mean scores of the Self-Concept Questionnaire Items: The mathematical self-concept of the under-achievers appears to be noticeably lower than the one of the selected high-achievers. In contrast, the point estimate for Cohen's d related to the mean scores of the Interest Questionnaire Items specifies a negligible effect, while effect sizes ranging from -0.62, a medium negative effect, to 0.35, a small positive effect, would be supported by the data as well.

Individual Questionnaire Items For a more detailed evaluation of the group comparison between the selected high-achievers and the under-achievers with respect to the Self-Concept Questionnaire Items and the Interest Questionnaire Items, the individual Questionnaire Items are collected in Table 9. Selfconcept_QI1 - Selfconcept_QI4 represent the Self-Concept Questionnaire Items, while Interest_QI5 - Interest_QI8 correspond to the Interest Questionnaire Items. The related descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are arranged in an analogous manner as in Table 8. While Selfconcept in Table 8 refers to the mean scores of Selfconcept_QI1 - Selfconcept_QI4, Interest in Table 8 describes the mean scores of Interest_QI5 - Interest_QI8.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Selfconcept_QI1	3.00 (0.98)	2.65 (1.01)	1.46	0.150	0.35 (−0.13, 0.84)
Selfconcept_QI2	3.21 (0.84)	2.44 (0.82)	3.78	0.000	0.92 (0.41, 1.43)
Selfconcept_QI3	3.00 (1.13)	1.97 (1.00)	3.98	0.000	0.97 (0.45, 1.48)
Selfconcept_QI4	3.38 (0.85)	2.91 (0.93)	2.17	0.034	0.53 (0.03, 1.02)
Interest_QI5	2.53 (0.86)	2.32 (1.12)	0.85	0.399	0.21 (−0.28, 0.69)
Interest_QI6	2.21 (0.81)	2.44 (0.79)	−1.22	0.228	−0.30 (−0.78, 0.19)
Interest_QI7	2.53 (0.79)	2.62 (0.82)	−0.45	0.652	−0.11 (−0.59, 0.37)
Interest_QI8	3.12 (0.64)	3.32 (0.64)	−1.33	0.189	−0.32 (−0.81, 0.17)

Table 9: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

The point estimate for Cohen's d related to the first item (Selfconcept_QI1) shows a small positive effect, while effect sizes ranging from negligible effects to large positive effects would be compatible with the data as well. In addition, there seem to be group differences with respect to the remaining Self-Concept Questionnaire Items (Selfconcept_QI2 - Selfconcept_QI4). In terms of the Interest Questionnaire Items, the results associated with the fifth item (Interest_QI5) specify a small positive effect. Nonetheless, effect sizes ranging from -0.28, a small negative effect, to 0.69, a medium positive effect, would also be in line with the data. On the other hand, the point estimates for Cohen's d related to the sixth and eighth item (Interest_QI6 and Interest_QI8) both suggest a small negative effect, while effect sizes ranging from medium negative effects to negligible effects would be supported by the data as well. Lastly, the results associated with the seventh item (Interest_QI7) show a negligible effect. However, effect sizes ranging from medium negative effects to small positive effects would also be compatible with the data.

5.3.3 Summary of the Results

The aim of this subsection is to summarise the results from Subsections 5.3.1 and 5.3.2, which focused on different group comparisons between high-achievers and under-achievers. These results are collected in Table 10 in descending order of the respective effect sizes in terms of Cohen's d.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Post	45.91 (5.64)	30.50 (3.16)	13.90	0.000	3.37 (2.62, 4.12)
Selfconcept_QI3	3.00 (1.13)	1.97 (1.00)	3.98	0.000	0.97 (0.45, 1.48)
Selfconcept_QI2	3.21 (0.84)	2.44 (0.82)	3.78	0.000	0.92 (0.41, 1.43)
TG1_n	0.51 (0.07)	0.42 (0.11)	3.61	0.001	0.87 (0.37, 1.38)
Speed	32.62 (4.37)	29.15 (4.05)	3.40	0.001	0.82 (0.32, 1.33)
Selfconcept	3.15 (0.87)	2.49 (0.84)	3.14	0.003	0.76 (0.26, 1.26)
TG1	0.67 (0.05)	0.62 (0.08)	2.90	0.005	0.70 (0.21, 1.20)
Power_T3	1.24 (0.74)	0.82 (0.64)	2.53	0.014	0.61 (0.12, 1.11)
Power_T6	1.07 (1.06)	0.57 (0.76)	2.20	0.032	0.53 (0.04, 1.03)
Selfconcept_QI4	3.38 (0.85)	2.91 (0.93)	2.17	0.034	0.53 (0.03, 1.02)
TG1_v	0.59 (0.08)	0.55 (0.09)	1.89	0.064	0.46 (−0.03, 0.95)
Power	8.03 (4.91)	6.17 (3.80)	1.75	0.085	0.42 (−0.07, 0.91)
Power_T5	0.84 (1.07)	0.49 (0.82)	1.49	0.140	0.36 (−0.13, 0.85)
Selfconcept_QI1	3.00 (0.98)	2.65 (1.01)	1.46	0.150	0.35 (−0.13, 0.84)
TG2_v	0.29 (0.12)	0.25 (0.11)	1.42	0.160	0.34 (−0.14, 0.83)
KFT_N	22.79 (1.82)	22.21 (2.11)	1.23	0.224	0.30 (−0.19, 0.78)
Power_T4	1.46 (0.95)	1.21 (0.99)	1.10	0.277	0.27 (−0.22, 0.75)
TG3	2.37 (0.31)	2.28 (0.40)	1.04	0.303	0.25 (−0.23, 0.74)
Power_T1	1.62 (1.09)	1.37 (0.89)	1.04	0.304	0.25 (−0.23, 0.74)
TG3_n	2.36 (0.60)	2.22 (0.65)	0.93	0.355	0.23 (−0.26, 0.71)
TG2_f	0.37 (0.16)	0.33 (0.21)	0.93	0.358	0.22 (−0.26, 0.71)
TG1_f	0.90 (0.08)	0.88 (0.12)	0.91	0.367	0.22 (−0.27, 0.71)
Interest_QI5	2.53 (0.86)	2.32 (1.12)	0.85	0.399	0.21 (−0.28, 0.69)
TG3_v	2.31 (0.39)	2.22 (0.68)	0.70	0.487	0.17 (−0.32, 0.66)
Power_T2	1.80 (1.02)	1.71 (1.00)	0.36	0.720	0.09 (−0.40, 0.57)
KFT_Q	18.03 (2.18)	17.85 (2.08)	0.34	0.734	0.08 (−0.40, 0.57)
TG2	0.32 (0.11)	0.31 (0.13)	0.33	0.742	0.08 (−0.40, 0.56)
TG3_f	2.44 (0.44)	2.41 (0.43)	0.30	0.768	0.07 (−0.41, 0.56)
KFT	56.50 (2.71)	56.41 (2.74)	0.13	0.894	0.03 (−0.45, 0.52)
Interest_QI7	2.53 (0.79)	2.62 (0.82)	−0.45	0.652	−0.11 (−0.59, 0.37)
Interest	2.60 (0.57)	2.68 (0.65)	−0.55	0.587	−0.13 (−0.62, 0.35)
TG2_n	0.31 (0.19)	0.36 (0.19)	−1.14	0.257	−0.28 (−0.76, 0.21)
Interest_QI6	2.21 (0.81)	2.44 (0.79)	−1.22	0.228	−0.30 (−0.78, 0.19)
Interest_QI8	3.12 (0.64)	3.32 (0.64)	−1.33	0.189	−0.32 (−0.81, 0.17)
KFT_V	15.68 (2.03)	16.35 (1.76)	−1.47	0.146	−0.36 (−0.84, 0.13)

Table 10: Summary of all group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Clear group differences in performance appear on the Mathematics Posttest (Post), which is in line with the proper assignment of the two groups. With respect to the questionnaire items, the largest effects in terms of Cohen's d are found for some of the Self-Concept Questionnaire Items (Selfconcept and Selfconcept_QI2 - Selfconcept_QI4 respectively). For the working memory tests, the point estimates for Cohen's d specify the largest effects on the tests associated with the functional category of Storage-Processing (TG1 and TG1_n, TG1_v respectively). In addition, it can be observed that the point estimate for Cohen's d related to the Mathematics Speed Test (Speed) is larger than the one referring to the Mathematics Power Test (Power). Within the Mathematics Power Test, the point estimates for Cohen's d show the largest effects on the third and sixth task (Power_T3 and Power_T6).

5.3.4 Concluding Remark

As a concluding remark on this section, it should be mentioned that in order to further analyse the group differences discussed in the previous subsections, two additional group comparisons between different subgroups of high-achievers and under-achievers were performed. The corresponding results are given in Sections A.3 and A.4 in the appendix. They illustrate that the main findings presented in this section can also be derived from other approaches of selecting subgroups of high-achieving students and under-achieving students for the comparison of the two groups.

5.4 Subgroup Comparison within High-Achievers

The following approach for a comparison within the group of high-achievers compares the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (16 students: 10 female students (63%), 5 male students (31%) and one student, who did not specify the gender (6%)) against the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (17 students: 12 female students (71%) and 5 male students (29%)). The two subgroups are depicted in Figure 33, where the top scoring high-achievers are marked with a green dot, while the lower scoring high-achievers are marked with a blue dot.

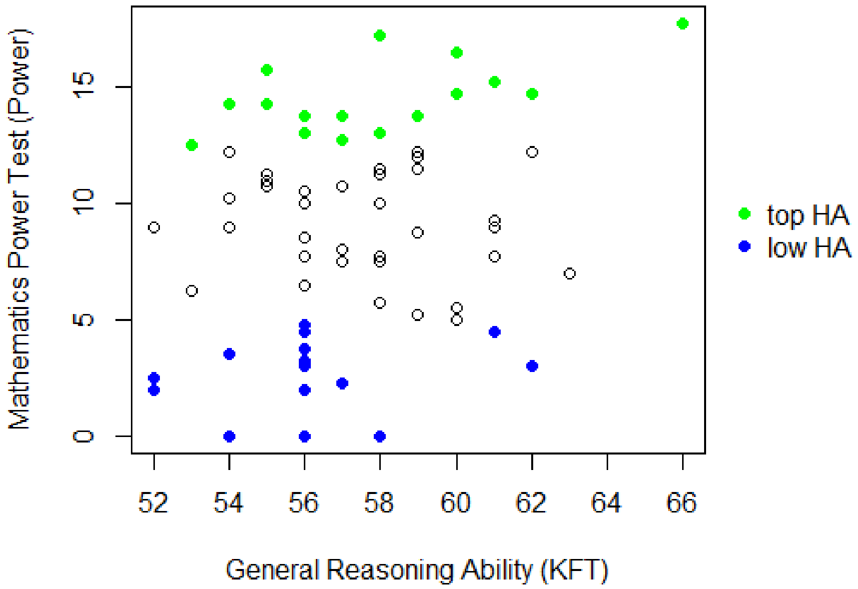


Figure 33: Top scoring high-achievers (marked with a green dot) and lower scoring high-achievers (marked with a blue dot).

5.4.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

The goal of this subsection is to present a comparison between the two subgroups shown in Figure 33 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms of the performances of the two subgroups in the various tests are shown in Figures 34, 35 and 36. While Figure 34 presents the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), Figure 35 depicts the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed). Lastly, the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) are arranged in Figure 36. Additionally to the different histograms, Table 11 lists the descriptive statistics of the performances of the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and of the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) in the second and third column respectively. The fourth and fifth column of Table 11 display the t-values and p-values of a two sample t-test comparing the two subgroups. Finally, the last column of Table 11 presents the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals.

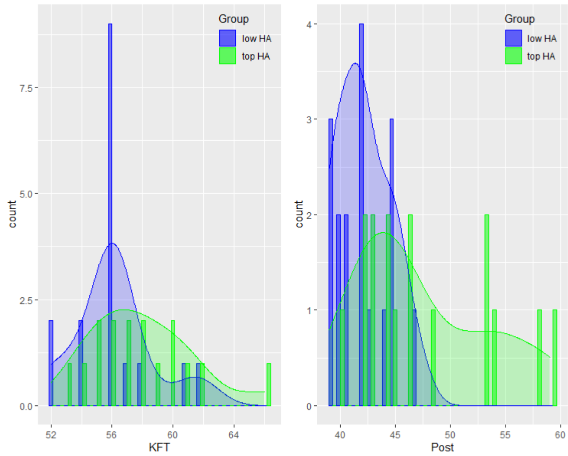


Figure 34: Histograms visualising the performances of the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and of the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

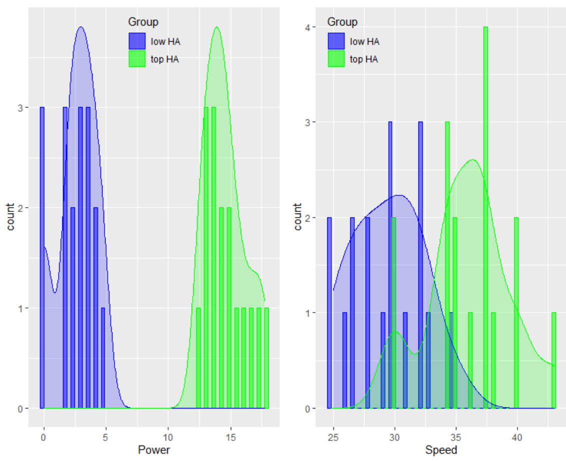


Figure 35: Histograms visualising the performances of the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and of the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

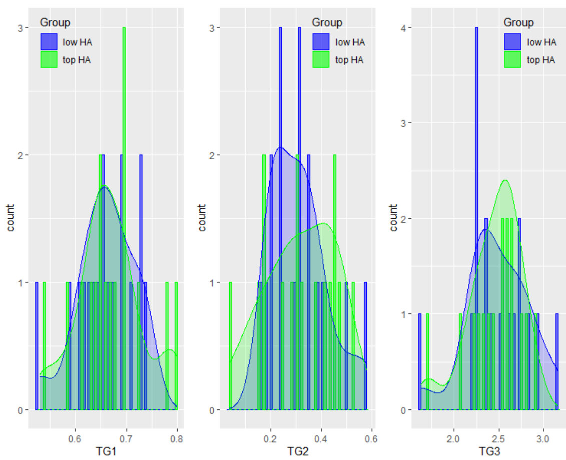


Figure 36: Histograms visualising the performances of the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and of the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{topHA} (SD_{topHA})	M_{lowHA} (SD_{lowHA})	t	p	d (CI)
General Reasoning Ability (KFT)	57.94 (3.34)	56.12 (2.57)	1.75	0.091	0.61 (−0.11, 1.34)
Mathematics Posttest (Post)	47.50 (5.99)	42.12 (2.42)	3.35	0.003	1.19 (0.42, 1.96)
Mathematics Power Test (Power)	14.56 (1.59)	2.63 (1.53)	21.97	0.000	7.66 (5.61, 9.71)
Mathematics Speed Test (Speed)	36.06 (3.42)	29.41 (2.90)	6.02	0.000	2.11 (1.22, 2.99)
WM Storage-Processing (TG1)	0.67 (0.07)	0.66 (0.06)	0.32	0.754	0.11 (−0.60, 0.82)
WM Supervision (TG2)	0.31 (0.14)	0.31 (0.11)	0.04	0.971	0.01 (−0.70, 0.72)
WM Relational Integration (TG3)	2.45 (0.29)	2.49 (0.35)	−0.30	0.768	−0.10 (−0.81, 0.61)

Table 11: Group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 12), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 12), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 12).

As a first observation, the proper assignment of the two subgroups should be noted: The two subgroups do not only substantially differ in performance on the Mathematics Power Test, but also on the other two mathematics tests.

The results associated with the General Reasoning Ability Test suggest a medium positive effect, while effect sizes ranging from −0.11, a negligible effect, to 1.34, a large positive effect, would be in line with the data as well. In terms of working memory functions, the point estimate for Cohen's d related to the functional category of Storage-Processing shows a negligible effect. However, effect sizes ranging from −0.60, a medium negative effect, to 0.82, a large positive effect, would also be compatible with the data. Moreover, the two subgroups do not seem to differ regarding the working memory function of Supervision. Finally, the point estimate for Cohen's d associated with the working memory function of Relational Integration indicates a negligible effect, while effect sizes ranging from −0.81, a large negative effect, to 0.61, a medium positive effect, would be supported by the data as well.

Individual Subtests or Tasks Table 12 lists the subtests or tasks of the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests for a more detailed group comparison between the two subgroups. While KFT_N, KFT_Q and KFT_V describe the non-verbal (N), quantitative (Q) and verbal (V) subscales of the General Reasoning Ability Test (KFT), Power_T1 - Power_T6 relate to the scores in Task 1 - Task 6 in the Mathematics Power Test (Power). In addition, for each functional category of working memory (TG1 - TG3), the corresponding figural (f), numerical (n) and verbal (v) subtests are presented. Table 12 arranges the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in the same way as Table 11. As a remark, it is important to mention that KFT and Power in Table 11 refer to the sum scores of the associated subtests or tasks displayed in Table 12, while TG1 - TG3 in Table 11 represent the mean scores of the related subtests collected in Table 12.

Test	M_{topHA} (SD_{topHA})	M_{lowHA} (SD_{lowHA})	t	p	d (CI)
KFT_N	23.44 (1.21)	22.29 (2.11)	1.92	0.066	0.66 (−0.07, 1.39)
KFT_Q	18.69 (1.54)	17.71 (2.39)	1.41	0.170	0.49 (−0.24, 1.21)
KFT_V	15.81 (2.74)	16.12 (2.29)	−0.35	0.732	−0.12 (−0.83, 0.59)
Power_T1	2.73 (0.47)	0.57 (0.58)	11.82	0.000	4.09 (2.84, 5.34)
Power_T2	2.84 (0.29)	0.63 (0.59)	13.75	0.000	4.70 (3.32, 6.07)
Power_T3	1.89 (0.63)	0.56 (0.86)	5.09	0.000	1.76 (0.92, 2.59)
Power_T4	2.58 (0.45)	0.57 (0.61)	10.74	0.000	3.71 (2.54, 4.88)
Power_T5	2.22 (0.71)	0.07 (0.25)	11.51	0.000	4.11 (2.86, 5.36)
Power_T6	2.30 (0.60)	0.22 (0.36)	11.93	0.000	4.22 (2.94, 5.49)
TG1_f	0.90 (0.09)	0.88 (0.11)	0.72	0.480	0.25 (−0.47, 0.96)
TG1_n	0.51 (0.10)	0.49 (0.08)	0.48	0.637	0.17 (−0.54, 0.88)
TG1_v	0.59 (0.10)	0.61 (0.08)	−0.62	0.541	−0.22 (−0.93, 0.50)
TG2_f	0.32 (0.16)	0.33 (0.22)	−0.14	0.893	−0.05 (−0.76, 0.66)
TG2_n	0.36 (0.18)	0.29 (0.20)	1.10	0.280	0.38 (−0.33, 1.10)
TG2_v	0.26 (0.13)	0.32 (0.12)	−1.34	0.189	−0.47 (−1.19, 0.25)
TG3_f	2.53 (0.41)	2.69 (0.50)	−1.02	0.317	−0.35 (−1.07, 0.36)
TG3_n	2.45 (0.33)	2.36 (0.52)	0.59	0.558	0.20 (−0.51, 0.92)
TG3_v	2.38 (0.56)	2.41 (0.53)	−0.14	0.887	−0.05 (−0.76, 0.66)

Table 12: Group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

The two subgroups clearly differ in performance on all tasks of the Mathematics Power Test (Power_T1 - Power_T6), which is in line with the proper assignment of the two subgroups.

While there could be group differences in performance on the non-verbal subtest of the General Reasoning Ability Test (KFT_N), the point estimate for Cohen's d related to the verbal subtest (KFT_V) indicates a negligible effect. Moreover, the results associated with the quantitative subtest (KFT_Q) show a small positive effect, with effect sizes ranging from small negative effects to large positive effects being in line with the data as well. The point estimates for Cohen's d related to the figural and verbal subtest of the working memory function Storage-Processing (TG1_f and TG1_v) specify a small positive effect and a small negative effect respectively. Moreover, for the numerical subtest (TG1_n) the results suggest a negligible effect, while effect sizes ranging from medium negative effects to large positive effects would also be supported by the data. With respect to the functional category of Supervision, the two subgroups do not appear to differ in performance on the figural subtest (TG2_f). On the other hand, the point estimate for Cohen's d associated with the numerical subtest (TG2_n) indicates a small positive effect. Nonetheless, effect sizes ranging from -0.33, a small negative effect, to 1.10, a large positive effect, would be compatible with the data as well. For the verbal subtest (TG2_v) the results show a small negative effect, with effect sizes ranging from large negative effects to small positive effects also being in line with the data. Lastly, substantial group differences in performance cannot be observed for the verbal subtest of the working memory function Relational Integration (TG3_v), while the point estimates for Cohen's d related to the figural and numerical subtest (TG3_f and TG3_n) specify a small negative effect and a small positive effect respectively.

5.4.2 Mathematical Self-Concept and Interest in Mathematics

As an additional comparison between the two subgroups shown in Figure 33, the two subgroups have been compared with respect to their mathematical self-concept and their interest in mathematics as well. This subsection presents the corresponding results.

Sum Scores and Mean Scores Figure 37 depicts the histograms related to the mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the two subgroups respectively. In addition to this graphical representation, Table 13 lists the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals. Moreover, to enable an overview on the various results,

Table 13 shows the results associated with the mathematics tests and the General Reasoning Ability Test already presented in Table 11. The structure of Table 13 is analogous to the one of Table 11.

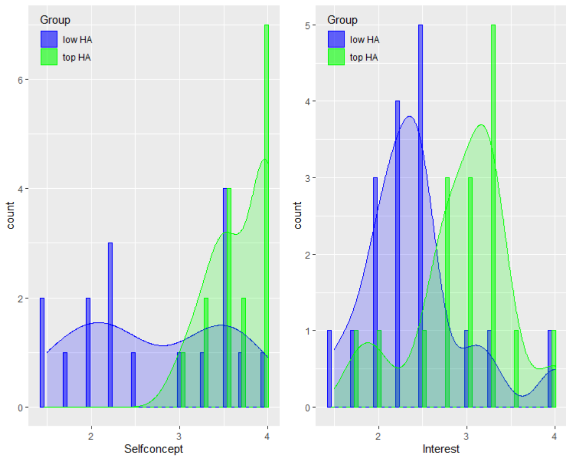


Figure 37: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{topHA} (SD_{topHA})	M_{lowHA} (SD_{lowHA})	t	p	d (CI)
General Reasoning Ability (KFT)	57.94 (3.34)	56.12 (2.57)	1.75	0.091	0.61 (−0.11, 1.34)
Mathematics Posttest (Post)	47.50 (5.99)	42.12 (2.42)	3.35	0.003	1.19 (0.42, 1.96)
Mathematics Power Test (Power)	14.56 (1.59)	2.63 (1.53)	21.97	0.000	7.66 (5.61, 9.71)
Mathematics Speed Test (Speed)	36.06 (3.42)	29.41 (2.90)	6.02	0.000	2.11 (1.22, 2.99)
Self-Concept Questionnaire Items (Selfconcept)	3.69 (0.34)	2.71 (0.83)	4.48	0.000	1.53 (0.72, 2.33)
Interest Questionnaire Items (Interest)	2.95 (0.55)	2.41 (0.59)	2.74	0.010	0.95 (0.20, 1.70)

Table 13: Group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 12), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 12), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 14) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 14) respectively.

Substantial group differences can be observed for both, the mean scores of the Self-Concept Questionnaire Items and the mean scores of the Interest Questionnaire Items.

Individual Questionnaire Items In order to analyse the group comparison between the two subgroups with respect to the Self-Concept Questionnaire Items and the Interest Questionnaire Items in more detail, Table 14 collects the individual Questionnaire Items. While Selfconcept_QI1 - Selfconcept_QI4 correspond to the Self-Concept Questionnaire Items, Interest_QI5 - Interest_QI8 represent the Interest Questionnaire Items. The structure of the related descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in Table 14 is analogous to the one in Table 13. As a remark, it is important to mention that Selfconcept and Interest in Table 13 refer to the mean scores of Selfconcept_QI1 - Selfconcept_QI4 and of Interest_QI5 - Interest_QI8 respectively.

Test	M_{topHA} (SD_{topHA})	M_{lowHA} (SD_{lowHA})	t	p	d (CI)
Selfconcept_QI1	3.75 (0.45)	2.71 (1.10)	3.60	0.002	1.22 (0.45, 2.00)
Selfconcept_QI2	3.63 (0.50)	2.65 (0.70)	4.63	0.000	1.60 (0.78, 2.41)
Selfconcept_QI3	3.50 (0.73)	2.35 (1.17)	3.40	0.002	1.17 (0.40, 1.94)
Selfconcept_QI4	3.88 (0.34)	3.12 (0.86)	3.37	0.003	1.15 (0.38, 1.91)
Interest_QI5	2.69 (0.79)	2.47 (0.94)	0.72	0.479	0.25 (−0.46, 0.96)
Interest_QI6	2.69 (0.70)	2.06 (0.83)	2.36	0.025	0.82 (0.08, 1.56)
Interest_QI7	2.94 (0.77)	2.12 (0.93)	2.77	0.010	0.96 (0.21, 1.71)
Interest_QI8	3.50 (0.63)	3.00 (0.79)	2.01	0.053	0.70 (−0.04, 1.43)

Table 14: Group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

As a first observation, it is important to note that the two subgroups differ substantially on all Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4). With respect to the Interest Questionnaire Items, the results related to the fifth item (Interest_QI5) show a small positive effect. Nonetheless, effect sizes ranging from small negative effects to large positive effects would be supported by the data as well. Finally, there seem to be group differences on the remaining items (Interest_QI6 - Interest_QI8).

5.4.3 Summary of the Results

The results from numerous group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test were reported in Subsections 5.4.1 and 5.4.2. Table 15 lists these results in descending order of the respective effect sizes in terms of Cohen's d. This subsection aims at giving a summary of them.

Test	M_{topHA} (SD_{topHA})	M_{lowHA} (SD_{lowHA})	t	p	d (CI)
Power	14.56 (1.59)	2.63 (1.53)	21.97	0.000	7.66 (5.61, 9.71)
Power_T2	2.84 (0.29)	0.63 (0.59)	13.75	0.000	4.70 (3.32, 6.07)
Power_T6	2.30 (0.60)	0.22 (0.36)	11.93	0.000	4.22 (2.94, 5.49)
Power_T5	2.22 (0.71)	0.07 (0.25)	11.51	0.000	4.11 (2.86, 5.36)
Power_T1	2.73 (0.47)	0.57 (0.58)	11.82	0.000	4.09 (2.84, 5.34)
Power_T4	2.58 (0.45)	0.57 (0.61)	10.74	0.000	3.71 (2.54, 4.88)
Speed	36.06 (3.42)	29.41 (2.90)	6.02	0.000	2.11 (1.22, 2.99)
Power_T3	1.89 (0.63)	0.56 (0.86)	5.09	0.000	1.76 (0.92, 2.59)
Selfconcept_QI2	3.63 (0.50)	2.65 (0.70)	4.63	0.000	1.60 (0.78, 2.41)
Selfconcept	3.69 (0.34)	2.71 (0.83)	4.48	0.000	1.53 (0.72, 2.33)
Selfconcept_QI1	3.75 (0.45)	2.71 (1.10)	3.60	0.002	1.22 (0.45, 2.00)
Post	47.50 (5.99)	42.12 (2.42)	3.35	0.003	1.19 (0.42, 1.96)
Selfconcept_QI3	3.50 (0.73)	2.35 (1.17)	3.40	0.002	1.17 (0.40, 1.94)
Selfconcept_QI4	3.88 (0.34)	3.12 (0.86)	3.37	0.003	1.15 (0.38, 1.91)
Interest_QI7	2.94 (0.77)	2.12 (0.93)	2.77	0.010	0.96 (0.21, 1.71)
Interest	2.95 (0.55)	2.41 (0.59)	2.74	0.010	0.95 (0.20, 1.70)
Interest_QI6	2.69 (0.70)	2.06 (0.83)	2.36	0.025	0.82 (0.08, 1.56)
Interest_QI8	3.50 (0.63)	3.00 (0.79)	2.01	0.053	0.70 (−0.04, 1.43)
KFT_N	23.44 (1.21)	22.29 (2.11)	1.92	0.066	0.66 (−0.07, 1.39)
KFT	57.94 (3.34)	56.12 (2.57)	1.75	0.091	0.61 (−0.11, 1.34)
KFT_Q	18.69 (1.54)	17.71 (2.39)	1.41	0.170	0.49 (−0.24, 1.21)
TG2_n	0.36 (0.18)	0.29 (0.20)	1.10	0.280	0.38 (−0.33, 1.10)
Interest_QI5	2.69 (0.79)	2.47 (0.94)	0.72	0.479	0.25 (−0.46, 0.96)
TG1_f	0.90 (0.09)	0.88 (0.11)	0.72	0.480	0.25 (−0.47, 0.96)
TG3_n	2.45 (0.33)	2.36 (0.52)	0.59	0.558	0.20 (−0.51, 0.92)
TG1_n	0.51 (0.10)	0.49 (0.08)	0.48	0.637	0.17 (−0.54, 0.88)
TG1	0.67 (0.07)	0.66 (0.06)	0.32	0.754	0.11 (−0.60, 0.82)
TG2	0.31 (0.14)	0.31 (0.11)	0.04	0.971	0.01 (−0.70, 0.72)
TG2_f	0.32 (0.16)	0.33 (0.22)	−0.14	0.893	−0.05 (−0.76, 0.66)
TG3_v	2.38 (0.56)	2.41 (0.53)	−0.14	0.887	−0.05 (−0.76, 0.66)
TG3	2.45 (0.29)	2.49 (0.35)	−0.30	0.768	−0.10 (−0.81, 0.61)
KFT_V	15.81 (2.74)	16.12 (2.29)	−0.35	0.732	−0.12 (−0.83, 0.59)
TG1_v	0.59 (0.10)	0.61 (0.08)	−0.62	0.541	−0.22 (−0.93, 0.50)
TG3_f	2.53 (0.41)	2.69 (0.50)	−1.02	0.317	−0.35 (−1.07, 0.36)
TG2_v	0.26 (0.13)	0.32 (0.12)	−1.34	0.189	−0.47 (−1.19, 0.25)

Table 15: Summary of all group comparisons between the 25 percent of high-achievers with the highest performance in the Mathematics Power Test (topHA) and the 25 percent of high-achievers with the lowest performance in the Mathematics Power Test (lowHA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Large effect sizes in terms of Cohen's d can be observed for the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), which indicates the proper assignment of the two subgroups. Furthermore, substantial group differences with respect to the Mathematics Speed Test

(Speed) seem to be present as well. The point estimate for Cohen's d referring to the Self-Concept Questionnaire Items (Selfconcept) is larger than the one related to the Interest Questionnaire Items (Interest). Moreover, for the Mathematics Posttest (Post), the point estimate for Cohen's d is larger than the one associated with the General Reasoning Ability Test (KFT). Finally, the point estimate for Cohen's d related to the General Reasoning Ability Test (KFT) is larger than the point estimates referring to the working memory tests.

5.4.4 Concluding Remark

To conclude Section 5.4, the following remark should be added: For a broader perspective on subgroup comparisons within the group of high-achievers, Section A.5 in the appendix presents an additional subgroup comparison.

6 General Discussion

The overarching research aim of the current study was to further assess the role of working memory in advanced school mathematics by studying the interplay between general reasoning abilities, working memory functions and math performance. In particular, the focus lay on investigating discrepancies between students' actual math performance and the respective math performance, which would be expected based on their general reasoning abilities. To this end, a sample consisting of high-achieving, over-achieving and under-achieving students at the Swiss Gymnasium, who formerly participated in the TraM-Study (a large scale study studying the learning transfer within mathematics and from mathematics to physics, see Subsection 4.1.1), was selected. The sample selection was based on students' performances in two of the measures of the TraM-Study, namely in the Mathematics Posttest on mathematical functions (see Subsection 4.1.2) and in the General Reasoning Ability Test (three subtests from the *Kognitiver Fähigkeits-Test (KFT)*, Heller & Perleth, 2000, see Subsection 4.1.2). More concretely, the sample of the current study comprised 67 high-achievers (66% female, 33% male), 19 over-achievers (53% female, 47% male) and 34 under-achievers (50% female, 50% male) with a mean age of 16.3 years. Despite the female majority in the group of high-achieving students, the three groups appear to be sufficiently balanced in terms of their gender ratios.

All study participants of the current study underwent the following tests: A working memory test battery comprising nine different tests (assembled by von Bastian & Oberauer, 2013, see Subsection 4.2.2), the Mathematics Speed Test (see Subsection 4.2.3) and the Mathematics Power Test (see Subsection 4.2.4). Moreover, they filled in a questionnaire on mathematical self-concept and interest in mathematics (most questionnaire items were utilised in *PISA 2000*, the items were taken from Gaspard et al., 2015, 2018; Mang et al., 2018; Trautwein et al., 2006, see Subsection 4.2.1). Even though the students participating in the current study came from different classes, it can be presumed that they were familiar with the mathematical contents covered in the mathematics measures of the current study, as all classes in the TraM-Study followed the same syllabus on these contents.

As described in Section 3.1 and for example demonstrated in the review by Raghubar et al. (2010), most studies focusing on the relation between working memory and math performance were conducted with primary school pupils or with children experiencing math difficulties. One of the contributions of the current study to ongoing research on the association between working memory functions and math achievement is therefore the continuative investigation of this relationship in adolescents. With respect to the assessment of math achievement, the current study tried to extend studies relying on arithmetic tasks by including mathematics tests, which cover other aspects of math performance as well: On the one hand, the Mathematics Speed Test (see Subsection 4.2.3) was designed in such a way that study participants, who built useful knowledge representations and knowledge chunks on mathematical functions would manage to solve the tasks quickly and correctly. On the other hand, the Mathematics Power Test (see Subsection 4.2.4) contains mathematical problem solving tasks, which draw on both, the conceptual knowledge and the procedural knowledge of the study participants. In terms of working memory measures, the current study made a contribution to a broader perspective on working memory by utilising an extensive working memory test battery suggested by von Bastian and Oberauer (2013), which targets all facets of working memory comprised in the facet model (Oberauer et al., 2000, 2003; Süß et al., 2002, a short description of the facet model is provided in Section 2.3).

6.1 Reasons for Over-Achievement

With respect to the group comparison between high-achievers and over-achievers, the main focus was to explore the following research question: Are there any indications that over-achievers may be able to partly compensate the difference in general reasoning abilities by efficiently chunking their knowledge and successfully organising and representing it? While the group comparison between high-achievers and over-achievers did not reveal substantial group differences regarding the Mathematics Power Test, it could indeed be observed that the point estimate for Cohen's d associated with the Mathematics Power Test appeared to be larger than the one related to the Mathematics Speed Test. Even though these findings do not provide a clear answer to the research question above, they could be seen as a first indication pointing towards the above-mentioned consideration: Given that the Mathematics Speed Test targets the knowledge representations of the students, successful chunking potentially explains why over-achievers managed to show a more similar performance to the high-achievers in the Mathematics Speed Test than in the Mathematics Power Test. This result could add to the research findings of Castejón et al. (2016), who reported that over-achievers heavily use different learning strategies in order to achieve higher scholastic performances.

Regarding the evaluation of the questionnaire on mathematical self-concept and interest in mathematics, the group comparison between high-achievers and over-achievers revealed the following: Both, the point estimate for Cohen's d referring to the mean scores of the Self-Concept Questionnaire Items as well as the one related to the mean scores of the Interest Questionnaire Items specified negligible effects. However, a noticeable result in this group comparison was that the over-achievers seemingly agreed more on the statement "it is important to me to be good at mathematics" compared to the high-achievers - the point estimate for Cohen's d with respect to this item (Interest_QI8, the item was taken from Gaspard et al., 2015, 2018) demonstrated a medium negative effect. Overall, the findings suggest that the mathematical self-concept of the over-achieving students in the sample of the current study is comparable to the one of the high-achieving students, and that the over-achievers are as interested in mathematics as the high-achievers. A possible interpretation of these results could be that the motivational aspects of mathematical self-concept and interest in mathematics play a role for over-achievement in mathematics, i.e. that these factors help students to over-achieve in mathematics.

6.2 Reasons for Under-Achievement

For the group comparison between high-achievers and under-achievers, the research aim was to further explore the question why under-achieving students do not (fully) translate their potential in terms of general reasoning abilities into math achievement. In particular, in addition to the investigation of the motivational factors of mathematical self-concept and interest in mathematics, the focus of the current study lay on assessing the potential role of working memory functions for under-achievement in mathematics.

Regarding working memory functions, an important observation was that substantial group differences with respect to the functional category of Storage-Processing were present in the group comparison between high-achievers and under-achievers. This finding revealed that differences in the working memory function Storage-Processing could be an additional explanatory factor for under-achievement in mathematics. At the same time, this result suggested that the role of the working memory function Storage-Processing appears to be complementary to the one of general reasoning abilities, as the group differences regarding the working memory function Storage-Processing emerged, even tough the two groups did not seem to differ in terms of general reasoning abilities. This observation is in line with other findings supporting the view that, although being closely related, general reasoning abilities and working memory functions are two separable constructs (e.g., Chooi, 2012; Conway et al., 2003). With respect to the other two functional categories, the point estimate for Cohen's d showed a negligible effect related to the working memory function Supervision and a small effect referring to the working memory function Relational Integration.

Another noticeable result was that in the group comparison between high-achievers and under-achievers, the two groups did not seem to differ with respect to their interest in mathematics. Thus, while motivational factors can play a role in explaining under-achievement in mathematics (e.g., Fong & Kremer, 2020), in the sample of the current study, under-achievement cannot be traced back to a lack of interest in mathematics. On the other hand, there were clear group differences related to the mathematical self-concept of the students in the group comparison between high-achievers and under-achievers (i.e., the mathematical self-concept of the under-achieving students appeared to be considerably lower than the one of the high-achieving students).

In summary, the findings associated with the group comparison between high-achieving and under-achieving students suggested that a poorer performance of the working memory function Storage-Processing as well as a lower mathematical self-concept could be possible sources of under-achievement in mathematics. At the same time, in the sample of the current study, under-achievement did not appear to be caused by a scarce interest in mathematics.

6.3 Who Are the Top High-Achievers? Differences within the High-Achievers

The research goal behind the subgroup comparison within high-achievers was to analyse how the following subgroups may be different from each other: The subgroup of high-achieving students with top scores in the Mathematics Power Test and the subgroup of high-achieving students with lower scores in the Mathematics Power Test.

The two subgroups appeared to differ in performance on the Mathematics Speed Test: The subgroup of high-achievers with top scores in the Mathematics Power Test clearly outperformed the subgroup of high-achievers with lower scores in the Mathematics Power Test. This result shows that the two subgroups can not only be distinguished with respect to their performances in the Mathematics Power Test, but also regarding their performances in the Mathematics Speed Test.

While the subgroup comparison revealed that the point estimate for Cohen's d related to the General Reasoning Ability Test indicated a medium positive effect in favour of the subgroup of high-achieving students with top scores in the Mathematics Power Test, it showed that the point estimates for Cohen's d referring to the measures of the different working memory functions suggested negligible effects. This result potentially provides another indication that general reasoning abilities and working memory functions seem to be distinguishable, which is in line with previous research (e.g., Chooi, 2012; Conway et al., 2003). Furthermore, it illustrated that the working memory measures did not discriminate between the two subgroups.

With respect to the mathematical self-concept of the two subgroups and their interest in mathematics, clear group differences were present regarding both, the mean scores of the Self-Concept Questionnaire Items as well as the mean scores of the Interest Questionnaire Items: The high-achievers with top scores in the Mathematics Power Test appeared to have a higher mathematical self-concept and to be more interested in mathematics compared to the high-achievers with lower scores in the Mathematics Power Test. These findings suggest that within the group of high-achieving students, the motivational aspects of mathematical self-concept and interest in mathematics play a role in differentiating between the two subgroups.

6.4 Pedagogical Consequences

The sample selection process of the current study (described in Subsection 4.1.3) revealed that there is a noticeable number of under-achieving students. For the under-achievers in the sample of the current study, the root of their under-achievement does not seem to be a lack of interest in mathematics (as explained in Section 6.2). The results regarding the group comparison between high-achievers and under-achievers should therefore also be discussed in terms of possible educational implications: How could the surrounding conditions in educational settings be optimised in order to help under-achieving students to better unfold their potential?

As discussed in Section 6.2, clear group differences in performance related to the working memory function Storage-Processing could be observed in the group comparison between high-achieving students and under-achieving students. For the preparation of learning activities it might therefore be helpful to keep the potential impact of these differences in mind. In particular, it could be worthwhile to consider the load on working memory of the respective learning activities and to analyse whether this load could be reduced to a certain degree. For example, could a mathematical problem solving task be restructured in such a way that it poses lower demands on the working memory function Storage-Processing while still covering the same mathematical content? In the context of cognitive load theory (Sweller, 1988, a historical review is given in Moreno & Park, 2010), different suggestions have been proposed to reduce the cognitive load of learning activities: Among other approaches, worked examples (for a review, see Sweller, Van Merrienboer, & Paas, 1998), completion problems (for an overview, see Sweller et al., 1998) and scaffolding (Van Merriënboer, Kirschner, & Kester, 2003) were reported as possible strategies.

As pointed out in Section 6.2, in the sample of the current study, the under-achievers showed a noticeably lower mathematical self-concept compared to the high-achievers. Regarding the relations between academic self-concept and academic achievement, Guay, Marsh, and Boivin (2003) found that the effects are reciprocal, i.e. that achievement influences self-concept and that self-concept has an impact on achievement. On the one hand, the observed group differences in mathematical self-concept between under-achieving students and high-achieving students in the current study might therefore reflect the group differences in math performance to a certain extent. On the other hand, given the influence of academic self-concept on academic achievement (see also Kriegbaum et al., 2015; Marsh, 1990), supporting under-achieving students in improving their mathematical self-concept could be a fruitful approach to help them unfolding their potential better.

6.5 Limitations of the Current Study and Suggestions for Future Research

As discussed in Subsection 1.2.1, general reasoning abilities play an important role for math performance. It was therefore to be expected that the group of over-achieving students would be smaller than other groups of students. Nevertheless, more research with a larger sample of over-achievers and possibly additional tests is needed to gain a deeper and more extensive understanding of over-achievement in mathematics and its characteristics. The findings of the current study indicated that different aspects, such as the way how over-achieving students organise and structure their knowledge or the mathematical self-concept of over-achieving students and their interest in mathematics, should be taken into account when assessing over-achievement in mathematics. In future studies, the following research questions should be investigated further: To what extent do over-achievers manage to (partly) compensate the difference in general reasoning abilities by efficiently structuring their knowledge, having a good mathematical self-concept, being interested in mathematics, and applying potential other approaches? What are the limitations of these strategies?

As a more general remark regarding the sample size of the sample of the current study, it should be considered that the data collection of the current study was conducted during the Covid-19 pandemic. Even though various safety and hygiene measures were implemented to protect the health of the study participants, it might be that under more "normal" circumstances, more students could have been recruited for the current study.

The results of the current study demonstrated that group differences in performance related to the working memory function Storage-Processing were present in both group comparisons, in the one between high-achievers and over-achievers as well as in the one between high-achievers and under-achievers. The functional category of Storage-Processing was measured by different complex span tasks, which are an established measure of working memory (e.g., Conway et al., 2005). Thus, some complex span tasks appeared to differentiate between high-achieving and over-achieving students and between high-achieving and under-achieving students respectively. It could therefore be reasonable to incorporate complex span tasks in further investigations of over-achievement and under-achievement in mathematics. At the same time, as recommended by Lewandowsky et al. (2010), it should be considered to include other types of working memory measures in future studies as well.

The results of the current study emphasised the importance of considering cognitive factors in addition to motivational factors when assessing under-achievement in mathematics, which was also pointed out by Owens et al. (2008). Furthermore, the findings demonstrated that different cognitive and motivational aspects should be taken into account for the investigation of over-achieving students and of high-achieving students as well.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological bulletin*, *131*(1), 30.

Allen, K., Giofrè, D., Higgins, S., & Adams, J. (2020a). Working memory predictors of mathematics across the middle primary school years. *British Journal of Educational Psychology*, *90*(3), 848–869.

Allen, K., Giofrè, D., Higgins, S., & Adams, J. (2020b). Working memory predictors of written mathematics in 7-to 8-year-old children. *Quarterly Journal of Experimental Psychology*, *73*(2), 239–248.

Allen, K., Giofrè, D., Higgins, S., & Adams, J. (2021). Using working memory performance to predict mathematics performance 2 years on. *Psychological Research*, *85*(5), 1986–1996.

Allen, K., Higgins, S., & Adams, J. (2019). The relationship between visuospatial working memory and mathematical performance in school-aged children: a systematic review. *Educational Psychology Review*, *31*(3), 509–531.

Alloway, T. P. (2010). *Improving working memory: Supporting students' learning*. Sage.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and iq in academic attainment. *Journal of experimental child psychology*, *106*(1), 20–29.

Alloway, T. P., Banner, G. E., & Smith, P. (2010). Working memory and cognitive styles in adolescents' attainment. *British Journal of Educational Psychology*, *80*(4), 567–581.

Alloway, T. P., & Copello, E. (2013). Working memory: The what, the why, and the how. *The Educational and Developmental Psychologist*, *30*(2), 105–118.

Alloway, T. P., Gathercole, S. E., Adams, A.-M., Willis, C., Eaglen, R., & Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *British Journal of Developmental Psychology*, *23*(3), 417–426.

Alloway, T. P., & Gregory, D. (2013). The predictive ability of iq and working memory scores in literacy in an adult population. *International Journal of Educational Research*, *57*, 51–56.

Auguie, B. (2017). gridextra: Miscellaneous functions for "grid" graphics [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=gridExtra` (R package version 2.3)

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417–423.

Baddeley, A. (2006). Working memory: An overview. *Working memory and education*, 1–31.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of*

learning and motivation (Vol. 8, pp. 47–89). Elsevier.
Berkowitz, M., Edelsbrunner, P., & Stern, E. (2022). The relation between working memory and mathematics performance among students in math-intensive stem programs. *Intelligence*, *92*, 101649.
Berkowitz, M., & Stern, E. (2018). Which cognitive abilities make the difference? predicting academic achievements in advanced stem studies. *Journal of Intelligence*, *6*(4), 48.
Blankson, A. N., & Blair, C. (2016). Cognition and classroom quality as predictors of math achievement in the kindergarten year. *Learning and Instruction*, *41*, 32–40.
Boehnke, K. (2008). Peer pressure: A cause of scholastic underachievement? a cross-cultural study of mathematical achievement among german, canadian, and israeli middle school students. *Social Psychology of Education*, *11*(2), 149–160.
Bonnot, V., & Croizet, J.-C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of experimental social psychology*, *43*(6), 857–866.
Bresgi, L., Alexander, D., & Seabi, J. (2017). The predictive relationships between working memory skills within the spatial and verbal domains and mathematical performance of grade 2 south african learners. *International Journal of Educational Research*, *81*, 1–10.
Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly journal of experimental psychology*, *10*(1), 12–21.
Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology*, *33*(3), 205–228.
Campos, I. S., Almeida, L. S., Ferreira, A. I., Martinez, L. F., & Ramalho, G. (2013). Cognitive processes and math performance: a study with children at third grade of basic education. *European Journal of Psychology of Education*, *28*(2), 421–436.
Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences*, *34*(3), 113.
Castejón, J. L., Gilar, R., Veas, A., & Miñano, P. (2016). Differences in learning strategies, goal orientations, and self-concept between overachieving, normal-achieving, and underachieving secondary students. *Frontiers in Psychology*, *7*, 1438.
Caviola, S., Colling, L. J., Mammarella, I. C., & Szűcs, D. (2020). Predictors of mathematics in primary school: Magnitude comparison, verbal and spatial working memory measures. *Developmental science*, *23*(6), e12957.
Chooi, W.-T. (2012). Working memory and intelligence: a brief review. *Journal of Educational and Developmental Psychology*, *2*(2), 42.
Cognitive ability & plasticity lab. (n.d.). `http://www.claudiavonbastian`

.com/people.html. (Accessed: 2022-09-13)
Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual differences*, *34*(1), 33–39.
Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769–786.
Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences*, *7*(12), 547–552.
Cormier, D. C., Bulut, O., McGrew, K. S., & Singh, D. (2017). Exploring the relations between cattell–horn–carroll (chc) cognitive abilities and mathematics achievement. *Applied Cognitive Psychology*, *31*(5), 530–538.
Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.
de Abreu, P. M. E., Conway, A. R., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552–561.
Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21.
De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of experimental child psychology*, *103*(2), 186–201.
Dobbs, J., Doctoroff, G. L., Fisher, P. H., & Arnold, D. H. (2006). The association between preschool children's socio-emotional functioning and their mathematical skills. *Journal of Applied Developmental Psychology*, *27*(2), 97–108.
Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, *128*(3), 309.
Fanari, R., Meloni, C., & Massidda, D. (2019). Visual and spatial working memory abilities predict early math skills: a longitudinal study. *Frontiers in psychology*, *10*, 2460.
Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of cattell-horn-carroll (chc) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155–171.
Fong, C. J., & Kremer, K. P. (2020). An expectancy-value approach to math underachievement: Examining high school achievement, college

attendance, and stem interest. *Gifted Child Quarterly*, *64*(2), 67–84.

Friso-Van den Bos, I., Van der Ven, S. H., Kroesbergen, E. H., & Van Luit, J. E. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational research review*, *10*, 29–44.

Froiland, J. M., & Davison, M. L. (2016). The longitudinal influences of peers, parents, motivation, and mathematics course-taking on high school math achievement. *Learning and Individual Differences*, *50*, 252–259.

Fry, A. F., & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological psychology*, *54*(1-3), 1–34.

García, T., Rodríguez, C., Betts, L., Areces, D., & González-Castro, P. (2016). How affective-motivational variables and approaches to learning predict mathematics achievement in upper elementary levels. *Learning and Individual Differences*, *49*, 25–31.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: gender differences in value beliefs about math. *Journal of educational psychology*, *107*(3), 663.

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, *52*, 1–14.

Gathercole, S. E., & Alloway, T. P. (2004). Working memory and classroom learning. *Dyslexia Review*, *15*, 4–9.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In *Working memory and education* (pp. 219–240). Elsevier.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six-and seven-year-old children. *Journal of educational psychology*, *92*(2), 377.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(1), 1–16.

Giofrè, D., Donolato, E., & Mammarella, I. C. (2018). The differential role of verbal and visuospatial working memory in mathematics and reading. *Trends in neuroscience and education*, *12*, 1–6.

Giofrè, D., Mammarella, I. C., & Cornoldi, C. (2013). The structure of working memory and how it relates to intelligence in children. *Intelligence*, *41*(5), 396–406.

Gottfredson, L. S. (1997). *Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography* (Vol. 24) (No. 1).

Citeseer.
Green, C. T., Bunge, S. A., Chiongbian, V. B., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of experimental child psychology*, *157*, 125–143.
Grigg, S., Perera, H. N., McIlveen, P., & Svetleff, Z. (2018). Relations among math self efficacy, interest, intentions, and achievement: A social cognitive perspective. *Contemporary Educational Psychology*, *53*, 73–86.
Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of educational psychology*, *95*(1), 124.
Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, *15*(3), 609–627.
Heller, K. A., & Perleth, C. (2000). Kognitiver fähigkeitstest für 4. bis 12. klassen, revision. beltz-test.
Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. *Conceptual and procedural knowledge: The case of mathematics*, *2*, 1–27.
Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. Retrieved from `https://www.jstatsoft.org/v42/i08/`
Hofer, S. I., & Stern, E. (2016). Underachievement in physics: When intelligent girls fail. *Learning and Individual Differences*, *51*, 119–131.
Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology*, *26*(3), 339–366.
Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology*, *20*(3), 123–140.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on ackerman, beier, and boyle (2005).
Kriegbaum, K., Jansen, M., & Spinath, B. (2015). Motivation: A predictor of pisa's mathematical competence beyond intelligence and prior test achievement. *Learning and individual differences*, *43*, 140–148.
Krouse, J. H., & Krouse, H. J. (1981). Toward a multimodal theory of academic underachievement. *Educational Psychologist*, *16*(3), 151–164.
Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., & Stanat, P. (2002). *Pisa 2000. dokumentation der erhebungsinstrumente.* Berlin, Germany: Max Planck Institut für Bildungsforschung.

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, *23*(1), 77–94.

LaRocque, M. (2008). Assessing perceptions of the environment in elementary classrooms: The link with achievement. *Educational Psychology in Practice*, *24*(4), 289–305.

Lauermann, F., Meißner, A., & Steinmayr, R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology*, *112*(2), 364.

León, J., Núñez, J. L., & Liew, J. (2015). Self-determination and stem education: Effects of autonomy, motivation, and self-regulated learning on high school math achievement. *Learning and Individual Differences*, *43*, 156–163.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. (2010). A working memory test battery for matlab. *Behavior Research Methods*, *42*(2), 571–585.

Li, Y., & Geary, D. C. (2013). Developmental gains in visuospatial memory predict gains in mathematics achievement. *PloS one*, *8*(7), e70160.

Li, Y., & Geary, D. C. (2017). Children's visuospatial memory predicts mathematics achievement through early adolescence. *PloS one*, *12*(2), e0172046.

Lotz, C., Schneider, R., & Sparfeldt, J. R. (2018). Differential relevance of intelligence and motivation for grades and competence tests in mathematics. *Learning and Individual Differences*, *65*, 30–40.

Mang, J., Ustjanzew, N., Schiepe-Tiska, A., Prenzel, M., Sälzer, C., Müller, K., & Gonzaléz Rodríguez, E. (2018). *Pisa 2012 skalenhandbuch: Dokumentation der erhebungsinstrumente*. Waxmann.

Männamaa, M., Kikas, E., Peets, K., & Palu, A. (2012). Cognitive correlates of math skills in third-grade students. *Educational Psychology*, *32*(1), 21–44.

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: a multiwave, longitudinal panel analysis. *Journal of educational psychology*, *82*(4), 646.

Mayes, S. D., Calhoun, S. L., Bixler, E. O., & Zimmerman, D. N. (2009). Iq and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, *19*(2), 238–241.

Mercader, J., Presentación, M.-J., Siegenthaler, R., Molinero, V., & Miranda, A. (2017). Motivation and mathematics performance: a longitudinal study in early educational stages. *Revista de Psicodidáctica (English ed.)*, *22*(2), 157–163.

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory compo-

nents to mathematics achievement in 2nd and 3rd graders. *Learning and individual differences*, *20*(2), 101–109.
Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.
Moenikia, M., & Zahed-Babelan, A. (2010). A study of simple and multiple relations between mathematics attitude, academic motivation and intelligence quotient with mathematics achievement. *Procedia-Social and Behavioral Sciences*, *2*(2), 1537–1542.
Monsell, S. (2003). Task switching. *Trends in cognitive sciences*, *7*(3), 134–140.
Moreno, R. E., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories.
Musso, M. F., Boekaerts, M., Segers, M., & Cascallar, E. C. (2019). Individual differences in basic cognitive processes and self-regulated learning: Their interaction effects on math performance. *Learning and Individual Differences*, *71*, 58–70.
Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence–their correlation and their relation: comment on ackerman, beier, and boyle (2005).
Oberauer, K., Süβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*(6), 641–652.
Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and individual differences*, *29*(6), 1017–1045.
Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193.
Owens, M., Stevenson, J., Norgate, R., & Hadwin, J. A. (2008). Processing efficiency theory in children: Working memory as a mediator between trait anxiety and academic performance. *Anxiety, Stress, & Coping*, *21*(4), 417–430.
Parhiala, P., Torppa, M., Vasalampi, K., Eklund, K., Poikkeus, A.-M., & Aro, T. (2018). Profiles of school motivation and emotional well-being among adolescents: Associations with math and reading performance. *Learning and Individual Differences*, *61*, 196–204.
Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology*, *82*(1), 42–63.
Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive development*, *22*(2), 165–184.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455.

Phillipson, S. N. (2008). The optimal achievement model and underachievement in hong kong: An application of the rasch model. *Psychology Science*, *50*(2), 147.

Portrat, S., Guida, A., Phénix, T., & Lemaire, B. (2016). Promoting the experimental dialogue between working memory and chunking: Behavioral data and simulation. *Memory & cognition*, *44*(3), 420–434.

Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, *20*(5), 446–451.

Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, *122*, 104–121.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and individual differences*, *20*(2), 110–122.

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, *21*(4), 387–399.

Revelle, W. (2020). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from `https://CRAN.R-project.org/package=psych` (R package version 2.0.12)

Ricker, T. J., AuBuchon, A. M., & Cowan, N. (2010). Working memory. *Wiley interdisciplinary reviews: Cognitive science*, *1*(4), 573–585.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of educational psychology*, *91*(1), 175.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM*, *49*(4), 599–611.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, *93*(2), 346.

RStudio Team. (2021). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from `http://www.rstudio.com/`

Schalk, L., Schumacher, R., Barth, A., & Stern, E. (2018). When problem-solving followed by instruction is superior to the traditional tell-and-practice sequence. *Journal of educational psychology*, *110*(4), 596.

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., … Crowley, J. (2021). Ggally: Extension to 'ggplot2' [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=GGally` (R package version 2.1.0)

Schneider, M., & Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: A multimethod approach. *Developmental psychology*, *46*(1), 178.

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of educational psychology*, *103*(4), 759.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426–445.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, *99*(2), 285.

Sepie, A. C., & Keeling, B. (1978). The relationship between types of anxiety and under-achievement in mathematics. *The Journal of Educational Research*, *72*(1), 15–19.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child development*, 973–980.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The journal of educational research*, *95*(6), 323–332.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly journal of experimental psychology*, *59*(4), 745–759.

Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and individual differences*, *19*(1), 80–90.

Stern, E. (2017). Individual differences in the learning potential of human beings. *npj Science of Learning*, *2*(1), 1–7.

Stevenson, C. E., Bergwerff, C. E., Heiser, W. J., & Resing, W. C. (2014). Working memory and dynamic measures of analogical reasoning as predictors of children's math and reading achievement. *Infant and Child Development*, *23*(1), 51–66.

Suárez-Álvarez, J., Fernández-Alonso, R., & Muñiz, J. (2014). Self-concept, motivation, expectations, and socioeconomic level as predictors of academic performance in mathematics. *Learning and Individual Differences*, *30*, 118–123.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288.

Swanson, L., & Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence*, *35*(2), 151–168.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257–285.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*(3), 251–296.

Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, *23*(2), 187.

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37.

Tikhomirova, T., Voronina, I., Marakshina, J., Nikulchev, E., Ayrapetyan, I., & Malykh, T. (2016). The relationship between non-verbal intelligence and mathematical achievement in high school students. In *Shs web of conferences* (Vol. 29, p. 02039).

Torchiano, M. (2020). effsize: Efficient effect size computation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=effsize` (R package version 0.8.1) doi: 10.5281/zenodo.1480624

Tosto, M. G., Asbury, K., Mazzocco, M. M., Petrill, S. A., & Kovas, Y. (2016). From classroom environment to mathematics achievement: The mediating role of self-perceived ability and subject interest. *Learning and individual differences*, *50*, 260–269.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of educational psychology*, *98*(4), 788.

van der Ven, S. H., van der Maas, H. L., Straatemeier, M., & Jansen, B. R. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learning and Individual Differences*, *27*, 182–192.

Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Memory & cognition*, *43*(3), 367.

Van Merriënboer, J. J., Kirschner, P. A., & Kester, L. (2003). Taking the

load off a learner's mind: Instructional design for complex learning. *Educational psychologist*, *38*(1), 5–13.
von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*(1), 36–58.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from `https://ggplot2.tidyverse.org`
Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A grammar of data manipulation [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=dplyr` (R package version 1.0.4)
Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence*, *53*, 166–180.
Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal of Educational Psychology*, *107*(4), 954.
Zheng, X., Swanson, H. L., & Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *Journal of experimental child psychology*, *110*(4), 481–498.
Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction*, *33*, 131–146.
Ziegler, E., & Stern, E. (2016). Consistent advantages of contrasted comparisons: Algebra learning under direct instruction. *Learning and Instruction*, *41*, 41–51.

A Appendix

A.1 Additional Group Comparison between High-Achievers and Over-Achievers I

For the first additional comparison between high-achievers and over-achievers, a subgroup of 32 high-achievers (22 female students (69%) and 10 male students (31%)) has been selected to be compared against the group of 19 over-achievers (10 female students (53%) and 9 male students (47%)). The subgroup of high-achievers has been determined using a cut-off point on the Mathematics Posttest scale. Figure 38 shows the selected high-achievers (marked with a green dot), as well as the over-achievers (marked with a red dot).

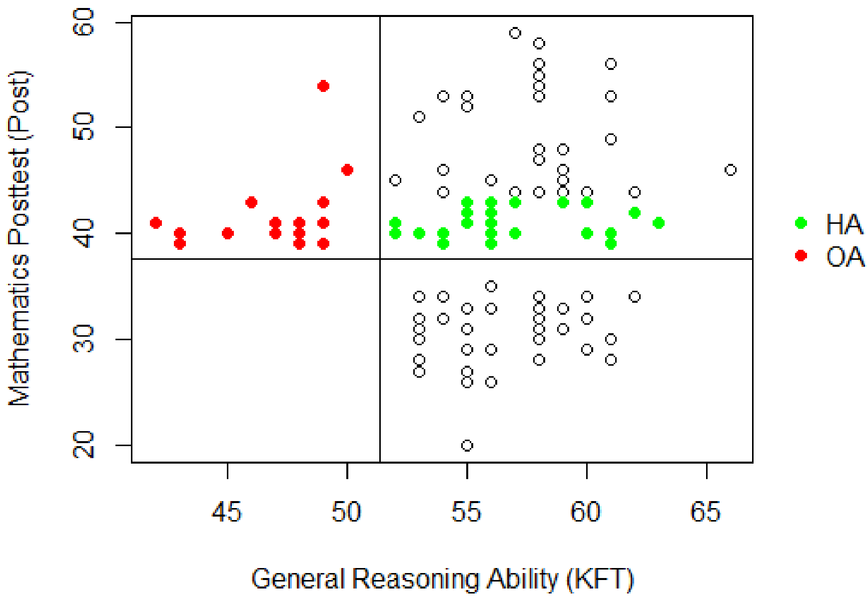


Figure 38: Selected high-achievers (marked with a green dot) and over-achievers (marked with a red dot).

A.1.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

This subsection discusses a comparison between the selected high-achievers and the over-achievers depicted in Figure 38 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms in Figures 39, 40 and 41 visualise the performances of the selected high-achievers (HA) and the over-achievers (OA) in the various tests. While Figure 39 shows the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed) are depicted in Figure 40. Finally, Figure 41 collects the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3). In addition to these graphical representations, the corresponding descriptive statistics of the performances of the selected high-achievers (HA) are given in the second column of Table 16, while the third column contains the descriptive statistics related to the performances of the over-achievers (OA). Moreover, the t-values and p-values of a two sample t-test comparing the two groups are presented in the fourth and fifth column of Table 16. Finally, the last column of Table 16 comprises the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals.

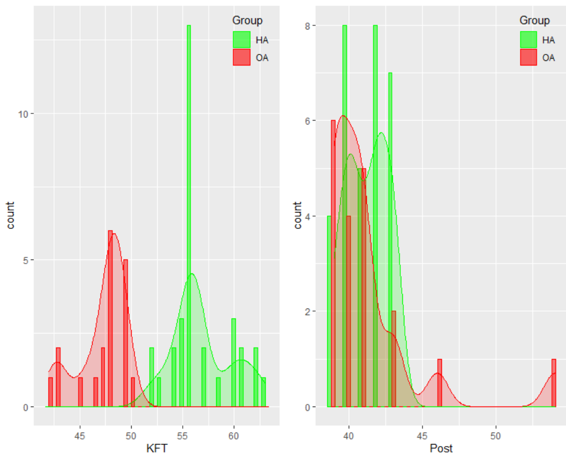


Figure 39: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

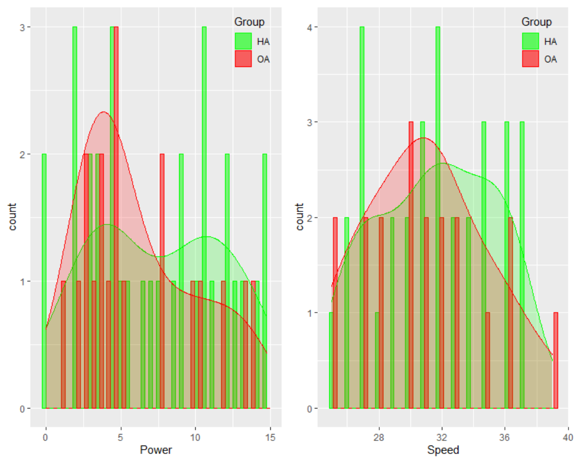


Figure 40: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

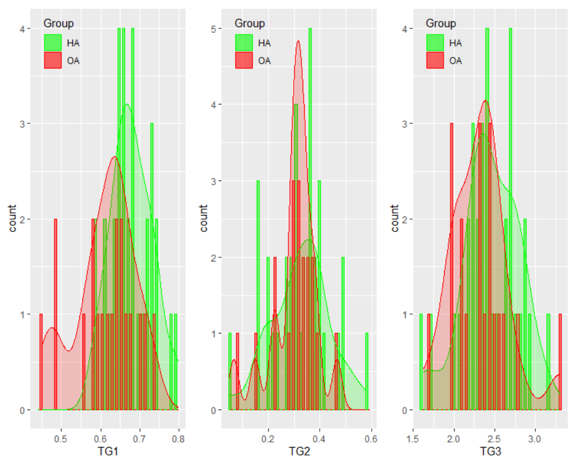


Figure 41: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.88 (2.88)	47.16 (2.32)	13.20	0.000	3.62 (2.69, 4.54)
Mathematics Posttest (Post)	41.19 (1.38)	41.32 (3.56)	−0.15	0.882	−0.05 (−0.64, 0.53)
Mathematics Power Test (Power)	7.46 (4.44)	6.13 (3.84)	1.13	0.266	0.31 (−0.27, 0.90)
Mathematics Speed Test (Speed)	31.56 (3.68)	30.95 (3.84)	0.56	0.578	0.16 (−0.42, 0.75)
WM Storage-Processing (TG1)	0.68 (0.05)	0.61 (0.08)	3.52	0.001	1.12 (0.50, 1.75)
WM Supervision (TG2)	0.33 (0.11)	0.30 (0.09)	0.85	0.402	0.23 (−0.35, 0.81)
WM Relational Integration (TG3)	2.46 (0.34)	2.32 (0.35)	1.36	0.182	0.40 (−0.19, 0.99)

Table 16: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 17), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 17), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 17).

As a first observation, it is important to note that the two groups have been properly assigned, as they clearly differ in their scores on the General Reasoning Ability Test, but they do not seem to differ in performance on the Mathematics Posttest.

It can be observed that the point estimate for Cohen's d associated with the Mathematics Power Test is larger than the one related to the Mathematics Speed Test. Regarding the working memory functions, the results indicate group differences in performance on the working memory test measuring the functional category of Storage-Processing. Moreover, the point estimate for Cohen's d related to the working memory function of Supervision suggests a small positive effect, while effect sizes ranging from −0.35, a small negative effect, to 0.81, a large positive effect, are also reasonably compatible with the data. Finally, the results with respect to the functional category of Relational Integration show a small positive effect as well. Nonetheless, effect sizes ranging from −0.19, a negligible effect, to 0.99, a large positive effect, would also be in line with the data.

Individual Subtests or Tasks For a more detailed evaluation of the group comparison between the selected high-achievers and the over-achievers with respect to the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests, Table 17 lists the respective subtests or tasks. KFT_N, KFT_Q and KFT_V relate to the non-verbal (N), quantitative (Q) and verbal (V) subscales of the General Reasoning Ability Test

(KFT), while Power_T1 - Power_T6 refer to the scores in Task 1 - Task 6 in the Mathematics Power Test (Power). Lastly, Table 17 presents the figural (f), numerical (n) and verbal (v) subtests for each functional category of working memory (TG1 - TG3). Table 17 structures the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in an analogous way as Table 16. As a remark, it is important to mention that KFT and Power in Table 16 represent the sum scores of the corresponding subtests or tasks given in Table 17, while TG1 - TG3 in Table 16 relate to the mean scores of the respective subtests collected in Table 17.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT_N	22.47 (1.81)	18.16 (2.48)	6.61	0.000	2.07 (1.36, 2.78)
KFT_Q	18.09 (1.86)	15.89 (2.90)	2.96	0.006	0.96 (0.34, 1.57)
KFT_V	16.31 (2.26)	13.11 (3.28)	3.76	0.001	1.20 (0.57, 1.82)
Power_T1	1.63 (1.08)	1.05 (0.85)	2.10	0.041	0.57 (−0.02, 1.16)
Power_T2	1.88 (1.09)	1.80 (0.96)	0.25	0.806	0.07 (−0.51, 0.65)
Power_T3	1.14 (0.83)	1.04 (0.74)	0.45	0.655	0.13 (−0.46, 0.71)
Power_T4	1.35 (0.84)	1.20 (0.85)	0.63	0.533	0.18 (−0.40, 0.77)
Power_T5	0.64 (0.90)	0.43 (0.76)	0.87	0.388	0.24 (−0.34, 0.83)
Power_T6	0.83 (0.83)	0.61 (0.84)	0.92	0.364	0.27 (−0.32, 0.85)
TG1_f	0.90 (0.09)	0.85 (0.13)	1.26	0.218	0.40 (−0.19, 0.99)
TG1_n	0.52 (0.08)	0.45 (0.09)	2.75	0.010	0.82 (0.22, 1.43)
TG1_v	0.63 (0.08)	0.53 (0.10)	3.49	0.001	1.08 (0.46, 1.70)
TG2_f	0.35 (0.18)	0.32 (0.11)	0.80	0.426	0.21 (−0.38, 0.79)
TG2_n	0.35 (0.15)	0.31 (0.15)	1.01	0.317	0.29 (−0.29, 0.88)
TG2_v	0.27 (0.13)	0.28 (0.13)	−0.15	0.879	−0.04 (−0.63, 0.54)
TG3_f	2.64 (0.44)	2.60 (0.59)	0.30	0.763	0.09 (−0.49, 0.68)
TG3_n	2.42 (0.59)	2.18 (0.65)	1.30	0.204	0.39 (−0.20, 0.97)
TG3_v	2.31 (0.53)	2.18 (0.53)	0.84	0.405	0.24 (−0.34, 0.83)

Table 17: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

As a further confirmation of the proper assignment of the two groups, substantial group differences in performance can be observed for all three subtests of the General Reasoning Ability Test.

While the two groups seem to differ in performance on the first task (Power_T1) of the Mathematics Power Test, they do not appear to differ in performance on the second task (Power_T2). The point estimates for Cohen's d associated with the third and fourth task (Power_T3 and Power_T4) both show a negligible effect, while effect sizes ranging from small negative effects to medium positive effects would also be in line with the data. Lastly, the results related to the fifth and sixth task (Power_T5 and Power_T6) both suggest small positive effects. Nonetheless, effect sizes ranging from small negative effects to large positive effects would be supported by the data as well. While substantial group differences in performance can be observed for the numerical as well as for the verbal subtest of the working memory function Storage-Processing (TG1_n and TG1_v), the point estimate for Cohen's d regarding the figural subtest (TG1_f) indicates a small positive effect, with effect sizes ranging from −0.19, a negligible effect, to 0.99, a large positive effect, being compatible with the data as well. With respect to the functional category of Supervision, the two groups do not seem to differ in performance on the verbal subtest (TG2_v). On the other hand, the point estimates for Cohen's d associated with the figural and numerical subtest (TG2_f and TG2_n) both specify a small positive effect, while effect sizes ranging from small negative effects to medium or large positive effects respectively would also be supported by the data. Finally, clear group differences in performance cannot be detected for the figural subtest of the working memory function Relational Integration (TG3_f), while the results for the numerical and verbal subtest (TG3_n and TG3_v) both suggest a small positive effect. However, effect sizes ranging from small negative effects to large positive effects would be in line with the data on the numerical and verbal subtest as well.

A.1.2 Mathematical Self-Concept and Interest in Mathematics

In addition to the comparisons collected in the previous subsection, the selected high-achievers and the over-achievers shown in Figure 38 have been compared with respect to their mathematical self-concept as well as their interest in mathematics. This subsection focuses on the presentation of the corresponding results.

Sum Scores and Mean Scores The histograms in Figure 42 visualise the mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the over-achievers respectively. To complement this graphical representation, Table 18 displays the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals. In addition, to facilitate the comparison between the various results, Table 18 contains the results related to the mathematics tests and the General Reasoning Ability Test already presented in Table 16. The structure of Table 18 is analogous to the one of Table 16.

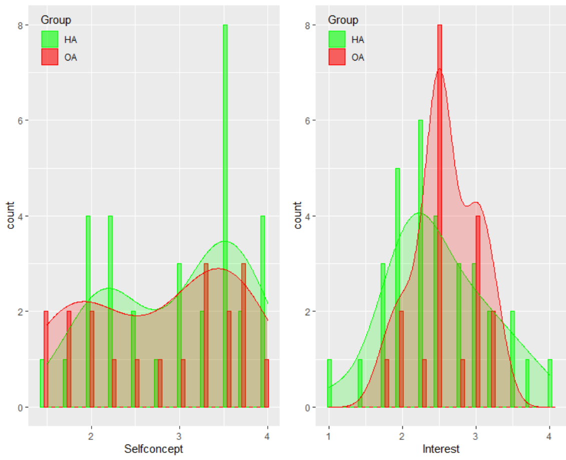


Figure 42: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the over-achievers (OA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.88 (2.88)	47.16 (2.32)	13.20	0.000	3.62 (2.69, 4.54)
Mathematics Posttest (Post)	41.19 (1.38)	41.32 (3.56)	−0.15	0.882	−0.05 (−0.64, 0.53)
Mathematics Power Test (Power)	7.46 (4.44)	6.13 (3.84)	1.13	0.266	0.31 (−0.27, 0.90)
Mathematics Speed Test (Speed)	31.56 (3.68)	30.95 (3.84)	0.56	0.578	0.16 (−0.42, 0.75)
Self-Concept Questionnaire Items (Selfconcept)	2.97 (0.76)	2.79 (0.85)	0.76	0.454	0.23 (−0.36, 0.81)
Interest Questionnaire Items (Interest)	2.49 (0.68)	2.59 (0.42)	−0.65	0.520	−0.17 (−0.75, 0.42)

Table 18: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 17), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 17), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 19) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 19) respectively.

It can be observed that the point estimate for Cohen's d associated with the mean scores of the Self-Concept Questionnaire Items indicates a small positive effect. Nonetheless, effect sizes ranging from −0.36, a small negative effect, to 0.81, a large positive effect, would also be supported by the data. On the other hand, the results related to the mean scores of the Interest Questionnaire Items suggest a negligible effect, while effect sizes ranging from −0.75, a medium negative effect, to 0.42, a small positive effect, would be compatible with the data as well.

Individual Questionnaire Items For a more detailed analysis of the group comparison between the selected high-achievers and the over-achievers with respect to the Self-Concept Questionnaire Items and the Interest Questionnaire Items, the individual Questionnaire Items are listed in Table 19. Selfconcept_QI1 - Selfconcept_QI4 relate to the Self-Concept Questionnaire Items, while Interest_QI5 - Interest_QI8 refer to the Interest Questionnaire Items. The corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are arranged in the same way as in Table 18. It is important to note that Selfconcept in Table 18 describes the mean scores of Selfconcept_QI1 - Selfconcept_QI4, while Interest in Table 18 represents the mean scores of Interest_QI5 - Interest_QI8.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
Selfconcept_QI1	3.06 (0.98)	2.53 (0.90)	1.98	0.054	0.56 (−0.03, 1.15)
Selfconcept_QI2	3.03 (0.69)	2.95 (1.03)	0.32	0.754	0.10 (−0.48, 0.68)
Selfconcept_QI3	2.53 (1.02)	2.42 (1.02)	0.37	0.710	0.11 (−0.47, 0.69)
Selfconcept_QI4	3.25 (0.84)	3.26 (0.93)	−0.05	0.960	−0.02 (−0.60, 0.57)
Interest_QI5	2.28 (0.99)	2.32 (0.89)	−0.13	0.898	−0.04 (−0.62, 0.55)
Interest_QI6	2.19 (0.82)	2.21 (0.63)	−0.11	0.911	−0.03 (−0.61, 0.55)
Interest_QI7	2.53 (0.98)	2.37 (0.68)	0.70	0.490	0.18 (−0.40, 0.77)
Interest_QI8	2.97 (0.78)	3.47 (0.61)	−2.56	0.014	−0.70 (−1.30, −0.10)

Table 19: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

Regarding the Self-Concept Questionnaire Items, there could be group differences with respect to the first item (Selfconcept_QI1), while there do not appear to be group differences on the fourth item (Selfconcept_QI4). The results associated with the second and third item (Selfconcept_QI2 and Selfconcept_QI3) both indicate a negligible effect, while effect sizes ranging from small negative effects to medium positive effects would also be supported by the data. In terms of the Interest Questionnaire Items, the two groups do not seem to differ on the fifth and sixth item (Interest_QI5 and

Interest_QI6), while they clearly differ on the eighth item (Interest_QI8). Lastly, the point estimate for Cohen's d related to the seventh item (Interest_QI7) suggests a negligible effect. Nonetheless, effect sizes ranging from -0.40, a small negative effect, to 0.77, a medium positive effect, would be compatible with the data as well.

A.1.3 Summary of the Results

This subsection aims at summarising the results from Subsections A.1.1 and A.1.2, which reported on various group comparisons between high-achievers and over-achievers. More specifically, these results are collected in Table 20, where they are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT	56.88 (2.88)	47.16 (2.32)	13.20	0.000	3.62 (2.69, 4.54)
KFT_N	22.47 (1.81)	18.16 (2.48)	6.61	0.000	2.07 (1.36, 2.78)
KFT_V	16.31 (2.26)	13.11 (3.28)	3.76	0.001	1.20 (0.57, 1.82)
TG1	0.68 (0.05)	0.61 (0.08)	3.52	0.001	1.12 (0.50, 1.75)
TG1_v	0.63 (0.08)	0.53 (0.10)	3.49	0.001	1.08 (0.46, 1.70)
KFT_Q	18.09 (1.86)	15.89 (2.90)	2.96	0.006	0.96 (0.34, 1.57)
TG1_n	0.52 (0.08)	0.45 (0.09)	2.75	0.010	0.82 (0.22, 1.43)
Power_T1	1.63 (1.08)	1.05 (0.85)	2.10	0.041	0.57 (−0.02, 1.16)
Selfconcept_QI1	3.06 (0.98)	2.53 (0.90)	1.98	0.054	0.56 (−0.03, 1.15)
TG1_f	0.90 (0.09)	0.85 (0.13)	1.26	0.218	0.40 (−0.19, 0.99)
TG3	2.46 (0.34)	2.32 (0.35)	1.36	0.182	0.40 (−0.19, 0.99)
TG3_n	2.42 (0.59)	2.18 (0.65)	1.30	0.204	0.39 (−0.20, 0.97)
Power	7.46 (4.44)	6.13 (3.84)	1.13	0.266	0.31 (−0.27, 0.90)
TG2_n	0.35 (0.15)	0.31 (0.15)	1.01	0.317	0.29 (−0.29, 0.88)
Power_T6	0.83 (0.83)	0.61 (0.84)	0.92	0.364	0.27 (−0.32, 0.85)
TG3_v	2.31 (0.53)	2.18 (0.53)	0.84	0.405	0.24 (−0.34, 0.83)
Power_T5	0.64 (0.90)	0.43 (0.76)	0.87	0.388	0.24 (−0.34, 0.83)
TG2	0.33 (0.11)	0.30 (0.09)	0.85	0.402	0.23 (−0.35, 0.81)
Selfconcept	2.97 (0.76)	2.79 (0.85)	0.76	0.454	0.23 (−0.36, 0.81)
TG2_f	0.35 (0.18)	0.32 (0.11)	0.80	0.426	0.21 (−0.38, 0.79)
Interest_QI7	2.53 (0.98)	2.37 (0.68)	0.70	0.490	0.18 (−0.40, 0.77)
Power_T4	1.35 (0.84)	1.20 (0.85)	0.63	0.533	0.18 (−0.40, 0.77)
Speed	31.56 (3.68)	30.95 (3.84)	0.56	0.578	0.16 (−0.42, 0.75)
Power_T3	1.14 (0.83)	1.04 (0.74)	0.45	0.655	0.13 (−0.46, 0.71)
Selfconcept_QI3	2.53 (1.02)	2.42 (1.02)	0.37	0.710	0.11 (−0.47, 0.69)
Selfconcept_QI2	3.03 (0.69)	2.95 (1.03)	0.32	0.754	0.10 (−0.48, 0.68)
TG3_f	2.64 (0.44)	2.60 (0.59)	0.30	0.763	0.09 (−0.49, 0.68)
Power_T2	1.88 (1.09)	1.80 (0.96)	0.25	0.806	0.07 (−0.51, 0.65)
Selfconcept_QI4	3.25 (0.84)	3.26 (0.93)	−0.05	0.960	−0.02 (−0.60, 0.57)
Interest_QI6	2.19 (0.82)	2.21 (0.63)	−0.11	0.911	−0.03 (−0.61, 0.55)
Interest_QI5	2.28 (0.99)	2.32 (0.89)	−0.13	0.898	−0.04 (−0.62, 0.55)
TG2_v	0.27 (0.13)	0.28 (0.13)	−0.15	0.879	−0.04 (−0.63, 0.54)
Post	41.19 (1.38)	41.32 (3.56)	−0.15	0.882	−0.05 (−0.64, 0.53)
Interest	2.49 (0.68)	2.59 (0.42)	−0.65	0.520	−0.17 (−0.75, 0.42)
Interest_QI8	2.97 (0.78)	3.47 (0.61)	−2.56	0.014	−0.70 (−1.30, −0.10)

Table 20: Summary of all group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Substantial group differences in performance appear on the General Reasoning Ability Test (KFT) and on its subscales (KFT_N, KFT_V, KFT_Q) respectively, which suggests that the two groups were properly assigned. Apart from these group differences in performance, the largest effect sizes in terms of Cohen's d can be observed for the working memory tests related to the working memory function of Storage-Processing (TG1 and TG1_v, TG1_n respectively). For the Mathematics Power Test, the absolute values of the point estimates for Cohen's d indicate the largest effect on the first task (Power_T1), while the ones associated with the questionnaire items show the largest effects on the first and eighth item (Selfconcept_QI1 and Interest_QI8). As already noted in Subsection A.1.1, the point estimate for Cohen's d related to the Mathematics Power Test (Power) is larger than the one associated with the Mathematics Speed Test (Speed).

A.2 Additional Group Comparison between High-Achievers and Over-Achievers II

As a second additional comparison between high-achievers and over-achievers, a subgroup of 24 high-achievers (18 female students (75%) and 6 male students (25%)) has been compared against the group of 19 over-achievers (10 female students (53%) and 9 male students (47%)). The selection of the 24 high-achievers was based on cut-off points on both scales, the Mathematics Posttest scale and the General Reasoning Ability scale. Similarly as before, Figure 43 depicts the selected high-achievers (marked with a green dot), as well as the over-achievers (marked with a red dot).

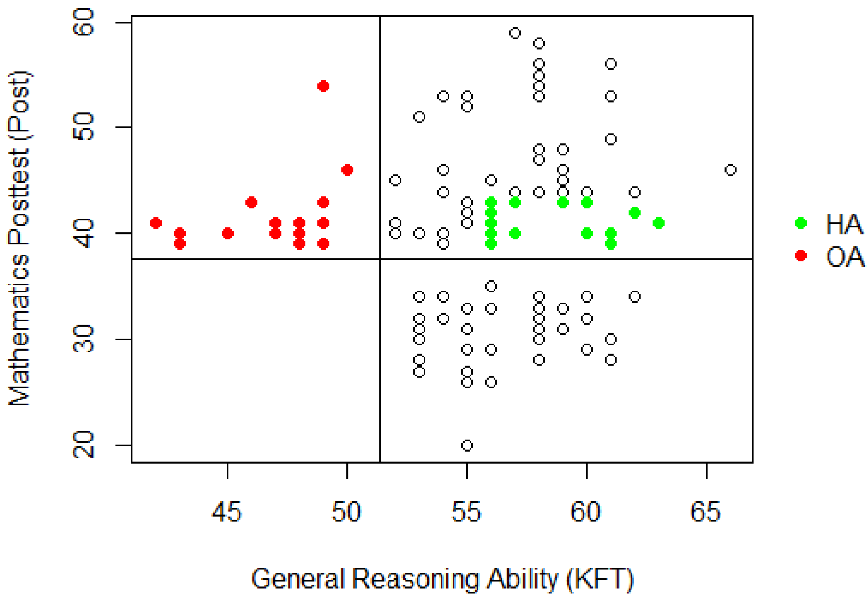


Figure 43: Selected high-achievers (marked with a green dot) and over-achievers (marked with a red dot).

A.2.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

The goal of this subsection is to present a comparison between the selected high-achievers and the over-achievers shown in Figure 43 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms of the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the various tests are collected in Figures 44, 45 and 46. The histograms belonging to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post) are given in Figure 44, while the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed) are shown in Figure 45. Lastly, Figure 46 arranges the histograms related to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3). In an analogous way as in the previous section, Table 21 complements these graphical representations. The second and third column of Table 21 contain the descriptive statistics of the performances of the selected high-achievers (HA) and of the over-achievers (OA) respectively. The fourth and fifth column of Table 21 present the t-values and p-values of a two sample t-test comparing the two groups. Finally, the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are shown in the last column of Table 21.

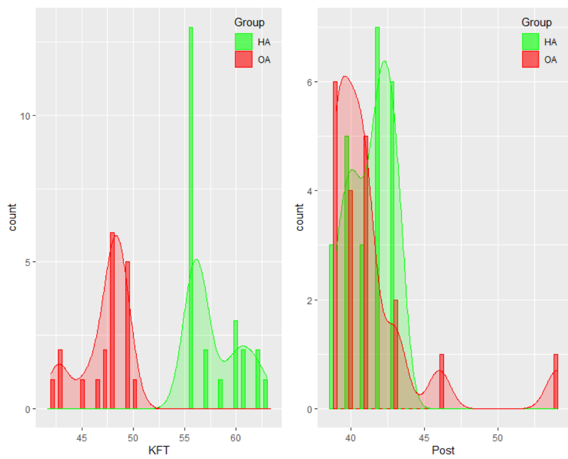


Figure 44: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

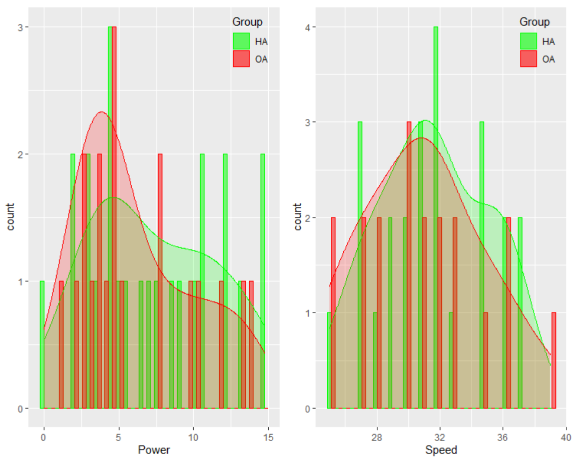


Figure 45: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

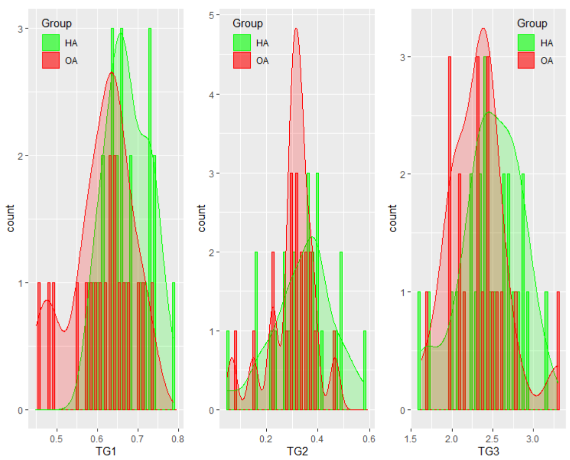


Figure 46: Histograms visualising the performances of the selected high-achievers (HA) and of the over-achievers (OA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	57.92 (2.48)	47.16 (2.32)	14.65	0.000	4.46 (3.31, 5.61)
Mathematics Posttest (Post)	41.33 (1.40)	41.32 (3.56)	0.02	0.984	0.01 (−0.61, 0.63)
Mathematics Power Test (Power)	7.31 (4.23)	6.13 (3.84)	0.96	0.344	0.29 (−0.33, 0.91)
Mathematics Speed Test (Speed)	31.54 (3.45)	30.95 (3.84)	0.53	0.601	0.16 (−0.46, 0.79)
WM Storage-Processing (TG1)	0.68 (0.05)	0.61 (0.08)	3.15	0.004	1.01 (0.35, 1.66)
WM Supervision (TG2)	0.34 (0.12)	0.30 (0.09)	1.25	0.219	0.37 (−0.26, 1.00)
WM Relational Integration (TG3)	2.48 (0.38)	2.32 (0.35)	1.41	0.167	0.43 (−0.20, 1.06)

Table 21: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 22), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 22), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 22).

The proper assignment of the two groups can be confirmed, as the results indicate substantial group differences in performance on the General Reasoning Ability Test, while there do not seem to be any substantial group differences regarding the Mathematics Posttest.

The point estimate for Cohen's d related to the Mathematics Power Test is larger than the one associated with the Mathematics Speed Test. For the working memory functions, it can be observed that the two groups differ in their scores on the test related to the functional category of Storage-Processing. On the other hand, the point estimates for Cohen's d with respect to the other two working memory functions both suggest a small positive effect, while effect sizes ranging from small negative effects to large positive effects would also be reasonably compatible with the data.

Individual Subtests or Tasks In order to analyse the group comparison between the selected high-achievers and the over-achievers with respect to the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests in more detail, Table 22 collects the respective subtests or tasks. More specifically, it presents the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subscales of the General Reasoning Ability Test as well as the scores on the different tasks in the Mathematics Power Test (Power_T1 - Power_T6). Moreover, for each functional category of working memory (TG1 - TG3) it displays the corresponding figural (f), numerical (n) and verbal (v) subtests respectively. In Table 22, the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are arranged in the same way as in Table 21. It is important to note that KFT and Power in Table 21 relate to the sum scores of the respective subtests or tasks listed in Table 22, while TG1 - TG3 in Table 21 describe the mean scores of the corresponding subtests shown in Table 22.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT_N	22.83 (1.52)	18.16 (2.48)	7.22	0.000	2.34 (1.54, 3.14)
KFT_Q	18.29 (1.40)	15.89 (2.90)	3.31	0.003	1.09 (0.43, 1.76)
KFT_V	16.79 (2.28)	13.11 (3.28)	4.16	0.000	1.33 (0.65, 2.02)
Power_T1	1.59 (1.08)	1.05 (0.85)	1.84	0.073	0.55 (−0.08, 1.18)
Power_T2	1.89 (1.05)	1.80 (0.96)	0.27	0.789	0.08 (−0.54, 0.70)
Power_T3	1.09 (0.87)	1.04 (0.74)	0.22	0.826	0.07 (−0.55, 0.69)
Power_T4	1.39 (0.74)	1.20 (0.85)	0.76	0.451	0.24 (−0.38, 0.86)
Power_T5	0.58 (0.86)	0.43 (0.76)	0.60	0.551	0.18 (−0.44, 0.80)
Power_T6	0.77 (0.85)	0.61 (0.84)	0.64	0.527	0.20 (−0.43, 0.82)
TG1_f	0.90 (0.09)	0.85 (0.13)	1.23	0.229	0.39 (−0.23, 1.02)
TG1_n	0.50 (0.07)	0.45 (0.09)	2.01	0.053	0.63 (0.00, 1.27)
TG1_v	0.63 (0.08)	0.53 (0.10)	3.49	0.001	1.10 (0.44, 1.77)
TG2_f	0.36 (0.20)	0.32 (0.11)	0.93	0.359	0.27 (−0.35, 0.89)
TG2_n	0.38 (0.13)	0.31 (0.15)	1.74	0.090	0.55 (−0.09, 1.18)
TG2_v	0.28 (0.14)	0.28 (0.13)	−0.03	0.980	−0.01 (−0.63, 0.61)
TG3_f	2.67 (0.48)	2.60 (0.59)	0.45	0.657	0.14 (−0.48, 0.76)
TG3_n	2.45 (0.61)	2.18 (0.65)	1.36	0.183	0.42 (−0.21, 1.05)
TG3_v	2.31 (0.54)	2.18 (0.53)	0.82	0.419	0.25 (−0.37, 0.87)

Table 22: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

As a first observation, it is important to note that the two groups clearly differ in performance on all three subtests of the General Reasoning Ability Test, providing further evidence for the proper assignment of the two groups.

While there could be group differences in performance on the first task (Power_T1) of the Mathematics Power Test, there do not seem to be any group differences in performance on the second and third task (Power_T2 and Power_T3). Moreover, the results associated with the fourth and sixth task (Power_T4 and Power_T6) both specify a small positive effect, with effect sizes ranging from small negative effects to large positive effects being supported by the data as well. Finally, the point estimate for Cohen's d related to the fifth task (Power_T5) indicates a negligible effect. However, effect sizes ranging from −0.44, a small negative effect, to 0.80, a medium positive effect, would also be in line with the data. With respect to the working memory subtests, the results suggest that the two groups differ in performance on the numerical and verbal subtest of the working memory function Storage-Processing (TG1_n and TG1_v). For the figural subtest (TG1_f), the point estimate for Cohen's d shows a small positive effect, while effect sizes ranging from −0.23, a small negative effect, to 1.02, a large positive effect, would be compatible with the data as well. The results regarding the figural subtest of the functional category of Supervision (TG2_f) suggest a small positive effect. Nonetheless, effect sizes ranging from −0.35, a small negative effect, to 0.89, a large positive effect, would also be supported by the data. Moreover, the point estimate for Cohen's d associated with the numerical subtest (TG2_n) specifies a medium positive effect, with effect sizes ranging from −0.09, a negligible effect, to 1.18, a large positive effect, being in line with the data as well, while the two groups do not appear to differ in performance on the verbal subtest (TG2_v). Lastly, the results related to the figural subtest of the working memory function Relational Integration (TG3_f) show a negligible effect. However, effect sizes ranging from −0.48, a small negative effect, to 0.76, a medium positive effect, would also be supported by the data. For the numerical and verbal subtest (TG3_n and TG3_v), the point estimates for Cohen's d both indicate a small positive effect, while effect sizes ranging from small negative effects to large positive effects would be compatible with the data as well.

A.2.2 Mathematical Self-Concept and Interest in Mathematics

Additionally to the comparisons presented in the previous subsection, the selected high-achievers and the over-achievers shown in Figure 43 have been compared with respect to their mathematical self-concept and their interest in mathematics as well. The corresponding results are discussed in this subsection.

Sum Scores and Mean Scores Figure 47 shows the histograms related to the mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the over-achievers respectively. In addition to this graphical representation, the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are given in Table 23. Moreover, to enable an overview on the various results, Table 23 also lists the results associated with the mathematics tests and the General Reasoning Ability Test already displayed in Table 21. Table 23 is structured in the same way as Table 21.

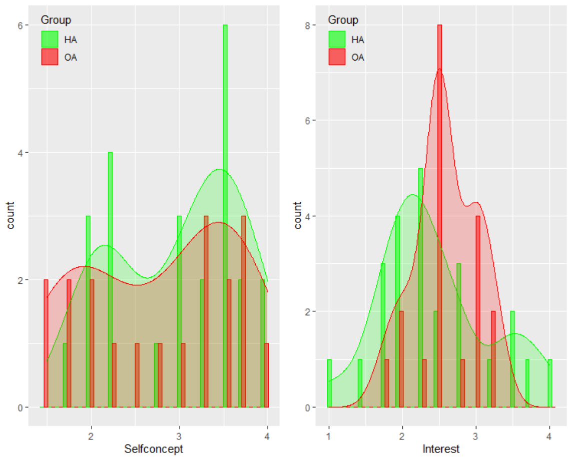


Figure 47: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the over-achievers (OA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
General Reasoning Ability (KFT)	57.92 (2.48)	47.16 (2.32)	14.65	0.000	4.46 (3.31, 5.61)
Mathematics Posttest (Post)	41.33 (1.40)	41.32 (3.56)	0.02	0.984	0.01 (−0.61, 0.63)
Mathematics Power Test (Power)	7.31 (4.23)	6.13 (3.84)	0.96	0.344	0.29 (−0.33, 0.91)
Mathematics Speed Test (Speed)	31.54 (3.45)	30.95 (3.84)	0.53	0.601	0.16 (−0.46, 0.79)
Self-Concept Questionnaire Items (Selfconcept)	2.98 (0.71)	2.79 (0.85)	0.78	0.441	0.24 (−0.38, 0.87)
Interest Questionnaire Items (Interest)	2.43 (0.74)	2.59 (0.42)	−0.92	0.364	−0.27 (−0.89, 0.36)

Table 23: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 22), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 22), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 24) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 24) respectively.

The point estimate for Cohen's d related to the mean scores of the Self-Concept Questionnaire Items specifies a small positive effect, while effect sizes ranging from −0.38, a small negative effect, to 0.87, a large positive effect, would be in line with the data as well. In addition, the results associated with the mean scores of the Interest Questionnaire Items show a small negative effect. However, effect sizes ranging from −0.89, a large negative effect, to 0.36, a small positive effect, would also be supported by the data.

Individual Questionnaire Items In order to analyse the group comparison between the selected high-achievers and the over-achievers with respect to the Self-Concept Questionnaire Items and the Interest Questionnaire Items in more detail, Table 24 collects the individual Questionnaire Items. While Selfconcept_QI1 - Selfconcept_QI4 correspond to the Self-Concept Questionnaire Items, Interest_QI5 - Interest_QI8 represent the Interest Questionnaire Items. Table 24 structures the related descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in the same way as Table 23. While Selfconcept in Table 23 refers to the mean scores of Selfconcept_QI1 - Selfconcept_QI4, Interest in Table 23 describes the mean scores of Interest_QI5 - Interest_QI8.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
Selfconcept_QI1	3.25 (0.74)	2.53 (0.90)	2.82	0.008	0.89 (0.24, 1.54)
Selfconcept_QI2	2.96 (0.69)	2.95 (1.03)	0.04	0.968	0.01 (−0.61, 0.63)
Selfconcept_QI3	2.50 (0.98)	2.42 (1.02)	0.26	0.799	0.08 (−0.54, 0.70)
Selfconcept_QI4	3.21 (0.88)	3.26 (0.93)	−0.20	0.846	−0.06 (−0.68, 0.56)
Interest_QI5	2.13 (1.08)	2.32 (0.89)	−0.64	0.527	−0.19 (−0.81, 0.43)
Interest_QI6	2.21 (0.83)	2.21 (0.63)	−0.01	0.992	0.00 (−0.62, 0.62)
Interest_QI7	2.46 (1.06)	2.37 (0.68)	0.34	0.739	0.10 (−0.52, 0.72)
Interest_QI8	2.92 (0.83)	3.47 (0.61)	−2.53	0.015	−0.75 (−1.39, −0.11)

Table 24: Group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

While the two groups seem to clearly differ on the first item (Selfconcept_QI1), they do not appear to differ on the remaining Self-Concept Questionnaire Items (Selfconcept_QI2 - Selfconcept_QI4). On the other hand, regarding the Interest Questionnaire Items, the point estimate for Cohen's d associated with the fifth item (Interest_QI5) specifies a negligible effect. However, effect sizes ranging from −0.81, a large negative effect, to 0.43, a small positive effect, would be in line with the data as well. Moreover, there do not seem to be any group differences on the sixth item (Interest_QI6), while substantial group differences can be observed with respect to the eighth item (Interest_QI8). Finally, the results related to the seventh item (Interest_QI7) show a negligible effect, with effect sizes ranging from medium negative effects to medium positive effects being also compatible with the data.

A.2.3 Summary of the Results

This subsection merges the results from Subsection A.2.1 and Subsection A.2.2, which focused on different group comparisons between high-achieving and over-achieving students. More concretely, Table 25 arranges these results in descending order of the respective effect sizes in terms of Cohen's d.

Test	M_{HA} (SD_{HA})	M_{OA} (SD_{OA})	t	p	d (CI)
KFT	57.92 (2.48)	47.16 (2.32)	14.65	0.000	4.46 (3.31, 5.61)
KFT_N	22.83 (1.52)	18.16 (2.48)	7.22	0.000	2.34 (1.54, 3.14)
KFT_V	16.79 (2.28)	13.11 (3.28)	4.16	0.000	1.33 (0.65, 2.02)
TG1_v	0.63 (0.08)	0.53 (0.10)	3.49	0.001	1.10 (0.44, 1.77)
KFT_Q	18.29 (1.40)	15.89 (2.90)	3.31	0.003	1.09 (0.43, 1.76)
TG1	0.68 (0.05)	0.61 (0.08)	3.15	0.004	1.01 (0.35, 1.66)
Selfconcept_QI1	3.25 (0.74)	2.53 (0.90)	2.82	0.008	0.89 (0.24, 1.54)
TG1_n	0.50 (0.07)	0.45 (0.09)	2.01	0.053	0.63 (0.00, 1.27)
Power_T1	1.59 (1.08)	1.05 (0.85)	1.84	0.073	0.55 (−0.08, 1.18)
TG2_n	0.38 (0.13)	0.31 (0.15)	1.74	0.090	0.55 (−0.09, 1.18)
TG3	2.48 (0.38)	2.32 (0.35)	1.41	0.167	0.43 (−0.20, 1.06)
TG3_n	2.45 (0.61)	2.18 (0.65)	1.36	0.183	0.42 (−0.21, 1.05)
TG1_f	0.90 (0.09)	0.85 (0.13)	1.23	0.229	0.39 (−0.23, 1.02)
TG2	0.34 (0.12)	0.30 (0.09)	1.25	0.219	0.37 (−0.26, 1.00)
Power	7.31 (4.23)	6.13 (3.84)	0.96	0.344	0.29 (−0.33, 0.91)
TG2_f	0.36 (0.20)	0.32 (0.11)	0.93	0.359	0.27 (−0.35, 0.89)
TG3_v	2.31 (0.54)	2.18 (0.53)	0.82	0.419	0.25 (−0.37, 0.87)
Selfconcept	2.98 (0.71)	2.79 (0.85)	0.78	0.441	0.24 (−0.38, 0.87)
Power_T4	1.39 (0.74)	1.20 (0.85)	0.76	0.451	0.24 (−0.38, 0.86)
Power_T6	0.77 (0.85)	0.61 (0.84)	0.64	0.527	0.20 (−0.43, 0.82)
Power_T5	0.58 (0.86)	0.43 (0.76)	0.60	0.551	0.18 (−0.44, 0.80)
Speed	31.54 (3.45)	30.95 (3.84)	0.53	0.601	0.16 (−0.46, 0.79)
TG3_f	2.67 (0.48)	2.60 (0.59)	0.45	0.657	0.14 (−0.48, 0.76)
Interest_QI7	2.46 (1.06)	2.37 (0.68)	0.34	0.739	0.10 (−0.52, 0.72)
Power_T2	1.89 (1.05)	1.80 (0.96)	0.27	0.789	0.08 (−0.54, 0.70)
Selfconcept_QI3	2.50 (0.98)	2.42 (1.02)	0.26	0.799	0.08 (−0.54, 0.70)
Power_T3	1.09 (0.87)	1.04 (0.74)	0.22	0.826	0.07 (−0.55, 0.69)
Selfconcept_QI2	2.96 (0.69)	2.95 (1.03)	0.04	0.968	0.01 (−0.61, 0.63)
Post	41.33 (1.40)	41.32 (3.56)	0.02	0.984	0.01 (−0.61, 0.63)
Interest_QI6	2.21 (0.83)	2.21 (0.63)	−0.01	0.992	0.00 (−0.62, 0.62)
TG2_v	0.28 (0.14)	0.28 (0.13)	−0.03	0.980	−0.01 (−0.63, 0.61)
Selfconcept_QI4	3.21 (0.88)	3.26 (0.93)	−0.20	0.846	−0.06 (−0.68, 0.56)
Interest_QI5	2.13 (1.08)	2.32 (0.89)	−0.64	0.527	−0.19 (−0.81, 0.43)
Interest	2.49 (0.68)	2.59 (0.42)	−0.65	0.520	−0.17 (−0.75, 0.42)
Interest_QI8	2.92 (0.83)	3.47 (0.61)	−2.53	0.015	−0.75 (−1.39, −0.11)

Table 25: Summary of all group comparisons between the selected high-achievers (HA) and the over-achievers (OA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Clear group differences in performance appear for the General Reasoning Ability Test (KFT) and its subscales (KFT_N, KFT_V, KFT_Q) respectively, which reflects the proper assignment of the two groups. Apart from these, the largest effects in terms of Cohen's d can be observed for the working memory tests related to the working memory function of Storage-Processing (TG1 and TG1_v, TG1_n respectively) as well as for the first and eighth item of the questionnaire (Selfconcept_QI1 and Interest_QI8). For the Mathematics Power Test, the point estimates for Cohen's d specify the largest effect on the first task (Power_T1), while the ones related to the subtests of the functional category of Supervision indicate the largest effect on the numerical subtest (TG2_n). As remarked in Subsection A.2.1, the point estimate for Cohen's d associated with the Mathematics Power Test (Power) is larger than the one related to the Mathematics Speed Test (Speed).

A.3 Additional Group Comparison between High-Achievers and Under-Achievers I

As a first additional comparison between high-achievers and under-achievers, a subgroup of 57 high-achievers (37 female students (65%) and 20 male students (35%)) has been compared against the group of 34 under-achievers (17 female students (50%) and 17 male students (50%)). The selection of high-achievers was based on a cut-off point on the General Reasoning Ability scale. Figure 48 depicts the selected high-achievers (marked with a green dot), as well as the under-achievers (marked with an orange dot).

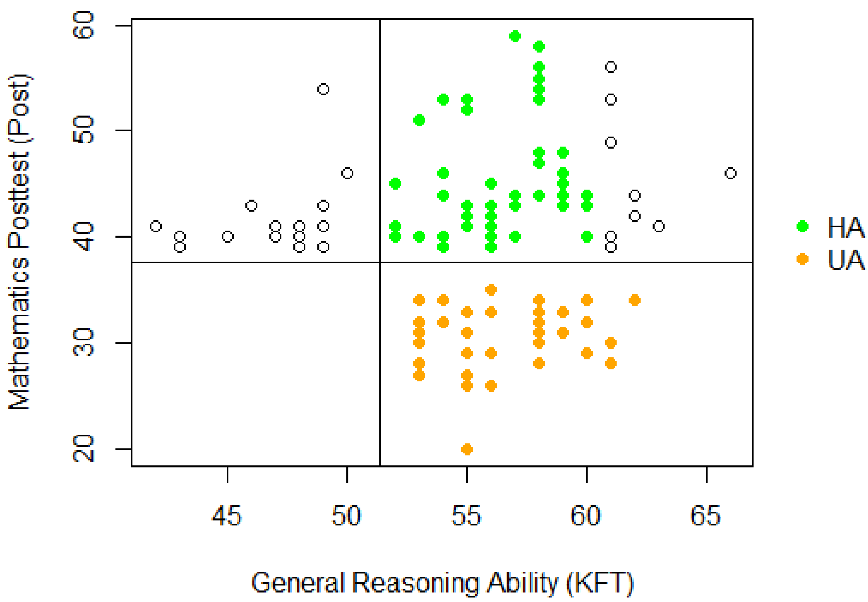


Figure 48: Selected high-achievers (marked with a green dot) and under-achievers (marked with an orange dot).

A.3.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

The aim of this subsection is to discuss a comparison between the selected high-achievers and the under-achievers depicted in Figure 48 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms of the performances of the selected high-achievers (HA) and the under-achievers (UA) in the different tests are given in Figures 49, 50 and 51. While Figure 49 shows the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), Figure 50 depicts the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed). Lastly, the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) are arranged in Figure 51. To complement all these histograms, Table 26 lists the descriptive statistics of the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the second and third column respectively. The fourth and fifth column of Table 26 contain the t-values and p-values of a two sample t-test comparing the two groups. Finally, the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are displayed in the last column of Table 26.

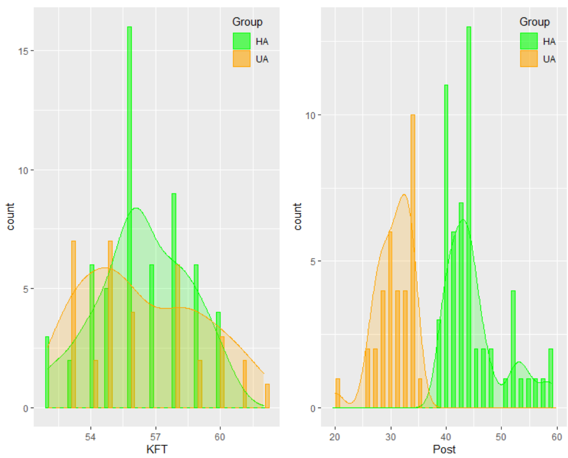


Figure 49: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

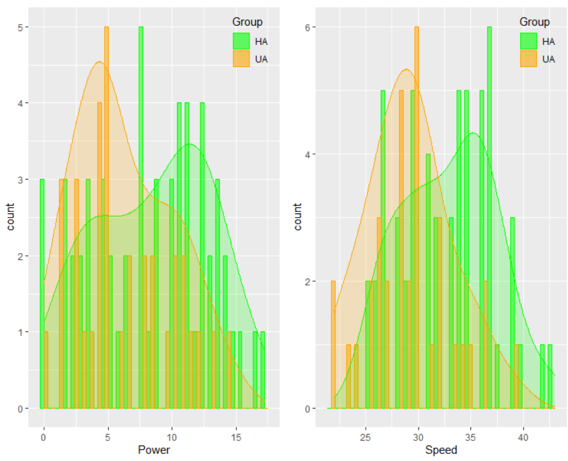


Figure 50: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

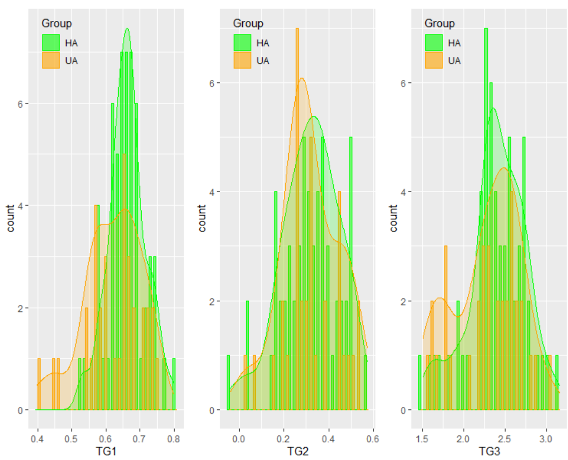


Figure 51: Histograms visualising the performances of the selected high-achievers (HA) and of the under-achievers (UA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.40 (2.09)	56.41 (2.74)	−0.02	0.988	0.00 (−0.43, 0.43)
Mathematics Posttest (Post)	45.21 (5.35)	30.50 (3.16)	16.49	0.000	3.16 (2.52, 3.79)
Mathematics Power Test (Power)	8.52 (4.60)	6.17 (3.80)	2.64	0.010	0.54 (0.11, 0.98)
Mathematics Speed Test (Speed)	32.89 (4.44)	29.15 (4.05)	4.11	0.000	0.87 (0.42, 1.32)
WM Storage-Processing (TG1)	0.66 (0.05)	0.62 (0.08)	2.66	0.011	0.64 (0.20, 1.08)
WM Supervision (TG2)	0.32 (0.13)	0.31 (0.13)	0.18	0.858	0.04 (−0.39, 0.47)
WM Relational Integration (TG3)	2.40 (0.34)	2.28 (0.40)	1.43	0.157	0.32 (−0.11, 0.75)

Table 26: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 27), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 27), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 27).

The results indicate that the two groups have been properly assigned, as they do not seem to differ in performance on the General Reasoning Ability Test, while they clearly differ in performance on the Mathematics Posttest. In addition, group differences in performance on the Mathematics Power Test as well as on the Mathematics Speed Test can also be observed.

Regarding working memory functions, the results show that the two groups differ in performance on the working memory test measuring the functional category of Storage-Processing. Moreover, the point estimate for Cohen's d with respect to the working memory function of Supervision specifies a negligible effect. Nonetheless, effect sizes ranging from −0.39, a small negative effect, to 0.47, a small positive effect, would also be in line with the data. Lastly, the results related to the functional category of Relational Integration indicate a small positive effect, while effect sizes ranging from −0.11, a negligible effect, to 0.75, a medium positive effect, would be supported by the data as well.

Individual Subtests or Tasks For a more detailed analysis of the group comparison between the selected high-achievers and the under-achievers with respect to the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests, the related subtests or tasks are displayed in Table 27. More specifically, it lists the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subscales of the General Reasoning Ability Test, as well as the scores in Task 1 - Task 6 in the Mathematics Power

Test (Power_T1 - Power_T6). Finally, Table 27 arranges the figural (f), numerical (n) and verbal (v) subtests for each functional category of working memory (TG1 - TG3). The structure of Table 27 regarding the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals is analogous to the one of Table 26. It is important to note that KFT and Power in Table 26 represent the sum scores of the corresponding subtests or tasks shown in Table 27, while TG1 - TG3 in Table 26 refer to the mean scores of the associated subtests given in Table 27.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
KFT_N	22.84 (1.68)	22.21 (2.11)	1.50	0.140	0.34 (−0.09, 0.78)
KFT_Q	17.89 (2.00)	17.85 (2.08)	0.09	0.925	0.02 (−0.41, 0.45)
KFT_V	15.67 (1.94)	16.35 (1.76)	−1.73	0.087	−0.37 (−0.80, 0.07)
Power_T1	1.71 (1.07)	1.37 (0.89)	1.65	0.102	0.34 (−0.09, 0.78)
Power_T2	1.96 (0.96)	1.71 (1.00)	1.18	0.243	0.26 (−0.17, 0.69)
Power_T3	1.23 (0.78)	0.82 (0.64)	2.75	0.007	0.57 (0.13, 1.01)
Power_T4	1.56 (0.88)	1.21 (0.99)	1.71	0.093	0.38 (−0.05, 0.82)
Power_T5	0.85 (0.98)	0.49 (0.82)	1.87	0.066	0.39 (−0.05, 0.82)
Power_T6	1.21 (1.07)	0.57 (0.76)	3.28	0.001	0.65 (0.21, 1.09)
TG1_f	0.90 (0.09)	0.88 (0.12)	0.70	0.485	0.17 (−0.27, 0.60)
TG1_n	0.50 (0.08)	0.42 (0.11)	3.51	0.001	0.82 (0.38, 1.27)
TG1_v	0.59 (0.09)	0.55 (0.09)	1.92	0.059	0.42 (−0.01, 0.86)
TG2_f	0.33 (0.19)	0.33 (0.21)	0.14	0.887	0.03 (−0.40, 0.46)
TG2_n	0.35 (0.20)	0.36 (0.19)	−0.44	0.665	−0.09 (−0.52, 0.34)
TG2_v	0.28 (0.12)	0.25 (0.11)	1.07	0.289	0.23 (−0.21, 0.66)
TG3_f	2.55 (0.48)	2.41 (0.43)	1.45	0.151	0.31 (−0.13, 0.74)
TG3_n	2.39 (0.53)	2.22 (0.65)	1.26	0.211	0.29 (−0.14, 0.72)
TG3_v	2.26 (0.54)	2.22 (0.68)	0.31	0.758	0.07 (−0.36, 0.50)

Table 27: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

While the two groups do not appear to differ in performance on the quantitative subtest of the General Reasoning Ability Test (KFT_Q), the point estimates for Cohen's d associated with the non-verbal (KFT_N) and verbal (KFT_V) subtest indicate a small positive effect and a small negative effect respectively.

Substantial group differences in performance can be observed for the third and sixth task of the Mathematics Power Test (Power_T3 and Power_T6). The results for the remaining tasks (Power_T1, Power_T2, Power_T4 and Power_T5) all show a small positive effect. Nonetheless, effect sizes ranging from negligible effects to medium or large positive effects would be in line with the data as well. In addition, the two groups seem to differ in performance on the numerical and verbal subtest of the functional category of Storage-Processing (TG1_n and TG1_v). For the figural subtest (TG1_f), the point estimate for Cohen's d specifies a negligible effect, while effect sizes ranging from -0.27, a small negative effect, to 0.60, a medium positive effect, would also be supported by the data. With respect to the working memory function Supervision, the two groups do not seem to differ in performance on the figural and numerical subtest (TG2_f and TG2_n). Moreover, the results related to the verbal subtest (TG2_v) suggest a small positive effect. However, effect sizes ranging from -0.21, a small negative effect, to 0.66, a medium positive effect, would be compatible with the data as well. Finally, for the figural and numerical subtest of the functional category of Relational Integration (TG3_f and TG3_n), the point estimates for Cohen's d both indicate a small positive effect, while effect sizes ranging from negligible effects to medium positive effects would be compatible with the data as well. On the other hand, substantial group differences in performance cannot be observed for the verbal subtest (TG3_v).

A.3.2 Mathematical Self-Concept and Interest in Mathematics

As an additional comparison between the selected high-achievers and the under-achievers depicted in Figure 48, the two groups have been compared with respect to their mathematical self-concept as well as their interest in mathematics. The focus of this subsection lies on the discussion of the corresponding results.

Sum Scores and Mean Scores Figure 52 depicts the histograms related to the mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the under-achievers respectively. To complement this graphical representation, the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are displayed in Table 28. In order to facilitate the comparison between the various results, Table 28 additionally contains the results associated with the mathematics tests and the General Reasoning Ability Test already shown in Table 26. The structure of Table 28 is analogous to the one of Table 26.

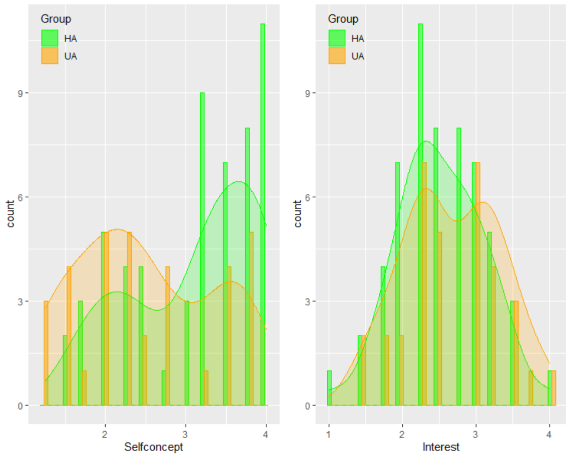


Figure 52: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the under-achievers (UA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.40 (2.09)	56.41 (2.74)	−0.02	0.988	0.00 (−0.43, 0.43)
Mathematics Posttest (Post)	45.21 (5.35)	30.50 (3.16)	16.49	0.000	3.16 (2.52, 3.79)
Mathematics Power Test (Power)	8.52 (4.60)	6.17 (3.80)	2.64	0.010	0.54 (0.11, 0.98)
Mathematics Speed Test (Speed)	32.89 (4.44)	29.15 (4.05)	4.11	0.000	0.87 (0.42, 1.32)
Self-Concept Questionnaire Items (Selfconcept)	3.10 (0.79)	2.49 (0.84)	3.41	0.001	0.75 (0.31, 1.20)
Interest Questionnaire Items (Interest)	2.52 (0.59)	2.68 (0.65)	−1.17	0.245	−0.26 (−0.69, 0.17)

Table 28: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 27), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 27), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 29) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 29) respectively.

As a first observation, it is important to note that the two groups differ with respect to their mean scores of the Self-Concept Questionnaire Items: The mathematical self-concept of the under-achievers seems to be considerably lower than the one of the selected high-achievers. In contrast, the results associated with the mean scores of the Interest Questionnaire Items indicate a small negative effect. Nonetheless, effect sizes ranging from −0.69, a medium negative effect, to 0.17, a negligible effect, would also be compatible with the data.

Individual Questionnaire Items For a more detailed evaluation of the group comparison between the selected high-achievers and the under-achievers, Table 29 shows the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and Interest Questionnaire Items (Interest_QI5 - Interest_QI8). The corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are structured in an analogous manner as in Table 28. While Selfconcept in Table 28 refers to the mean scores of Selfconcept_QI1 - Selfconcept_QI4, Interest in Table 28 relates to the mean scores of Interest_QI5 - Interest_QI8.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Selfconcept_QI1	3.07 (0.92)	2.65 (1.01)	1.99	0.050	0.44 (0.01, 0.88)
Selfconcept_QI2	3.16 (0.77)	2.44 (0.82)	4.11	0.000	0.90 (0.45, 1.35)
Selfconcept_QI3	2.84 (1.10)	1.97 (1.00)	3.88	0.000	0.82 (0.37, 1.27)
Selfconcept_QI4	3.33 (0.81)	2.91 (0.93)	2.19	0.032	0.49 (0.06, 0.93)
Interest_QI5	2.35 (0.86)	2.32 (1.12)	0.12	0.903	0.03 (−0.40, 0.46)
Interest_QI6	2.21 (0.80)	2.44 (0.79)	−1.35	0.182	−0.29 (−0.72, 0.14)
Interest_QI7	2.40 (0.82)	2.62 (0.82)	−1.21	0.231	−0.26 (−0.69, 0.17)
Interest_QI8	3.11 (0.72)	3.32 (0.64)	−1.50	0.138	−0.31 (−0.75, 0.12)

Table 29: Group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

As a first observation, it is important to note that there seem to be group differences with respect to all Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4). In contrast, the two groups do not appear to differ on the fifth item (Interest_QI5). The results for the remaining Interest Questionnaire Items (Interest_QI6 - Interest_QI8) all suggest a small negative effect. Nonetheless, effect sizes ranging from medium negative effects to negligible effects would be compatible with the data as well.

A.3.3 Summary of the Results

The results from Subsection A.3.1 and Subsection A.3.2 regarding the numerous group comparisons between high-achievers and under-achievers are collected in Table 30 in descending order of the respective effect sizes in terms of Cohen's d. This subsection aims at giving a summary of these results.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Post	45.21 (5.35)	30.50 (3.16)	16.49	0.000	3.16 (2.52, 3.79)
Selfconcept_QI2	3.16 (0.77)	2.44 (0.82)	4.11	0.000	0.90 (0.45, 1.35)
Speed	32.89 (4.44)	29.15 (4.05)	4.11	0.000	0.87 (0.42, 1.32)
TG1_n	0.50 (0.08)	0.42 (0.11)	3.51	0.001	0.82 (0.38, 1.27)
Selfconcept_QI3	2.84 (1.10)	1.97 (1.00)	3.88	0.000	0.82 (0.37, 1.27)
Selfconcept	3.10 (0.79)	2.49 (0.84)	3.41	0.001	0.75 (0.31, 1.20)
Power_T6	1.21 (1.07)	0.57 (0.76)	3.28	0.001	0.65 (0.21, 1.09)
TG1	0.66 (0.05)	0.62 (0.08)	2.66	0.011	0.64 (0.20, 1.08)
Power_T3	1.23 (0.78)	0.82 (0.64)	2.75	0.007	0.57 (0.13, 1.01)
Power	8.52 (4.60)	6.17 (3.80)	2.64	0.010	0.54 (0.11, 0.98)
Selfconcept_QI4	3.33 (0.81)	2.91 (0.93)	2.19	0.032	0.49 (0.06, 0.93)
Selfconcept_QI1	3.07 (0.92)	2.65 (1.01)	1.99	0.050	0.44 (0.01, 0.88)
TG1_v	0.59 (0.09)	0.55 (0.09)	1.92	0.059	0.42 (−0.01, 0.86)
Power_T5	0.85 (0.98)	0.49 (0.82)	1.87	0.066	0.39 (−0.05, 0.82)
Power_T4	1.56 (0.88)	1.21 (0.99)	1.71	0.093	0.38 (−0.05, 0.82)
KFT_N	22.84 (1.68)	22.21 (2.11)	1.50	0.140	0.34 (−0.09, 0.78)
Power_T1	1.71 (1.07)	1.37 (0.89)	1.65	0.102	0.34 (−0.09, 0.78)
TG3	2.40 (0.34)	2.28 (0.40)	1.43	0.157	0.32 (−0.11, 0.75)
TG3_f	2.55 (0.48)	2.41 (0.43)	1.45	0.151	0.31 (−0.13, 0.74)
TG3_n	2.39 (0.53)	2.22 (0.65)	1.26	0.211	0.29 (−0.14, 0.72)
Power_T2	1.96 (0.96)	1.71 (1.00)	1.18	0.243	0.26 (−0.17, 0.69)
TG2_v	0.28 (0.12)	0.25 (0.11)	1.07	0.289	0.23 (−0.21, 0.66)
TG1_f	0.90 (0.09)	0.88 (0.12)	0.70	0.485	0.17 (−0.27, 0.60)
TG3_v	2.26 (0.54)	2.22 (0.68)	0.31	0.758	0.07 (−0.36, 0.50)
TG2	0.32 (0.13)	0.31 (0.13)	0.18	0.858	0.04 (−0.39, 0.47)
TG2_f	0.33 (0.19)	0.33 (0.21)	0.14	0.887	0.03 (−0.40, 0.46)
Interest_QI5	2.35 (0.86)	2.32 (1.12)	0.12	0.903	0.03 (−0.40, 0.46)
KFT_Q	17.89 (2.00)	17.85 (2.08)	0.09	0.925	0.02 (−0.41, 0.45)
KFT	56.40 (2.09)	56.41 (2.74)	−0.02	0.988	0.00 (−0.43, 0.43)
TG2_n	0.35 (0.20)	0.36 (0.19)	−0.44	0.665	−0.09 (−0.52, 0.34)
Interest	2.52 (0.59)	2.68 (0.65)	−1.17	0.245	−0.26 (−0.69, 0.17)
Interest_QI7	2.40 (0.82)	2.62 (0.82)	−1.21	0.231	−0.26 (−0.69, 0.17)
Interest_QI6	2.21 (0.80)	2.44 (0.79)	−1.35	0.182	−0.29 (−0.72, 0.14)
Interest_QI8	3.11 (0.72)	3.32 (0.64)	−1.50	0.138	−0.31 (−0.75, 0.12)
KFT_V	15.67 (1.94)	16.35 (1.76)	−1.73	0.087	−0.37 (−0.80, 0.07)

Table 30: Summary of all group comparisons between the selected high-achievers (HA) and the under-achievers (UA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Clear group differences in performance appear for the Mathematics Posttest (Post), which is in line with the proper assignment of the two groups. For the questionnaire items, the largest effects in terms of Cohen's d can be observed for the Self-Concept Questionnaire Items (Selfconcept and Selfconcept_QI1 - Selfconcept_QI4 respectively). Regarding the working memory tests, the point estimates for Cohen's d show the largest effects on the tests related to the functional category of Storage-Processing (TG1 and TG1_n, TG1_v respectively). The point estimate for Cohen's d associated with the Mathematics Speed Test (Speed) is larger than the one related to the Mathematics Power Test (Power). Within the Mathematics Power Test, the point estimates for Cohen's d indicate the largest effects on the sixth and third task (Power_T6 and Power_T3).

A.4 Additional Group Comparison between High-Achievers and Under-Achievers II

For the second additional comparison between high-achievers and under-achievers, a subgroup of 19 high-achievers (13 female students (68%) and 6 male students (32%)) has been compared against a subgroup of 19 under-achievers (10 female students (53%) and 9 male students (47%)). The subgroup of high-achievers has been determined based on cut-off points on both scales, the Mathematics Posttest scale and the General Reasoning Ability scale, while the subgroup of under-achievers has been chosen based on a cut-off point on the Mathematics Posttest scale. The two subgroups are shown in Figure 53, where the selected high-achievers are marked with a green dot and the selected under-achievers are marked with an orange dot.

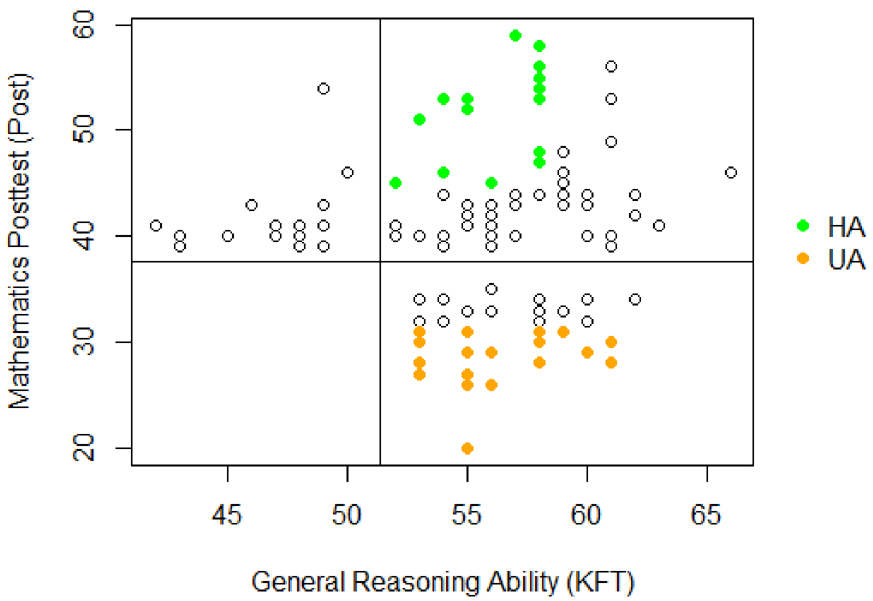


Figure 53: Selected high-achievers (marked with a green dot) and selected under-achievers (marked with an orange dot).

A.4.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

The focus of this subsection lies on the discussion of a comparison between the selected high-achievers and the selected under-achievers shown in Figure 53 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores Figures 54, 55 and 56 show the histograms of the performances of the selected high-achievers (HA) and of the selected under-achievers (UA) in the different tests. More concretely, the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post) are given in Figure 54, while the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed) are displayed in Figure 55. Moreover, the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) are collected in Figure 56. In addition to the various histograms, Table 31 indicates the descriptive statistics of the performances of the selected high-achievers (HA) and of the selected under-achievers (UA) in the second and third column respectively. The t-values and p-values of a two sample t-test comparing the two groups are listed in the fourth and fifth column of Table 31. Finally, the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are given in the last column of Table 31.

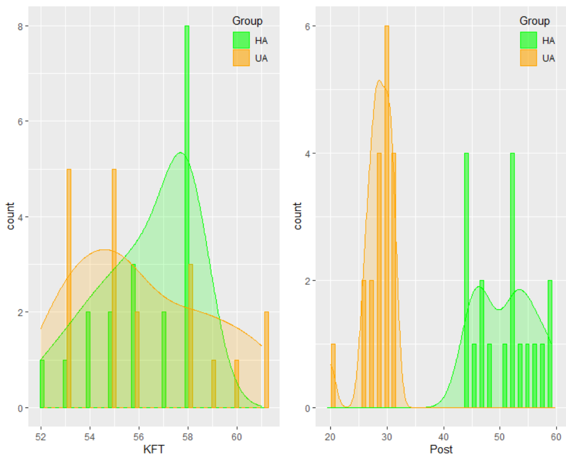


Figure 54: Histograms visualising the performances of the selected high-achievers (HA) and of the selected under-achievers (UA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

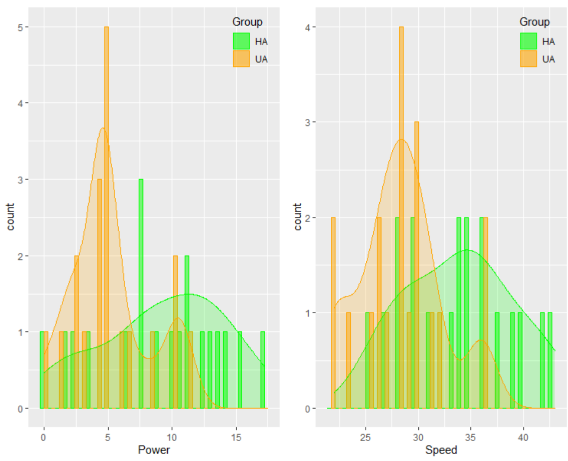


Figure 55: Histograms visualising the performances of the selected high-achievers (HA) and of the selected under-achievers (UA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

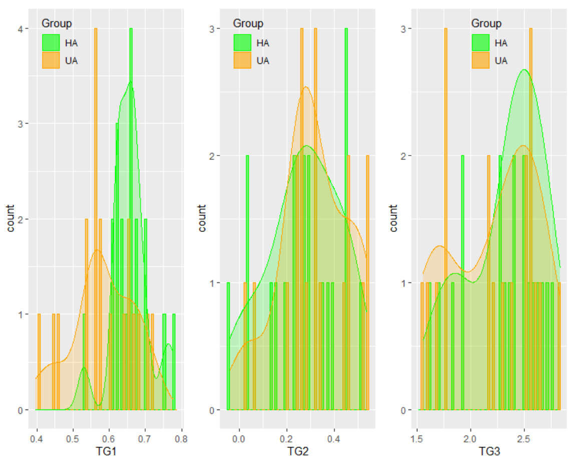


Figure 56: Histograms visualising the performances of the selected high-achievers (HA) and of the selected under-achievers (UA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.26 (1.94)	56.16 (2.75)	0.14	0.893	0.04 (−0.61, 0.70)
Mathematics Posttest (Post)	51.11 (5.00)	28.37 (2.61)	17.58	0.000	5.70 (4.22, 7.18)
Mathematics Power Test (Power)	9.30 (4.84)	5.12 (3.06)	3.18	0.003	1.03 (0.33, 1.73)
Mathematics Speed Test (Speed)	33.89 (5.14)	28.26 (3.94)	3.79	0.001	1.23 (0.51, 1.95)
WM Storage-Processing (TG1)	0.66 (0.05)	0.58 (0.09)	3.25	0.003	1.05 (0.35, 1.75)
WM Supervision (TG2)	0.27 (0.16)	0.32 (0.15)	−0.93	0.359	−0.30 (−0.96, 0.36)
WM Relational Integration (TG3)	2.32 (0.35)	2.21 (0.41)	0.94	0.355	0.30 (−0.36, 0.97)

Table 31: Group comparisons between the selected high-achievers (HA) and the selected under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 32), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 32), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 32).

As a first observation, it is important to note the proper assignment of the two groups: While there do not seem to be any substantial group differences in performance on the General Reasoning Ability Test, the results show group differences in performance on all three mathematics tests.

In terms of working memory functions, the results suggest that the two groups differ in performance on the working memory test associated with the functional category of Storage-Processing. On the other hand, the point estimate for Cohen's d related to the working memory function of Supervision indicates a small negative effect, while effect sizes ranging from −0.96, a large negative effect, to 0.36, a small positive effect, would be compatible with the data as well. Finally, the results regarding the functional category of Relational Integration specify a small positive effect. However, effect sizes ranging from −0.36, a small negative effect, to 0.97, a large positive effect, would also be supported by the data.

Individual Subtests or Tasks In order to evaluate the group comparison between the selected high-achievers and the selected under-achievers with respect to the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests in more detail, the respective subtests or tasks are collected in Table 32. The non-verbal (N), quantitative (Q) and verbal (V) subscales of the General Reasoning Ability Test (KFT) are denoted by KFT_N, KFT_Q and KFT_V respectively, while Power_T1 - Power_T6 describe the scores in Task 1 - Task 6 in the Mathematics Power Test (Power).

Moreover, for each functional category of working memory (TG1 - TG3), the corresponding figural (f), numerical (n) and verbal (v) subtests are given as well. Table 32 structures the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in the same way as Table 31. While KFT and Power in Table 31 relate to the sum scores of the associated subtests or tasks displayed in Table 32, TG1 - TG3 in Table 31 refer to the mean scores of the respective subtests listed in Table 32.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
KFT_N	23.53 (1.50)	22.16 (1.95)	2.42	0.021	0.79 (0.10, 1.47)
KFT_Q	17.68 (1.95)	17.84 (2.32)	−0.23	0.821	−0.07 (−0.73, 0.58)
KFT_V	15.05 (1.99)	16.16 (1.80)	−1.80	0.081	−0.58 (−1.25, 0.09)
Power_T1	1.87 (1.10)	1.29 (0.83)	1.83	0.076	0.59 (−0.08, 1.27)
Power_T2	2.01 (0.85)	1.43 (1.05)	1.87	0.071	0.61 (−0.07, 1.28)
Power_T3	1.24 (0.76)	0.61 (0.47)	3.07	0.005	1.00 (0.30, 1.69)
Power_T4	1.71 (0.85)	1.01 (0.73)	2.70	0.011	0.88 (0.19, 1.56)
Power_T5	0.96 (1.06)	0.41 (0.81)	1.81	0.079	0.59 (−0.09, 1.26)
Power_T6	1.51 (1.18)	0.37 (0.61)	3.76	0.001	1.22 (0.50, 1.94)
TG1_f	0.91 (0.09)	0.84 (0.15)	1.84	0.076	0.60 (−0.07, 1.27)
TG1_n	0.49 (0.07)	0.39 (0.10)	3.40	0.002	1.10 (0.40, 1.81)
TG1_v	0.58 (0.08)	0.52 (0.09)	2.11	0.042	0.68 (0.01, 1.36)
TG2_f	0.28 (0.23)	0.36 (0.25)	−1.04	0.303	−0.34 (−1.00, 0.32)
TG2_n	0.27 (0.22)	0.35 (0.22)	−1.10	0.279	−0.36 (−1.02, 0.31)
TG2_v	0.27 (0.13)	0.24 (0.12)	0.56	0.578	0.18 (−0.48, 0.84)
TG3_f	2.45 (0.43)	2.39 (0.51)	0.36	0.724	0.12 (−0.54, 0.77)
TG3_n	2.29 (0.37)	2.25 (0.64)	0.27	0.792	0.09 (−0.57, 0.74)
TG3_v	2.23 (0.62)	1.98 (0.50)	1.36	0.183	0.44 (−0.22, 1.11)

Table 32: Group comparisons between the selected high-achievers (HA) and the selected under-achievers (UA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

There do not seem to be any group differences in performance on the quantitative subtest of the General Reasoning Ability Test (KFT_Q), while there could be group differences in performance with respect to the non-verbal (KFT_N) and verbal (KFT_V) subtest.

The two groups clearly differ in performance on the third, fourth and sixth task of the Mathematics Power Test (Power_T3, Power_T4 and Power_T6). For the remaining tasks (Power_T1, Power_T2 and Power_T5), the results all suggest a medium positive effect, while effect sizes ranging from negligible effects to large positive effects would be compatible with the data as well. Regarding the working memory subtests, there appear to be group differences in performance on all three subtests of the working memory function Storage-Processing (TG1_f, TG1_n and TG1_v). For the figural and numerical subtest of the functional category of Supervision (TG2_f and TG2_n), the point estimates for Cohen's d both specify a small negative effect. However, effect sizes ranging from large negative effects to small positive effects would also be supported by the data. In addition, the results related to the verbal subtest (TG2_v) show a negligible effect, with effect sizes ranging from -0.48, a small negative effect, to 0.84, a large positive effect, being in line with the data as well. Lastly, the point estimates for Cohen's d associated with the figural and numerical subtest of the functional category of Relational Integration (TG3_f and TG3_n) both indicate a negligible effect, while effect sizes ranging from medium negative effects to medium positive effects would be compatible with the data as well. For the verbal subtest (TG3_v), the results suggest a small positive effect. Nonetheless, effect sizes ranging from -0.22, a small negative effect, to 1.11, a large positive effect, would also be supported by the data.

A.4.2 Mathematical Self-Concept and Interest in Mathematics

In addition to the comparisons discussed in the previous subsection, the selected high-achievers and the selected under-achievers shown in Figure 53 have been compared with respect to their mathematical self-concept as well as their interest in mathematics. This subsection presents the corresponding results.

Sum Scores and Mean Scores The mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the selected high-achievers and the selected under-achievers are visualised in Figure 57. In addition to this graphical representation, Table 33 indicates the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals. Moreover, to enable an overview on the different results, Table 33 shows the results associated with the mathematics tests and the General Reasoning Ability Test already presented in Table 31. Table 33 is structured in the same way as Table 31.

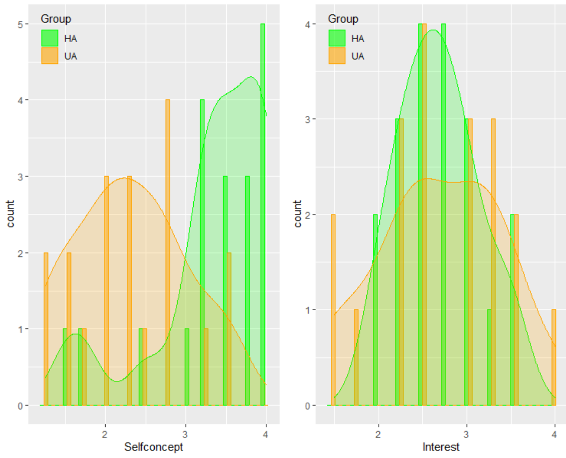


Figure 57: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the selected high-achievers (HA) and the selected under-achievers (UA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
General Reasoning Ability (KFT)	56.26 (1.94)	56.16 (2.75)	0.14	0.893	0.04 (−0.61, 0.70)
Mathematics Posttest (Post)	51.11 (5.00)	28.37 (2.61)	17.58	0.000	5.70 (4.22, 7.18)
Mathematics Power Test (Power)	9.30 (4.84)	5.12 (3.06)	3.18	0.003	1.03 (0.33, 1.73)
Mathematics Speed Test (Speed)	33.89 (5.14)	28.26 (3.94)	3.79	0.001	1.23 (0.51, 1.95)
Self-Concept Questionnaire Items (Selfconcept)	3.34 (0.73)	2.30 (0.70)	4.50	0.000	1.46 (0.72, 2.20)
Interest Questionnaire Items (Interest)	2.68 (0.45)	2.70 (0.70)	−0.07	0.945	−0.02 (−0.68, 0.64)

Table 33: Group comparisons between the selected high-achievers (HA) and the selected under-achievers (UA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 32), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 32), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 34) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 34) respectively.

Clear group differences regarding the mean scores of the Self-Concept Questionnaire Items can be observed: The selected under-achievers appear to have a substantially lower mathematical self-concept compared to the selected high-achievers. On the other hand, the two groups do not seem to differ with respect to their mean scores of the Interest Questionnaire Items.

Individual Questionnaire Items Table 34 lists the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and Interest Questionnaire Items (Interest_QI5 - Interest_QI8) in order to enable a more detailed group comparison between the selected high-achievers and the selected under-achievers. Table 34 arranges the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in the same way as Table 33. As a remark, it is important to mention that Selfconcept and Interest in Table 33 represent the mean scores of Selfconcept_QI1 - Selfconcept_QI4 and of Interest_QI5 - Interest_QI8 respectively.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Selfconcept_QI1	3.26 (0.87)	2.42 (0.90)	2.93	0.006	0.95 (0.26, 1.64)
Selfconcept_QI2	3.32 (0.82)	2.37 (0.60)	4.07	0.000	1.32 (0.59, 2.05)
Selfconcept_QI3	3.26 (0.99)	1.74 (0.73)	5.40	0.000	1.75 (0.98, 2.52)
Selfconcept_QI4	3.53 (0.70)	2.68 (0.95)	3.12	0.004	1.01 (0.31, 1.71)
Interest_QI5	2.47 (0.77)	2.58 (1.17)	−0.33	0.746	−0.11 (−0.76, 0.55)
Interest_QI6	2.53 (0.84)	2.37 (0.83)	0.58	0.564	0.19 (−0.47, 0.85)
Interest_QI7	2.53 (0.70)	2.58 (0.69)	−0.23	0.817	−0.08 (−0.73, 0.58)
Interest_QI8	3.21 (0.54)	3.26 (0.65)	−0.27	0.788	−0.09 (−0.75, 0.57)

Table 34: Group comparisons between the selected high-achievers (HA) and the selected under-achievers (UA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

The two groups substantially differ on all Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4). Regarding the Interest Questionnaire Items, the results related to the fifth, seventh and eighth item (Interest_QI5, Interest_QI7 and Interest_QI8) all indicate negligible effects. However, effect sizes ranging from medium negative effects to medium positive effects would also be in line with the data. Finally, the point estimate for Cohen's d associated with the sixth item (Interest_QI6) suggests a negligible effect, while effect sizes ranging from small negative effects to large positive effects would be supported by the data as well.

A.4.3 Summary of the Results

The results from Subsection A.4.1 and Subsection A.4.2 with respect to various group comparisons between high-achieving and under-achieving students are merged in this subsection. Table 35 arranges these results in descending order of the respective effect sizes in terms of Cohen's d.

Test	M_{HA} (SD_{HA})	M_{UA} (SD_{UA})	t	p	d (CI)
Post	51.11 (5.00)	28.37 (2.61)	17.58	0.000	5.70 (4.22, 7.18)
Selfconcept_QI3	3.26 (0.99)	1.74 (0.73)	5.40	0.000	1.75 (0.98, 2.52)
Selfconcept	3.34 (0.73)	2.30 (0.70)	4.50	0.000	1.46 (0.72, 2.20)
Selfconcept_QI2	3.32 (0.82)	2.37 (0.60)	4.07	0.000	1.32 (0.59, 2.05)
Speed	33.89 (5.14)	28.26 (3.94)	3.79	0.001	1.23 (0.51, 1.95)
Power_T6	1.51 (1.18)	0.37 (0.61)	3.76	0.001	1.22 (0.50, 1.94)
TG1_n	0.49 (0.07)	0.39 (0.10)	3.40	0.002	1.10 (0.40, 1.81)
TG1	0.66 (0.05)	0.58 (0.09)	3.25	0.003	1.05 (0.35, 1.75)
Power	9.30 (4.84)	5.12 (3.06)	3.18	0.003	1.03 (0.33, 1.73)
Selfconcept_QI4	3.53 (0.70)	2.68 (0.95)	3.12	0.004	1.01 (0.31, 1.71)
Power_T3	1.24 (0.76)	0.61 (0.47)	3.07	0.005	1.00 (0.30, 1.69)
Selfconcept_QI1	3.26 (0.87)	2.42 (0.90)	2.93	0.006	0.95 (0.26, 1.64)
Power_T4	1.71 (0.85)	1.01 (0.73)	2.70	0.011	0.88 (0.19, 1.56)
KFT_N	23.53 (1.50)	22.16 (1.95)	2.42	0.021	0.79 (0.10, 1.47)
TG1_v	0.58 (0.08)	0.52 (0.09)	2.11	0.042	0.68 (0.01, 1.36)
Power_T2	2.01 (0.85)	1.43 (1.05)	1.87	0.071	0.61 (−0.07, 1.28)
TG1_f	0.91 (0.09)	0.84 (0.15)	1.84	0.076	0.60 (−0.07, 1.27)
Power_T1	1.87 (1.10)	1.29 (0.83)	1.83	0.076	0.59 (−0.08, 1.27)
Power_T5	0.96 (1.06)	0.41 (0.81)	1.81	0.079	0.59 (−0.09, 1.26)
TG3_v	2.23 (0.62)	1.98 (0.50)	1.36	0.183	0.44 (−0.22, 1.11)
TG3	2.32 (0.35)	2.21 (0.41)	0.94	0.355	0.30 (−0.36, 0.97)
Interest_QI6	2.53 (0.84)	2.37 (0.83)	0.58	0.564	0.19 (−0.47, 0.85)
TG2_v	0.27 (0.13)	0.24 (0.12)	0.56	0.578	0.18 (−0.48, 0.84)
TG3_f	2.45 (0.43)	2.39 (0.51)	0.36	0.724	0.12 (−0.54, 0.77)
TG3_n	2.29 (0.37)	2.25 (0.64)	0.27	0.792	0.09 (−0.57, 0.74)
KFT	56.26 (1.94)	56.16 (2.75)	0.14	0.893	0.04 (−0.61, 0.70)
Interest	2.68 (0.45)	2.70 (0.70)	−0.07	0.945	−0.02 (−0.68, 0.64)
KFT_Q	17.68 (1.95)	17.84 (2.32)	−0.23	0.821	−0.07 (−0.73, 0.58)
Interest_QI7	2.53 (0.70)	2.58 (0.69)	−0.23	0.817	−0.08 (−0.73, 0.58)
Interest_QI8	3.21 (0.54)	3.26 (0.65)	−0.27	0.788	−0.09 (−0.75, 0.57)
Interest_QI5	2.47 (0.77)	2.58 (1.17)	−0.33	0.746	−0.11 (−0.76, 0.55)
TG2	0.27 (0.16)	0.32 (0.15)	−0.93	0.359	−0.30 (−0.96, 0.36)
TG2_f	0.28 (0.23)	0.36 (0.25)	−1.04	0.303	−0.34 (−1.00, 0.32)
TG2_n	0.27 (0.22)	0.35 (0.22)	−1.10	0.279	−0.36 (−1.02, 0.31)
KFT_V	15.05 (1.99)	16.16 (1.80)	−1.80	0.081	−0.58 (−1.25, 0.09)

Table 35: Summary of all group comparisons between the selected high-achievers (HA) and the selected under-achievers (UA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

Substantial group differences in performance appear on the Mathematics Posttest (Post), which indicates that the two groups were properly assigned.

The point estimates for Cohen's d referring to the questionnaire items specify the largest effects on the Self-Concept Questionnaire Items (Selfconcept and Selfconcept_QI1 - Selfconcept_QI4 respectively). For the Mathematics Speed Test (Speed), the point estimate for Cohen's d is larger than the one related to the Mathematics Power Test (Power), while medium to large positive effects are suggested by the point estimates for all tasks of the Mathematics Power Test (Power_T1 - Power_T6). With respect to the working memory tests, the largest effects in terms of Cohen's d are found for the tests measuring the working memory function Storage-Processing (TG1 and TG1_n, TG1_v, TG1_f respectively). Lastly, while the point estimate for Cohen's d related to the non-verbal subtest of the General Reasoning Ability Test (KFT_N) indicates a medium positive effect, the one associated with the verbal subtest (KFT_V) suggests a medium negative effect.

A.5 Additional Subgroup Comparison within High-Achievers

As an additional comparison within the group of high-achievers, a subgroup consisting of the 12 high-achievers with the highest scores in the Mathematics Power Test (9 female students (75%), 2 male students (17%) and one student, who did not specify the gender (8%)) has been compared against the remaining high-achievers (55 students: 35 female students (64%) and 20 male students (36%)). Figure 58 visualises the two subgroups: The top scoring high-achievers are marked with a green dot, while the remaining high-achievers are marked with a blue dot.

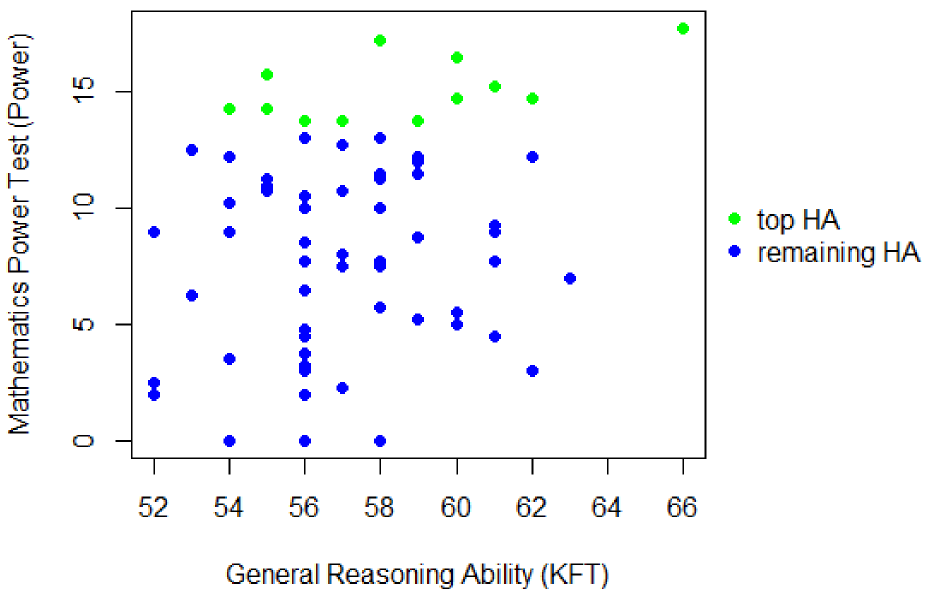


Figure 58: Top scoring high-achievers (marked with a green dot) and remaining high-achievers (marked with a blue dot).

A.5.1 General Reasoning Ability Test, Mathematics Tests and Working Memory Tests

This subsection aims at discussing a comparison between the two subgroups depicted in Figure 58 with respect to their performances in the General Reasoning Ability Test, the different mathematics tests and the working memory tests.

Sum Scores and Mean Scores The histograms of the performances of the two subgroups in the different tests are presented in Figures 59, 60 and 61. More specifically, Figure 59 shows the histograms related to the General Reasoning Ability Test (KFT) and the Mathematics Posttest (Post), while Figure 60 depicts the histograms of the Mathematics Power Test (Power) and the Mathematics Speed Test (Speed). In addition, Figure 61 arranges the histograms belonging to the tests of the working memory functions Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3). In order to complement all these histograms, Table 36 contains the descriptive statistics of the performances of the the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and of the remaining high-achievers (rmngHA) in the second and third column respectively. The t-values and p-values of a two sample t-test comparing the two subgroups are listed in the fourth and fifth column of Table 36. Finally, the corresponding effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are given in the last column of Table 36.

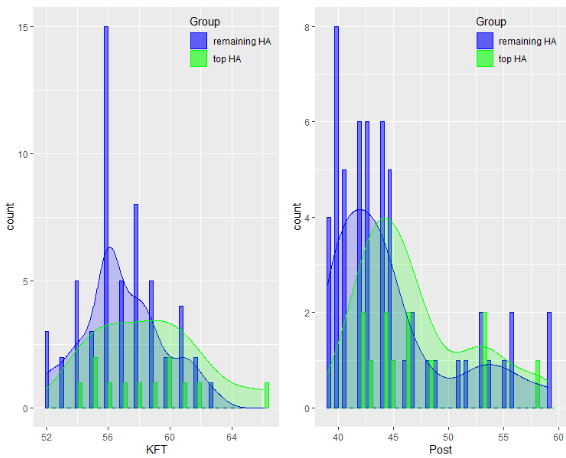


Figure 59: Histograms visualising the performances of the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and of the remaining high-achievers (rmngHA) in the **General Reasoning Ability Test (KFT)** and in the **Mathematics Posttest (Post)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

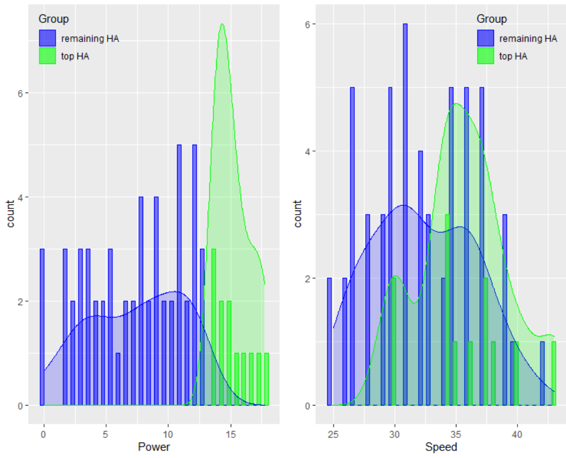


Figure 60: Histograms visualising the performances of the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and of the remaining high-achievers (rmngHA) in the **Mathematics Power Test (Power)** and in the **Mathematics Speed Test (Speed)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

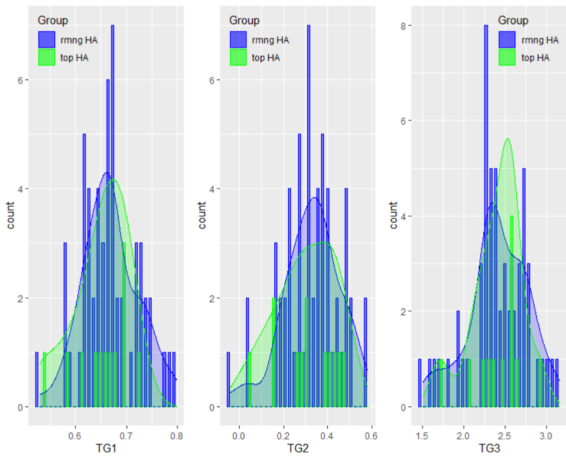


Figure 61: Histograms visualising the performances of the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and of the remaining high-achievers (rmngHA) in the working memory tests related to the functional categories of **Storage-Processing (TG1)**, **Supervision (TG2)** and **Relational Integration (TG3)** respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different tests.

Test	M_{topHA} (SD_{topHA})	M_{rmngHA} (SD_{rmngHA})	t	p	d (CI)
General Reasoning Ability (KFT)	58.58 (3.48)	56.95 (2.63)	1.54	0.146	0.59 (−0.06, 1.23)
Mathematics Posttest (Post)	47.00 (5.08)	44.82 (5.39)	1.33	0.200	0.41 (−0.23, 1.05)
Mathematics Power Test (Power)	15.15 (1.39)	7.35 (3.84)	11.90	0.000	2.20 (1.46, 2.94)
Mathematics Speed Test (Speed)	35.67 (3.75)	32.33 (4.22)	2.73	0.014	0.81 (0.15, 1.46)
WM Storage-Processing (TG1)	0.65 (0.05)	0.67 (0.05)	−1.04	0.315	−0.33 (−0.97, 0.31)
WM Supervision (TG2)	0.30 (0.13)	0.33 (0.13)	−0.71	0.489	−0.23 (−0.87, 0.40)
WM Relational Integration (TG3)	2.40 (0.31)	2.41 (0.36)	−0.12	0.910	−0.03 (−0.67, 0.60)

Table 36: Group comparisons between the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 37), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 37), in the Mathematics Speed Test (sum score) and in the working memory tests (TG1 - TG3 describe the mean scores of the related subtests given in Table 37).

The substantial group differences in performance on the Mathematics Power Test confirm the proper assignment of the two subgroups.

The point estimate for Cohen's d associated with the General Reasoning Ability Test shows a medium positive effect, while effect sizes ranging from −0.06, a negligible effect, to 1.23, a large positive effect, would be compatible with the data as well. On the other hand, the results related to the Mathematics Posttest indicate a small positive effect. Nonetheless, effect sizes ranging from −0.23, a small negative effect, to 1.05, a large positive effect, would also be supported by the data. Furthermore, there seem to be group differences in performance on the Mathematics Speed Test. Regarding working memory functions, the point estimates for Cohen's d related to the functional categories of Storage-Processing and Supervision both specify a small negative effect, while effect sizes ranging from large negative effects to small positive effects would also be in line with the data. Lastly, substantial group differences in performance cannot be detected with respect to the working memory function of Relational Integration.

Individual Subtests or Tasks For a more detailed analysis of the group comparison between the two subgroups with respect to the General Reasoning Ability Test, the Mathematics Power Test and the working memory tests, the respective subtests or tasks are given in Table 37. More specifically, Table 37 arranges the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subscales of the General Reasoning Ability Test as well as the scores in Task 1 - Task 6 in the Mathematics Power Test (Power_T1 - Power_T6). Moreover, for each functional category of working memory (TG1 - TG3), Table 37 lists the corresponding figural (f), numerical (n) and verbal (v) subtests. The structure of Table 37 regarding the descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals is analogous to the one of Table 36. While KFT and Power in Table 36 refer to the sum scores of the associated subtests or tasks displayed in Table 37, TG1 - TG3 in Table 36 describe the mean scores of the related subtests collected in Table 37.

Test	M_{topHA} (SD_{topHA})	M_{rmngHA} (SD_{rmngHA})	t	p	d (CI)
KFT_N	23.50 (1.17)	22.82 (1.69)	1.68	0.108	0.42 (−0.22, 1.06)
KFT_Q	18.83 (1.59)	17.96 (1.99)	1.64	0.117	0.45 (−0.19, 1.09)
KFT_V	16.25 (3.02)	16.16 (2.10)	0.09	0.926	0.04 (−0.60, 0.67)
Power_T1	2.67 (0.53)	1.58 (1.02)	5.31	0.000	1.14 (0.47, 1.81)
Power_T2	2.94 (0.16)	1.78 (0.95)	8.52	0.000	1.33 (0.66, 2.01)
Power_T3	1.96 (0.67)	1.14 (0.75)	3.75	0.002	1.11 (0.45, 1.78)
Power_T4	2.71 (0.35)	1.36 (0.77)	9.29	0.000	1.88 (1.16, 2.59)
Power_T5	2.48 (0.59)	0.57 (0.72)	9.73	0.000	2.71 (1.92, 3.50)
Power_T6	2.40 (0.53)	0.93 (0.96)	7.33	0.000	1.62 (0.93, 2.32)
TG1_f	0.90 (0.06)	0.90 (0.09)	0.03	0.975	0.01 (−0.63, 0.64)
TG1_n	0.48 (0.08)	0.50 (0.08)	−0.76	0.460	−0.26 (−0.89, 0.38)
TG1_v	0.57 (0.09)	0.60 (0.08)	−1.18	0.256	−0.41 (−1.05, 0.23)
TG2_f	0.32 (0.15)	0.34 (0.19)	−0.39	0.704	−0.10 (−0.74, 0.53)
TG2_n	0.32 (0.19)	0.35 (0.19)	−0.40	0.695	−0.12 (−0.76, 0.51)
TG2_v	0.25 (0.14)	0.29 (0.12)	−1.07	0.301	−0.38 (−1.02, 0.26)
TG3_f	2.43 (0.41)	2.56 (0.49)	−1.00	0.331	−0.29 (−0.92, 0.35)
TG3_n	2.40 (0.27)	2.41 (0.58)	−0.14	0.887	−0.03 (−0.67, 0.61)
TG3_v	2.37 (0.59)	2.26 (0.55)	0.63	0.537	0.21 (−0.43, 0.85)

Table 37: Group comparisons between the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) with respect to their performances in the non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests of the General Reasoning Ability Test, in the individual tasks of the Mathematics Power Test (Power_T1 - Power_T6) and in the figural (f), numerical (n) and verbal (v) working memory subtests associated with the functional categories of Storage-Processing (TG1, performances on the individual subtests are measured by the respective average proportion of correctly recalled items), Supervision (TG2, performances on the individual subtests are measured by the respective switch costs) and Relational Integration (TG3, performances on the individual subtests are measured by the respective detection performances).

Regarding the Mathematics Power Test, substantial group differences in performance can be observed for all tasks (Power_T1 - Power_T6), which is a further indication of the proper assignment of the two subgroups.

The results related to the non-verbal and quantitative subtest of the General Reasoning Ability Test (KFT_N and KFT_Q) both show a small positive effect. Nonetheless, effect sizes ranging from small negative or negligible effects to large positive effects would also be compatible with the data. In addition, the two subgroups do not appear to differ in performance on the verbal subtest (KFT_V). The two subgroups do not seem to differ in performance on the figural subtest of the working memory function Storage-Processing (TG1_f). For the numerical and verbal subtest (TG1_n and TG1_v), the point estimates for Cohen's d both specify small negative effects, while effect sizes ranging from large negative effects to small positive effects would be in line with the data as well. Regarding the functional category of Supervision, the results show negligible effects for both, the figural and numerical subtest (TG2_f and TG2_n). However, effect sizes ranging from medium negative effects to medium positive effects would also be supported by the data. The point estimate for Cohen's d associated with the verbal subtest (TG2_v) indicates a small negative effect, with effect sizes ranging from -1.02, a large negative effect, to 0.26, a small positive effect, being in line with the data as well. Finally, there do not seem to be any group differences in performance on the numerical subtest of the functional category of Relational Integration (TG3_n), while the point estimates for Cohen's d related to the figural and verbal subtest (TG3_f and TG3_v) suggest a small negative effect and a small positive effect respectively.

A.5.2 Mathematical Self-Concept and Interest in Mathematics

To further compare the two subgroups depicted in Figure 58, a comparison of the two subgroups with respect to their mathematical self-concept as well as their interest in mathematics is discussed in this subsection.

Sum Scores and Mean Scores The mean scores of the Self-Concept Questionnaire Items (Selfconcept) as well as of the Interest Questionnaire Items (Interest) for the two subgroups are visualised in Figure 62. To complement this graphical representation, the corresponding descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals are given in Table 38. To facilitate the comparison between the different results, Table 38 additionally contains the results related to the mathematics tests and the General Reasoning Ability Test already shown in Table 36. Table 38 is structured in the same way as Table 36.

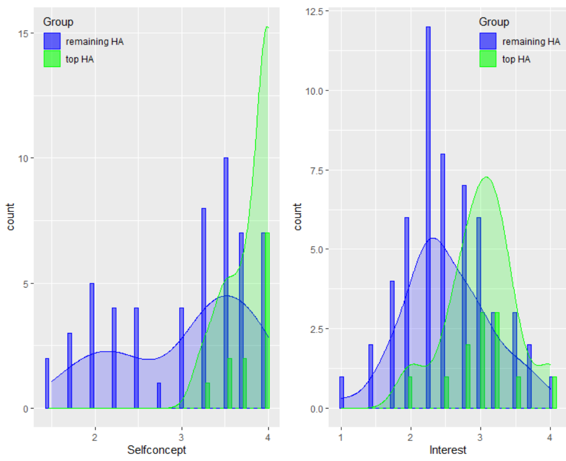


Figure 62: Histograms visualising the mean scores of the **Self-Concept Questionnaire Items (Selfconcept)** and of the **Interest Questionnaire Items (Interest)** for the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) respectively. Note that the bars of the histograms are dodged and that the scaling might be different for different histograms.

Test	M_{topHA} (SD_{topHA})	M_{rmngHA} (SD_{rmngHA})	t	p	d (CI)
General Reasoning Ability (KFT)	58.58 (3.48)	56.95 (2.63)	1.54	0.146	0.59 (−0.06, 1.23)
Mathematics Posttest (Post)	47.00 (5.08)	44.82 (5.39)	1.33	0.200	0.41 (−0.23, 1.05)
Mathematics Power Test (Power)	15.15 (1.39)	7.35 (3.84)	11.90	0.000	2.20 (1.46, 2.94)
Mathematics Speed Test (Speed)	35.67 (3.75)	32.33 (4.22)	2.73	0.014	0.81 (0.15, 1.46)
Self-Concept Questionnaire Items (Selfconcept)	3.81 (0.26)	3.04 (0.76)	6.03	0.000	1.10 (0.43, 1.76)
Interest Questionnaire Items (Interest)	3.02 (0.51)	2.53 (0.62)	2.94	0.008	0.82 (0.17, 1.47)

Table 38: Group comparisons between the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) with respect to their performances in the General Reasoning Ability Test (KFT relates to the sum score of the respective subtests displayed in Table 37), in the Mathematics Posttest (sum score), in the Mathematics Power Test (Power refers to the sum score of the corresponding tasks presented in Table 37), in the Mathematics Speed Test (sum score) as well as regarding their mean scores of the Self-Concept Questionnaire Items (Selfconcept describes the mean score of the individual Self-Concept Questionnaire Items listed in Table 39) and of the Interest Questionnaire Items (Interest corresponds to the mean score of the individual Interest Questionnaire Items given in Table 39) respectively.

The results suggest that the two subgroups differ with respect to the mean scores of the Self-Concept Questionnaire Items as well as regarding the mean scores of the Interest Questionnaire Items.

Individual Questionnaire Items For a more detailed group comparison between the two subgroups with respect to the Self-Concept Questionnaire Items and the Interest Questionnaire Items, Table 39 lists the individual Questionnaire Items. Selfconcept_QI1 - Selfconcept_QI4 correspond to the Self-Concept Questionnaire Items, while Interest_QI5 - Interest_QI8 represent the Interest Questionnaire Items. Table 39 arranges the related descriptive statistics, the t-values and p-values of a two sample t-test and the effect sizes in terms of Cohen's d as well as the respective 95% confidence intervals in an analogous manner as Table 38. While Selfconcept in Table 38 relates to the mean scores of Selfconcept_QI1 - Selfconcept_QI4, Interest in Table 38 refers to the mean scores of Interest_QI5 - Interest_QI8.

Test	M_{topHA} (SD_{topHA})	M_{rmngHA} (SD_{rmngHA})	t	p	d (CI)
Selfconcept_QI1	3.75 (0.45)	3.05 (0.93)	3.84	0.001	0.80 (0.15, 1.45)
Selfconcept_QI2	3.83 (0.39)	3.05 (0.73)	5.21	0.000	1.14 (0.47, 1.80)
Selfconcept_QI3	3.75 (0.45)	2.75 (1.06)	5.19	0.000	1.02 (0.36, 1.68)
Selfconcept_QI4	3.92 (0.29)	3.31 (0.81)	4.41	0.000	0.81 (0.16, 1.46)
Interest_QI5	2.67 (0.89)	2.38 (0.91)	1.00	0.331	0.31 (−0.33, 0.95)
Interest_QI6	2.83 (0.58)	2.24 (0.84)	2.96	0.007	0.75 (0.10, 1.40)
Interest_QI7	3.00 (0.60)	2.42 (0.88)	2.77	0.011	0.70 (0.05, 1.34)
Interest_QI8	3.58 (0.51)	3.07 (0.72)	2.88	0.009	0.74 (0.09, 1.39)

Table 39: Group comparisons between the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) with respect to the individual Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4) and the individual Interest Questionnaire Items (Interest_QI5 - Interest_QI8).

First of all, substantial group differences can be observed for all Self-Concept Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4). The point estimate for Cohen's d associated with the fifth item (Interest_QI5) specifies a small positive effect, while effect sizes ranging from small negative effects to large positive effects would be compatible with the data as well. Lastly, the two subgroups seem to differ on all the remaining Interest Questionnaire Items (Interest_QI6 - Interest_QI8).

A.5.3 Summary of the Results

This subsection aims at summarising the results from Subsections A.5.1 and A.5.2, which reported on different group comparisons between the 12 high-achievers, who scored the highest in the Mathematics Power Test, and the remaining high-achievers. Table 40 arranges these results in descending order of the respective effect sizes in terms of Cohen's d.

Test	M_{topHA} (SD_{topHA})	M_{rmngHA} (SD_{rmngHA})	t	p	d (CI)
Power_T5	2.48 (0.59)	0.57 (0.72)	9.73	0.000	2.71 (1.92, 3.50)
Power	15.15 (1.39)	7.35 (3.84)	11.90	0.000	2.20 (1.46, 2.94)
Power_T4	2.71 (0.35)	1.36 (0.77)	9.29	0.000	1.88 (1.16, 2.59)
Power_T6	2.40 (0.53)	0.93 (0.96)	7.33	0.000	1.62 (0.93, 2.32)
Power_T2	2.94 (0.16)	1.78 (0.95)	8.52	0.000	1.33 (0.66, 2.01)
Power_T1	2.67 (0.53)	1.58 (1.02)	5.31	0.000	1.14 (0.47, 1.81)
Selfconcept_QI2	3.83 (0.39)	3.05 (0.73)	5.21	0.000	1.14 (0.47, 1.80)
Power_T3	1.96 (0.67)	1.14 (0.75)	3.75	0.002	1.11 (0.45, 1.78)
Selfconcept	3.81 (0.26)	3.04 (0.76)	6.03	0.000	1.10 (0.43, 1.76)
Selfconcept_QI3	3.75 (0.45)	2.75 (1.06)	5.19	0.000	1.02 (0.36, 1.68)
Interest	3.02 (0.51)	2.53 (0.62)	2.94	0.008	0.82 (0.17, 1.47)
Selfconcept_QI4	3.92 (0.29)	3.31 (0.81)	4.41	0.000	0.81 (0.16, 1.46)
Speed	35.67 (3.75)	32.33 (4.22)	2.73	0.014	0.81 (0.15, 1.46)
Selfconcept_QI1	3.75 (0.45)	3.05 (0.93)	3.84	0.001	0.80 (0.15, 1.45)
Interest_QI6	2.83 (0.58)	2.24 (0.84)	2.96	0.007	0.75 (0.10, 1.40)
Interest_QI8	3.58 (0.51)	3.07 (0.72)	2.88	0.009	0.74 (0.09, 1.39)
Interest_QI7	3.00 (0.60)	2.42 (0.88)	2.77	0.011	0.70 (0.05, 1.34)
KFT	58.58 (3.48)	56.95 (2.63)	1.54	0.146	0.59 (−0.06, 1.23)
KFT_Q	18.83 (1.59)	17.96 (1.99)	1.64	0.117	0.45 (−0.19, 1.09)
KFT_N	23.50 (1.17)	22.82 (1.69)	1.68	0.108	0.42 (−0.22, 1.06)
Post	47.00 (5.08)	44.82 (5.39)	1.33	0.200	0.41 (−0.23, 1.05)
Interest_QI5	2.67 (0.89)	2.38 (0.91)	1.00	0.331	0.31 (−0.33, 0.95)
TG3_v	2.37 (0.59)	2.26 (0.55)	0.63	0.537	0.21 (−0.43, 0.85)
KFT_V	16.25 (3.02)	16.16 (2.10)	0.09	0.926	0.04 (−0.60, 0.67)
TG1_f	0.90 (0.06)	0.90 (0.09)	0.03	0.975	0.01 (−0.63, 0.64)
TG3_n	2.40 (0.27)	2.41 (0.58)	−0.14	0.887	−0.03 (−0.67, 0.61)
TG3	2.40 (0.31)	2.41 (0.36)	−0.12	0.910	−0.03 (−0.67, 0.60)
TG2_f	0.32 (0.15)	0.34 (0.19)	−0.39	0.704	−0.10 (−0.74, 0.53)
TG2_n	0.32 (0.19)	0.35 (0.19)	−0.40	0.695	−0.12 (−0.76, 0.51)
TG2	0.30 (0.13)	0.33 (0.13)	−0.71	0.489	−0.23 (−0.87, 0.40)
TG1_n	0.48 (0.08)	0.50 (0.08)	−0.76	0.460	−0.26 (−0.89, 0.38)
TG3_f	2.43 (0.41)	2.56 (0.49)	−1.00	0.331	−0.29 (−0.92, 0.35)
TG1	0.65 (0.05)	0.67 (0.05)	−1.04	0.315	−0.33 (−0.97, 0.31)
TG2_v	0.25 (0.14)	0.29 (0.12)	−1.07	0.301	−0.38 (−1.02, 0.26)
TG1_v	0.57 (0.09)	0.60 (0.08)	−1.18	0.256	−0.41 (−1.05, 0.23)

Table 40: Summary of all group comparisons between the 12 high-achievers (topHA), who scored the highest in the Mathematics Power Test, and the remaining high-achievers (rmngHA) with respect to the General Reasoning Ability Test (KFT) and its non-verbal (KFT_N), quantitative (KFT_Q) and verbal (KFT_V) subtests, the Mathematics Posttest (Post), the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), the Mathematics Speed Test (Speed), the working memory measures related to the functional categories of Storage-Processing (TG1), Supervision (TG2) and Relational Integration (TG3) and the respective figural (f), numerical (n) and verbal (v) subtests, the Self-Concept measure (Selfconcept) and the corresponding individual Questionnaire Items (Selfconcept_QI1 - Selfconcept_QI4), and the Interest measure (Interest) and the corresponding individual Questionnaire Items (Interest_QI5 - Interest_QI8). The results are arranged in descending order of the respective effect sizes in terms of Cohen's d.

The largest effects in terms of Cohen's d are found for the Mathematics Power Test (Power) and its individual tasks (Power_T1 - Power_T6), which shows the proper assignment of the two subgroups. With respect to the questionnaire items, it can be observed that the point estimate for Cohen's d related to the Self-Concept Questionnaire Items (Selfconcept) is larger than the one associated with the Interest Questionnaire Items (Interest). In addition, the results indicate that for the Mathematics Speed Test (Speed), the point estimate for Cohen's d is larger than the one referring to the General Reasoning Ability Test (KFT). In turn, the point estimate for Cohen's d associated with the General Reasoning Ability Test (KFT) is larger than the point estimates related to the working memory tests.

A.6 German Version of the Questionnaire on Mathematical Self-Concept and Interest in Mathematics

Fragebogen

Persönlicher Code:

Bitte kreuzen Sie bei den folgenden Aussagen jeweils das Zutreffende an.

1) Ich war schon immer gut in Mathematik.
   - o trifft nicht zu
   - o trifft eher nicht zu
   - o trifft eher zu
   - o trifft zu

2) Im Fach Mathematik bekomme ich gute Noten.
   - o trifft nicht zu
   - o trifft eher nicht zu
   - o trifft eher zu
   - o trifft zu

3) Mathematik ist eines meiner besten Fächer.
   - o trifft nicht zu
   - o trifft eher nicht zu
   - o trifft eher zu
   - o trifft zu

4) Ich bin einfach nicht gut in Mathematik.
   - o trifft nicht zu
   - o trifft eher nicht zu
   - o trifft eher zu
   - o trifft zu

5) Wenn ich mich mit Mathematik beschäftige, vergesse ich manchmal alles um mich herum.
   - o trifft nicht zu
   - o trifft eher nicht zu
   - o trifft eher zu
   - o trifft zu

Bitte blättern Sie um!

6) Weil mir die Beschäftigung mit Mathematik Spass macht, würde ich das nicht gerne aufgeben.

- o trifft nicht zu
- o trifft eher nicht zu
- o trifft eher zu
- o trifft zu

7) Mathematik ist mir persönlich wichtig.

- o trifft nicht zu
- o trifft eher nicht zu
- o trifft eher zu
- o trifft zu

8) Es ist mir wichtig, gut in Mathematik zu sein.

- o trifft nicht zu
- o trifft eher nicht zu
- o trifft eher zu
- o trifft zu

A.7 English Version of the Questionnaire on Mathematical Self-Concept and Interest in Mathematics

Questionnaire

Personal Code:

Please tick the appropriate box for the following statements.

1) I have always done well in mathematics.
   - o disagree
   - o disagree somewhat
   - o agree somewhat
   - o agree

2) I get good marks in mathematics.
   - o disagree
   - o disagree somewhat
   - o agree somewhat
   - o agree

3) Mathematics is one of my best subjects.
   - o disagree
   - o disagree somewhat
   - o agree somewhat
   - o agree

4) I am simply not good in mathematics.
   - o disagree
   - o disagree somewhat
   - o agree somewhat
   - o agree

5) When I do mathematics, I sometimes get totally absorbed.
   - o disagree
   - o disagree somewhat
   - o agree somewhat
   - o agree

Please turn the page!

6) Because doing mathematics is fun, I wouldn’t want to give it up.
- ○ disagree
- ○ disagree somewhat
- ○ agree somewhat
- ○ agree

7) Mathematics is important to me personally.
- ○ disagree
- ○ disagree somewhat
- ○ agree somewhat
- ○ agree

8) It is important to me to be good at mathematics.
- ○ disagree
- ○ disagree somewhat
- ○ agree somewhat
- ○ agree

A.8 German Version of the Mathematics Speed Test

Speed-Test

1) Jede quadratische Funktion hat mindestens eine Nullstelle.
   - Wahr
   - Falsch

2) Jede lineare Funktion hat genau eine Nullstelle.
   - Wahr
   - Falsch

3) Eine lineare Funktion ist immer monoton steigend.
   - Wahr
   - Falsch

4) Es gibt Funktionen, bei denen unterschiedlichen Inputs der gleiche Output zugeordnet wird.
   - Wahr
   - Falsch

5) Die Zahl -7 hat unter der Funktion $f(x) = 3x - 7$ genau ein Urbild.
   - Wahr
   - Falsch

6) Der Graph der Funktion $f(x) = 2x + 3x^3 - 5x^9$ ist punktsymmetrisch bezüglich dem Ursprung.
   - Wahr
   - Falsch

7) Es gibt quadratische Funktionen, die negativ sind.
   - Wahr
   - Falsch

8) Es gibt Funktionen der Art $y = f(x)$, die zwei unterschiedliche y-Achsenabschnitte besitzen.
   - Wahr
   - Falsch

9) Es gibt eine Funktion, für welche $f(-1) = -\frac{1}{2}$ und $f(1) = \frac{1}{2}$ gilt, und welche zwischen -1 und 1 keine Nullstelle besitzt.
   - Wahr
   - Falsch

10) Es gibt Funktionen, die genau 7 Nullstellen besitzen.
    - Wahr
    - Falsch

11) Eine Funktion, deren Graph durch den Ursprung geht, besitzt immer mindestens eine Nullstelle.
    - Wahr
    - Falsch

12) Es gibt eine Funktion der Art $y = f(x)$, deren Graph durch den Ursprung geht und welche den y-Achsenabschnitt -4 hat.
   - Wahr
   - Falsch

13) Es ist möglich, dass zwei lineare Funktionen den gleichen y-Achsenabschnitt haben, aber nicht die gleiche Steigung.
   - Wahr
   - Falsch

14) Es ist möglich, dass zwei lineare Funktionen, deren Graphen parallel verlaufen, nicht die gleiche Steigung haben.
   - Wahr
   - Falsch

15) Es ist möglich, dass sich die Graphen von zwei linearen Funktionen in genau zwei Punkten schneiden.
   - Wahr
   - Falsch

16) Es ist möglich, dass sich der Graph einer linearen Funktion und der Graph einer quadratischen Funktion in zwei Punkten schneiden.
   - Wahr
   - Falsch

17) Es ist möglich, dass sich die Graphen einer linearen und einer quadratischen Funktion nicht schneiden.
   - Wahr
   - Falsch

18) Eine Gerade, die auf 4 Einheiten in x-Richtung 20 Einheiten in y-Richtung ansteigt, hat eine Steigung von $1/5$.
   - Wahr
   - Falsch

19) Eine lineare Funktion, deren Graph durch die Punkte $A(-2, 2)$ und $B(2,2)$ geht, hat Steigung 0.
   - Wahr
   - Falsch

20) Es ist möglich, dass eine lineare Funktion, deren Graph durch den Punkt $E(-1,7)$ geht, auch durch den Punkt $F(2, -7)$ geht.
   - Wahr
   - Falsch

21) Es ist möglich, dass eine lineare Funktion, deren Graph durch den Punkt $E(-1, 7)$ geht, auch durch den Punkt $F(-1, -7)$ geht.
- Wahr
- Falsch

22) Der Graph einer linearen Funktion geht durch die Punkte $A(2,4)$ und $B(4,8)$. Dann liegt der Punkt $C(3,6)$ auch auf diesem Graphen.
- Wahr
- Falsch

23) Die Graphen der linearen Funktionen $f(x) = 2x - 1$ und $g(x) = \frac{1}{2}x + 2$ sind orthogonal.
- Wahr
- Falsch

24) Der Graph der Funktion $f(x) = \frac{3}{2}x - 3$ verläuft überall oberhalb vom Graphen der Funktion $g(x) = \frac{3}{2}x - 1$.
- Wahr
- Falsch

25) Die Funktion $f(x) = -\frac{x+4}{2}$ ist linear.
- Wahr
- Falsch

26) Die Funktion $g(x) = -4(2 + 3x)$ ist linear.
- Wahr
- Falsch

27) Die folgende Datenliste enthält einige Messwerte, die in gleichen Zeitabständen erhoben worden sind: $U = \{1,4,9,16,25\}$. In diesem Fall liegt lineares Wachstum vor.
- Wahr
- Falsch

28) Wenn $f(x)$ eine lineare Funktion ist, dann ist $g(x) = 4 \cdot f(x) + f(x)$ auch eine lineare Funktion.
- Wahr
- Falsch

29) Der Graph der Funktion $f(x) = -(x - 5)(1 + x)$ ist nach oben geöffnet.
- Wahr
- Falsch

30) Die Funktion $g(x) = x^2 + 2$ hat genau zwei Nullstellen.
- Wahr
- Falsch

31) Alle Funktionswerte der Funktion $g(x) = -2(x+2)^2 + 2$ sind kleiner oder gleich 2.
- Wahr
- Falsch

32) Der kleinste Funktionswert der Funktion $f(x) = -3(x-5)^2 - 7$ ist -7.
- Wahr
- Falsch

33) Der Graph der Funktion $g(x) = -x^2 + 8$ ist achsensymmetrisch bezüglich der y-Achse.
- Wahr
- Falsch

34) Es gibt quadratische Funktionen, welche 3 Nullstellen besitzen.
- Wahr
- Falsch

35) Der Funktionsgraph von $f(x) = -(x+3)(x-5)x + 7$ verläuft überall oberhalb vom Graphen der Funktion $g(x) = -(x+3)(x-5)x + 6$.
- Wahr
- Falsch

36) Der Funktionsgraph von $g(x) = (x-2)^3 - (x-2)^2 + 1$ ist gegenüber dem Funktionsgraphen von $h(x) = x^3 - x^2 + 1$ um 2 Einheiten in x-Richtung verschoben.
- Wahr
- Falsch

37) Den Graphen der Funktion $g(x) = x^2 + 5$ erhält man, indem man die Normalparabel um 5 Einheiten in x-Richtung verschiebt.
- Wahr
- Falsch

38) Die Funktion $f(x) = 5x^8 - \sqrt{7}x + \frac{\pi}{2}$ ist eine Polynomfunktion.
- Wahr
- Falsch

39) Es gibt quadratische Funktionen, deren Graphen keinen Schnittpunkt mit der x-Achse haben.
- Wahr
- Falsch

40) Lineare Funktionen besitzen immer eine Umkehrfunktion.
- Wahr
- Falsch

41) Die Betragsfunktion mit $\mathbb{D} = \mathbb{R}$ ist nicht injektiv.
- Wahr
- Falsch

42) Die Signumsfunktion mit $\mathbb{D} = \mathbb{R}$ besitzt eine Umkehrfunktion.

- Wahr
- Falsch

43) Die Funktion $g(x) = -5(x+1)(x-3)$ mit $\mathbb{D} = \mathbb{R}$ ist bijektiv.

- Wahr
- Falsch

44) Die Funktion $h(x) = 3x^3$ mit $\mathbb{D} = \mathbb{R}$ ist bijektiv.

- Wahr
- Falsch

45) Es gibt eine Funktion, die ihre eigene Umkehrfunktion ist.

- Wahr
- Falsch

A.9 English Version of the Mathematics Speed Test

Speed Test

1) Every quadratic function has at least one root.
   - True
   - False

2) Every linear function has exactly one root.
   - True
   - False

3) A linear function is always monotonically increasing.
   - True
   - False

4) There are functions, for which different inputs yield the same output.
   - True
   - False

5) With respect to the function $f(x) = 3x - 7$, the number -7 has exactly one preimage.
   - True
   - False

6) The graph of the function $f(x) = 2x + 3x^3 - 5x^9$ is point symmetrical with respect to the origin.
   - True
   - False

7) There are negative quadratic functions.
   - True
   - False

8) There are functions of the form $y = f(x)$, which have two different y-intercepts.
   - True
   - False

9) There is a function, for which $f(-1) = -\frac{1}{2}$ and $f(1) = \frac{1}{2}$, and which does not have any roots between -1 and 1.
   - True
   - False

10) There are functions with exactly 7 roots.
    - True
    - False

11) A function whose graph passes through the origin always has at least one root.
    - True
    - False

12) There is a function of the form $y = f(x)$ whose graph passes through the origin and which has the y –intercept -4.
   - True
   - False

13) It is possible that two linear functions have the same y-intercept but not the same slope.
   - True
   - False

14) It is possible that two linear functions with parallel graphs do not have the same slopes.
   - True
   - False

15) It is possible that the graphs of two linear functions intersect in exactly two points.
   - True
   - False

16) It is possible that the graph of a linear function and the graph of a quadratic function intersect in two points.
   - True
   - False

17) It is possible that the graph of a linear function and the graph of a quadratic function do not intersect.
   - True
   - False

18) A straight line, which increases 20 units in the y-direction over 4 units in the x-direction, has a slope of $1/5$.
   - True
   - False

19) A linear function whose graph passes through the points $A(-2, 2)$ and $B(2,2)$ has slope 0.
   - True
   - False

20) It is possible that a linear function whose graph passes through the point $E(-1,7)$ also passes through the point $F(2, -7)$.
   - True
   - False

21) It is possible that a linear function whose graph passes through the point $E(-1, 7)$ also passes through the point $F(-1, -7)$.
   - True
   - False

22) If the graph of a linear function passes through the points $A(2,4)$ and $B(4,8)$ then the point $C(3,6)$ also lies on this graph.
- o True
- o False

23) The graphs of the linear functions $f(x) = 2x - 1$ and $g(x) = \frac{1}{2}x + 2$ are orthogonal.
- o True
- o False

24) The graph of the function $f(x) = \frac{3}{2}x - 3$ lies above the graph of the function $g(x) = \frac{3}{2}x - 1$ in every point.
- o True
- o False

25) The function $f(x) = -\frac{x+4}{2}$ is linear.
- o True
- o False

26) The function $g(x) = -4(2 + 3x)$ is linear.
- o True
- o False

27) The following data set contains some measured values that were collected during equal time intervals: $U = \{1,4,9,16,25\}$. The data suggests linear growth.
- o True
- o False

28) If $f(x)$ is a linear function, then $g(x) = 4 \cdot f(x) + f(x)$ is a linear function as well.
- o True
- o False

29) The graph of the function $f(x) = -(x - 5)(1 + x)$ is opened upwards.
- o True
- o False

30) The function $g(x) = x^2 + 2$ has exactly two roots.
- o True
- o False

31) All values of the function $g(x) = -2(x + 2)^2 + 2$ are less than or equal to 2.
- o True
- o False

32) The smallest value of the function $f(x) = -3(x - 5)^2 - 7$ is -7.
- o True
- o False

33) The graph of the function $g(x) = -x^2 + 8$ is axially symmetric with respect to the y-axis.
- True
- False

34) There are quadratic functions, which have 3 roots.
- True
- False

35) The graph of the function $f(x) = -(x+3)(x-5)x + 7$ lies above the graph of the function $g(x) = -(x+3)(x-5)x + 6$ in every point.
- True
- False

36) The graph of the function $g(x) = (x-2)^3 - (x-2)^2 + 1$ is shifted by 2 units in x-direction compared to the graph of the function $h(x) = x^3 - x^2 + 1$.
- True
- False

37) The graph of the function $g(x) = x^2 + 5$ can be obtained by shifting the unit parabola by 5 units in x-direction.
- True
- False

38) The function $f(x) = 5x^8 - \sqrt{7}x + \frac{\pi}{2}$ is a polynomial function.
- True
- False

39) There are quadratic functions whose graphs do not have any intersection points with the x-axis.
- True
- False

40) Linear functions always have an inverse function.
- True
- False

41) The absolute value function with $\mathbb{D} = \mathbb{R}$ is not injective.
- True
- False

42) The sign function with $\mathbb{D} = \mathbb{R}$ has an inverse function.
- True
- False

43) The function $g(x) = -5(x+1)(x-3)$ with $\mathbb{D} = \mathbb{R}$ is bijective.
- True
- False

44) The function $h(x) = 3x^3$ with $\mathbb{D} = \mathbb{R}$ is bijective.

- True
- False

45) There is a function, which is its own inverse function.

- True
- False

A.10 German Version of the Mathematics Power Test

Power-Test

1) Betrachten Sie die lineare Funktion $f(x) = -3x + 4$.
   a) Der Graph der linearen Funktion g verläuft parallel zum Funktionsgraphen von f. Zudem besitzt g den y-Achsenabschnitt -1. An welchem Punkt schneidet der Funktionsgraph von g die x-Achse?
   b) Der Graph der linearen Funktion h ist orthogonal zum Graphen der Funktion f. Zudem besitzt h die Nullstelle $x = 6$. An welchem Punkt schneidet der Funktionsgraph von h die y-Achse?
   c) Der Graph der linearen Funktion k schneidet den Funktionsgraphen von f in keinem einzigen Punkt. Der Funktionsgraph von k geht durch den Punkt $K(2,\ 2)$. Wie lauten die Nullstelle und der y-Achsenabschnitt von k?

2) Betrachten Sie die linearen Funktionen $f(x) = 2x + 2$, $g(x) = -3x + 1$ und $h(x) = -x - 8$.
   a) Wie müsste die Steigung von f angepasst werden, damit der Funktionsgraph von f neu durch den Punkt $A(4, 14)$ gehen würde (unter der Annahme, dass der y-Achsenabschnitt von f gleich bleibt)?
   b) Wie müsste der y-Achsenabschnitt von g verändert werden, damit g neu die Nullstelle $x = 3$ besitzen würde (unter der Annahme, dass die Steigung von g gleich bleibt)?
   c) Um wie viele Einheiten in y – Richtung müsste der Funktionsgraph von h verschoben werden, damit h neu die Nullstelle $x = 5$ besitzen würde?

3) Betrachten Sie die Funktionen $f(x) = 2x^2 - 16x + 24$, $h(x) = -\frac{1}{2}x - 6$ und $k(x) = x + 8$.
   a) Wie lautet die Funktionsgleichung der Funktion g, deren Graph gegenüber dem Graphen von f an der x-Achse gespiegelt ist?
   b) Um wie viele Einheiten in x – Richtung müsste der Funktionsgraph von f verschoben werden, damit f neu den Scheitelpunkt $S(8,\ -8)$ besitzen würde? Wie lautet die Funktionsgleichung des verschobenen Graphen?
   c) Wie müsste der y-Achsenabschnitt von k verändert werden, damit die Funktionsgraphen von h und k sich neu im Punkt $B(-4,\ -4)$ schneiden würden (unter der Annahme, dass die Steigung von k gleich bleibt)?

Bitte blättern Sie um!

4) Betrachten Sie die linearen Funktionen $f(x) = -2x + 3$ und $g(x) = 8x + 17$.
   a) Der Graph der linearen Funktion h ist orthogonal zum Funktionsgraphen von f. Zudem besitzt h die Nullstelle $x = 4$. Wie lautet der Schnittpunkt der Funktionsgraphen von f und h?
   b) Der Graph der linearen Funktion k ist parallel zur x-Achse und der Funktionsgraph von k schneidet den Funktionsgraphen von f im Punkt $B(1, 1)$. Wie lautet der Schnittpunkt der Funktionsgraphen von k und g?
   c) Die quadratische Funktion q besitzt den Scheitelpunkt $B(1, 1)$. Einer der beiden Schnittpunkte der Funktionsgraphen von f und q ist der Punkt $P(3, -3)$. Wie lautet die Funktionsgleichung von q?

5) Die quadratische Funktion f besitzt die Nullstellen $x = 1$ und $x = 5$ sowie den y-Achsenabschnitt 5. Der Graph der Funktion f ist symmetrisch bezüglich der vertikalen Achse $x = 3$. Über dem Intervall $[5\,, 10]$ ist f streng monoton steigend, während f über dem Intervall $[-1\,,1]$ streng monoton fallend ist. Der Graph der quadratischen Funktion h ist gegenüber dem Graphen der Funktion f am Ursprung gespiegelt. Die Funktion h ist über dem Intervall $[-1\,,1]$ streng monoton fallend. Zudem ist der Graph der Funktion h symmetrisch bezüglich der vertikalen Achse $x = -3$.
   a) Bestimmen Sie den Scheitelpunkt von f.
   b) Bestimmen Sie die Funktionsgleichung von h.
   c) Betrachten Sie die Funktion $i(x) = -4x + p$. Wie muss der Parameter p gewählt werden, damit der Funktionsgraph von i im Punkt $I(-1\,, 0)$ eine Tangente an den Funktionsgraphen von h ist?

6) Die Graphen der linearen Funktionen f und g sind orthogonal zueinander und schneiden sich in genau einem Punkt, nämlich dem Punkt $P(1, 3)$. Die Funktion f ist streng monoton steigend, während die Funktion g streng monoton fallend ist. Die Funktion f besitzt bei $x = \frac{1}{4}$ eine Nullstelle und ist über dem Intervall $[1{,}10]$ positiv. Die Umkehrfunktion der Funktion g besitzt den y-Achsenabschnitt 13. Der Graph der linearen Funktion k verläuft parallel zum Graphen der Funktion f und schneidet die x-Achse bei $x = -1$. Die Funktion k besitzt einen positiven y-Achsenabschnitt und ihr Graph schliesst gemeinsam mit der x-Achse und der y-Achse ein Dreieck der Fläche 2 ein.
   a) Die Funktionsgraphen von f und g schliessen gemeinsam mit der x-Achse ein Dreieck ein. Wie gross ist die Fläche dieses Dreiecks?
   b) Bestimmen Sie die Funktionsgleichung von k.
   c) Bestimmen Sie den Schnittpunkt des Funktionsgraphen von f mit dem Funktionsgraphen der Umkehrfunktion von f.

A.11 English Version of the Mathematics Power Test

Power Test

1) Consider the linear function $f(x) = -3x + 4$.
   a) The graph of the linear function g is parallel to the graph of the function f. Moreover, g has the y-intercept -1. Where is the intersection point between the graph of g and the x-axis?
   b) The graph of the linear function h is orthogonal to the graph of the function f. Moreover, h has a root at $x = 6$. Where is the intersection point between the graph of h and the y-axis?
   c) The graph of the linear function k does not intersect the graph of the function f in any point. The graph of k passes through the point $K(2,\ 2)$. Which root and which y-intercept does k have?

2) Consider the linear functions $f(x) = 2x + 2$, $g(x) = -3x + 1$ and $h(x) = -x - 8$.
   a) How should the slope of f be adjusted in order for the graph of f to pass through the point $A(4, 14)$ (assuming that the y-intercept of f remained the same)?
   b) How should the y-intercept of g be changed in order for g to have the root $x = 3$ (assuming that the slope of g remained the same)?
   c) By how many units in y-direction should the graph of the function h be shifted, in order for h to have the root $x = 5$?

3) Consider the functions $f(x) = 2x^2 - 16x + 24$, $h(x) = -\frac{1}{2}x - 6$ and $k(x) = x + 8$.
   a) What is the functional equation of the function g, whose graph is the reflection of the graph of f along the x-axis?
   b) By how many units in the x-direction should the graph of f be shifted in order for f to have the angular point $S(8,\ -8)$? What is the functional equation of the shifted graph?
   c) How should the y-intercept of k be adjusted in order for the graphs of the functions h and k to intersect at the point $B(-4,\ -4)$ (assuming that the slope of k remained the same)?

Please turn the page!

4) Consider the linear functions $f(x) = -2x + 3$ and $g(x) = 8x + 17$.
   a) The graph of the linear function h is orthogonal to the graph of the function f. Moreover, h has a root at $x = 4$. Where do the graphs of f and h intersect?
   b) The graph of the linear function k is parallel to the x-axis and it intersects the graph of f at the point $B(1, 1)$. Where is the intersection point between the graphs of k and g?
   c) The quadratic function q has the angular point $B(1, 1)$. One of the two intersection points between the graphs of f and q is the point $P(3, -3)$. What is the functional equation of q?

5) The quadratic function f has roots at $x = 1$ and $x = 5$ as well as the y-intercept 5. The graph of the function f is symmetric with respect to the vertical axis $x = 3$. Over the interval $[5\,, 10]$ f is strictly monotonically increasing, while over the interval $[-1\,,1]$ it is strictly monotonically decreasing. The graph of the quadratic function h is point symmetric to the graph of the function f with respect to the origin. The function h is strictly monotonically decreasing over the interval $[-1\,,1]$. Moreover, the graph of the function h is symmetric with respect to the vertical axis $x = -3$.
   a) Determine the angular point of f.
   b) Determine the functional equation of h .
   c) Consider the function $i(x) = -4x + p$. How should the parameter p be chosen in order for the graph of the function i to be tangent to the graph of the function h at the point $I(-1\,, 0)$?

6) The graphs of the linear functions f and g are orthogonal to each other and intersect at exactly one point, namely at the point $P(1, 3)$. The function f is strictly monotonically increasing, while the function g is strictly monotonically decreasing. The function f has a root at $x = \frac{1}{4}$ and is positive over the interval $[1,10]$. The inverse function of the function g has the y-intercept 13. The graph of the linear function k is parallel to the graph of the function f and intersects the x-axis at $x = -1$. The function k has a positive y-intercept and, together with the x-axis and the y-axis, its graph encloses a triangle of area 2.
   a) The graphs of the functions f and g enclose a triangle together with the x-axis. Determine the area of this triangle.
   b) Determine the functional equation of k.
   c) Determine the intersection point between the graph of the function f and the graph of the inverse function of f.

A.12 German Version of the List of Mathematical Terms for the Preparatory Task

Liste mit wichtigen Begriffen zur Vorbereitung auf den Testtag

- achsensymmetrisch
- Betragsfunktion
- bijektiv
- Definitionsbereich $\mathbb{D}$
- Dreieck
- Einheiten
- Fläche
- Funktionsgleichung
- Funktionsgraph
- Funktionswert
- Gerade
- gleichschenkliges Dreieck
- injektiv
- Intervall
- linear
- lineare Funktion
- negativ
- Normalparabel
- Nullstelle
- orthogonal
- parallel
- Parameter
- Polynomfunktion
- positiv
- punktsymmetrisch
- Quadrant
- quadratisch
- quadratische Funktion
- Scheitelpunkt
- Schnittpunkte von Funktionsgraphen
- Signumsfunktion
- Spiegelung
- Steigung
- (streng) monoton fallend
- (streng) monoton steigend
- surjektiv
- Tangente
- Umkehrfunktion
- Urbild
- Ursprung
- Winkelhalbierende
- x-Achse
- y-Achse
- y-Achsenabschnitt

A.13 English Version of the List of Mathematical Terms for the Preparatory Task

Glossar für Immersionsklassen

Deutscher Begriff	Englischer Begriff
achsensymmetrisch	axially symmetric
Betragsfunktion	absolute value function
bijektiv	bijective
Definitionsbereich $\mathbb{D}$	domain $\mathbb{D}$
Dreieck	triangle
Einheiten	units
Fläche	area
Funktionsgleichung	functional equation
Funktionsgraph	graph of a function
Funktionswert	value of the function
Gerade	straight line
gleichschenkliges Dreieck	isosceles triangle
injektiv	injective
Intervall	interval
linear	linear
lineare Funktion	linear function
negativ	negative
Normalparabel	unit parabola
Nullstelle	root of a function
orthogonal	orthogonal
parallel	parallel
Parameter	parameter
Polynomfunktion	polynomial function
positiv	positive
punktsymmetrisch	point symmetric
Quadrant	quadrant
quadratisch	quadratic
quadratische Funktion	quadratic function
Scheitelpunkt	angular point / vertex
Schnittpunkte von Funktionsgraphen	intersection points of functions
Signumsfunktion	sign function
Spiegelung	reflection
Steigung	slope
(streng) monoton fallend	(strictly) monotonically decreasing
(streng) monoton steigend	(strictly) monotonically increasing
surjektiv	surjective
Tangente	tangent
Umkehrfunktion	inverse function
Urbild	preimage
Ursprung	origin
Winkelhalbierende	bisecting line of an angle
x-Achse	x- axis
y-Achse	y- axis
y-Achsenabschnitt	y- intercept