Maastricht University

An evidence-based approach to working-memory based training in secondary education to improve reasoning test achievements

Citation for published version (APA):

Ariës, R. J. F. (2018). *An evidence-based approach to working-memory based training in secondary education to improve reasoning test achievements* (no. XII ed.). [Doctoral Thesis, Maastricht University]. TIER. https://doi.org/10.26481/dis.20180517ra

Document status and date:
Published: 01/01/2018

DOI:
10.26481/dis.20180517ra

Document Version:
Publisher's PDF, also known as Version of record

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 25 Apr. 2024

An evidence-based approach to working-memory-based training in secondary education to improve reasoning test achievements

Roel J.F.J. Ariës

© Roel J.F.J. Ariës, Maastricht 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing from the author.

ISBN 978-94-003-0137-5

Cover design: Raadhuis voor creatieve communicatie, Alkmaar

This book is no. **XII** of the TIER Research Series, a PhD thesis series published by TIER.

Acknowledgements

This work was supported by the Netherlands Organisation for Scientific Research (NWO) under the Doctoral Grant for Teachers.

AN EVIDENCE-BASED APPROACH TO WORKING-MEMORY-BASED TRAINING IN SECONDARY EDUCATION TO IMPROVE REASONING TEST ACHIEVEMENTS

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Universiteit Maastricht
op gezag van de Rector Magnificus,
Prof. Dr. Rianne M. Letschert
volgens het besluit van het College van Decanen

in het openbaar te verdedigen

op donderdag 17 mei 2018 om 16.00 uur

door

Roel Jacobus Florence Jozef Ariës
Geboren te Maastricht, Nederland

Promotores:
Prof. Dr. W.N.J. Groot
Prof. Dr. H. Maassen van den Brink

Co-promotor:
Dr. J. Ghysels

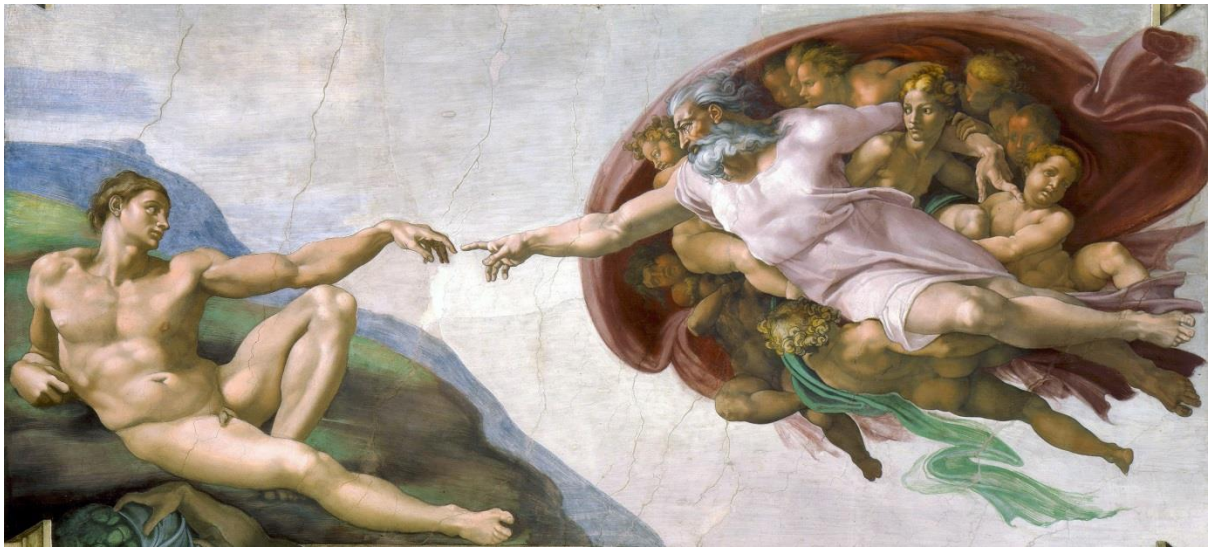
Beoordelingscommissie:
Prof. Dr. Kristof De Witte (voorzitter)
Prof. Dr. Bernadette Jansma
Prof. Dr. Paul Kirschner (Open Universiteit)
Prof. Dr. Lydia Krabbendam (Vrije Universiteit)
Prof. Dr. Jeroen van Merriënboer

Dedication

For Pip and Lou

Wie an dem Tag, der dich der Welt verliehen,
Die Sonne stand zum Gruße der Planeten,
Bist alsobald und fort und fort gediehen,
Nach dem Gesetz, wonach du angetreten.
So mußt du sein, dir kannst du nicht entfliehen,
So sagten schon Sibyllen, so Propheten;
Und keine Zeit und keine Macht zerstückelt
Geprägte Form, die lebend sich entwickelt.

Johan Wolfgang Goethe



Preface

In contemporary history, the frontal lobes have been accredited with the signature of divinity, interconnected to complex thinking, and thus reasoning. Julian Jaynes (2000), a former teacher of psychology at Princeton University, suggested that primitive humans once considered internally generated rational ideas as being the whispers of the gods. This implies that the origins of coordinating functions in the early stages of human civilization could have resulted in the formation of religious convictions. Some art historians even suggest that the early awareness of the importance of the frontal lobes is exemplified in Michelangelo's 16th century fresco 'The creation of Adam'. In this fresco, God's cloak is thought to portray the outline of the human brain, his feet rest on the brainstem while his head is portrayed in the frontal lobes (Meschberger, 1990). God's finger, which points to Adam and makes of him a conscious human being, emerges from the prefrontal cortex. One can not imagine a more powerful symbol that portrays the comprehensive effect of the frontal lobes on the existence of humanity more vividly (Goldberg, 2010, p.40).

In modern history, empirical reasoning has become increasingly valuable, running in parallel with the rise of modern science as being

'infallible' (Spinoza, 1677/2002 p.83). The definitive triumph of empirical reasoning became apparent when Kant, contrary to Hume who states that the existence of causality can not be proven, proposed that one can reason because man's mental structure consists of a set of *a priori* concepts (i.e. categories) through which he interprets the world, one category being *causality and dependence* (Kant, 1781/2004). And reasoning became such an important part of Man's general and skills development, that Arthur Schopenhauer (1804-1818/1988) for example, even metaphorically argued in the early 18th century that reasoning – and not factual knowledge - is the only key to find universal truths: 'One person tells a lie; another who knows the truth says this is a pack of lies and here you have the truth. A third person who does not *know* the truth, but is very acute and clever, points out the contradictions and impossible statements in that lie and says that it is therefore fraud and falsehood.' And thus, the irrefragable relationship between searching for truth and systematic thinking, which forms the basis of Western progress, is connected through the prefrontal cortex which is held responsible for these complex activities.

Indeed, systematic thinking processes depend strongly on the prefrontal lobe functions. The introduction of a rational method when problems have to be solved triggers the prefrontal cortex. And this activity increases when problems become more complex, and various components of a problem have to be connected to each other to form a solution to the problem. And, as such, we enter contemporary times in which scientists are increasingly convinced that the brain, and thus complex thinking, can be trained. Over the past two decades there has been much greater understanding of the role of working memory in reasoning, the development of working memory and, in particular, how working memory can be trained to adapt to reasoning demands more effectively. We still do not know much about the interconnectivity of the brain regions, in which working memory resides, and the possibilities and limitations that may result from these individual mechanisms. We also

still do not know much about reproducible training exercises to develop reasoning more effectively through working-memory training. A few training exercises have resulted in positive effects, but evidence of a clear transfer of these general reasoning skills to content-based domain-specific reasoning is lacking.

The majority of training effects are measured in laboratory settings. But what might result from working-memory training in classroom settings? This thesis addresses this last question by focusing on humanities reasoning training in secondary education.

Frequently-used terms

Cognitive load	According to cognitive load theory (CLT), the human cognitive system has a limited working-memory (WM)-capacity which is able to deal with novel information that is obtained through sensory memory. WM has no known limitations when it deals with information that is retrieved from long-term memory (LTM) which, in its turn, alters the characteristics of WM.
Executive Functions	Ef are a set of cognitive processes, and are conceptualized as having four components: the abilities of goal formation; planning; carrying out a goal-directed plan; and effective performance (Jurado and Rosselli, 2007)
Far transfer	Performance improvements in tests that are dissimilar to the trained task.
Fluid intelligence	The ability to reason and to solve new problems independently of previous acquired knowledge.
Humanities	Academic disciplines that study human culture, for example history and social studies.
Long-term memory	Holds cognitive schemas that vary in their degree of complexity and automation. Human expertise comes from knowledge organized by these schemas.
Meta-cognition	Awareness and management of one's own thought, or 'thinking about thinking'. One can learn how to think critically by reflecting on critical thinking which makes it increasingly more effective. WM involves executing meta-cognitive processes. A meta-cognitive process can serve as a long-term learning strategy to develop reasoning skills, because the awareness of the relevance of learning activities enhances the long-term memorization and re-use of these learning activities.

N-back	The n-back task involves the serial representation of a stimulus, spaced several seconds apart. The participant must decide whether the current stimulus matches the one displayed *n* trials ago, where *n* is a variable number that can be adjusted up or down to, respectively, increase or decrease cognitive load. In the context of WM-training, efforts have focused on flexibly adapting the task difficulty in accordance with the participants' fluctuating performance level by increasing or decreasing the level of *n*. The idea is to keep the participants' WM-system perpetually engaged at its limit, thereby stimulating an increase in WM-function.
Near transfer	Performance improvements in tests that are similar, but not identical, to the trained task.
Odd One Out	This task tests the storage capacity and manipulation of information in the WM because words or pictures have to be stored and processed simultaneously in the WM. During the experiments in this dissertation, four historical words or pictures were successively shown on a video screen in a PowerPoint presentation. Each word or picture was shown for 3 seconds. Three words or pictures in each series were related to each other, and one word or picture was the 'odd one out'. Participants had to remember all words or pictures within a sequence. After all words or pictures had been shown, the participants had to write down the word or picture that was the odd one out on an answer form.
Short-term memory	The ability to retain information during short periods of time.
Working memory	A brain system that provides short-term storage and manipulation of information.

List of abbreviations

CLT Cognitive load theory

ef Executive functions

Gf Fluid intelligence

KIT Assessment model. KIT stands for Knowledge, Comprehension, Application questions [Dutch: Kennis, Inzicht, Toepassing]

LTM Long-term memory

RTTI Assessment model. RTTI stands for Reproduction, Application from training, Application by transfer, Comprehension questions [Dutch: Reproductie, Toepassing 1, Toepassing 2, Inzicht]

STM Short-term memory

WM Working memory

Contents

Preface vi

Frequently-used terms ix

List of abbreviations xi

List of Tables xv

List of Figures xvii

1 General Introduction 1

1.1 Background 4

1.2 Aims, outline and research questions 16

2 Does n-back training in a school-based context improve reasoning? A literature review 21

Summary 21

2.1 Introduction 22

2.2 Method and search strategy 27

2.3 Results 33

2.4 Discussion and conclusion 40

3 Is working-memory capacity and metacognitive training effective in enhancing school-based reasoning achievements? A synthesis of the research 47

Summary 47

3.1 Introduction 48

3.2 Method and search strategy 51

3.3 Results 53

3.4 Discussion and conclusion 64

4 Improving reasoning skills in secondary history education by working-memory training 72

Summary 72

4.1 Introduction 73
4.2 Experiment 1: Complete training and durable effects 75
4.3 Experiment 2: The effects of the components of working-memory training on short-term test results 86
4.4 Discussion and conclusion 92
5 Combined working-memory capacity and reasoning-strategy training improves reasoning skills in secondary social studies education: evidence from an experimental study 97
Summary 97
5.1 Introduction 98
5.2 Methods 103
5.3 Results 110
5.4 Discussion and conclusion 121
6 Improving reasoning abilities in secondary history education: The effect of motivational components in combined reasoning-strategy and working-memory training 127
Summary 127
6.1 Introduction 128
6.2 Methods 132
6.3 Results 142
6.4 Discussion and Conclusion 149
7 Conclusions and Discussion 154
7.1 Key Conclusions 155
7.2 Methodological issues 163
7.3 Study limitations and future research 165
7.4 Policy implications 169
Bibliography 173
Appendix A 187

Appendix B.....188
Appendix C.....189
Appendix D 190
Appendix E 191
Summary.....195
Curriculum Vitae 208
List of publications in peer-reviewed journals.....210
Acknowledgements.....211

List of Tables

Table 2.1 Redick et al. (2015). 'Criteria for strong evidence of working memory training efficacy in educationally relevant context'. 30
Table 2.2 Evaluation of the criteria per intervention and far transfer results on reasoning. 32
Table 2.3 Overview of interventions assessing the impact of n-back training on reasoning. 34
Table 2.4 Significant effects of far transfer to reasoning: additional categories. 35

Table 3.1 Overview of studies assessing the impact of WM-capacity training on school reasoning skills. 57
Table 3.2 Overview of studies assessing the impact of metacognitive training on school reasoning skills. 61

Table 4.1 Independent T-test results for pretest scores per group (Experiment 1). 84
Table 4.2A ANOVA results and descriptive statistics for cognition intermediate and posttests between groups (Experiment 1). 84
Table 4.2B ANOVA results and descriptive statistics for strategy intermediate and posttests between groups (Experiment 1). 84
Table 4.2C ANOVA results and descriptive statistics for achievements intermediate and posttests between groups (Experiment 1). 85
Table 4.3 ANOVA results for pretests between groups (Experiment 2)...89
Table 4.4 Multiple comparisons and mean differences in test scores by instruction type controlling for group (Experiment 2).....90

Table 5.1 Research design. 104
Table 5.2 Independent T-test results for pretest scores: for the complete sample and per group. 110
Table 5.3A Independent T-test results and descriptive statistics for cognition intermediate and posttests between groups. 111

Table 5.3B Independent T-test results and descriptive statistics for reasoning achievements intermediate and posttests between groups..112
Table 5.4A Summary of regression effects of independent variables on reasoning achievements posttest. 119
Table 5.4B Summary of regression effects of independent variables for Geleen groups. 120

Table 6.1 Research design. 132
Table 6.2 Descriptive statistics and t-tests for equality in expectation per group. 133
Table 6.3 Structure of tests and training. 141
Table 6.4 Independent samples T-test results and descriptive statistics for cognition and reasoning pretests and motivation inquiry between experimental and control groups. 143
Table 6.5 Independent samples T-test results and descriptive statistics for cognition and reasoning posttests and motivation inquiry between experimental and control groups. 143
Table 6.6 Paired samples t-test results and descriptive statistics for cognition and reasoning pre- and posttests and motivation inquiry. 144
Table 6.7 Regression effects of independent variables on standardized reasoning achievements posttest. 146
Table 6.8 Regression effects of independent variables on GAINS in reasoning achievements pre- to posttest. 147

Table 7.1 Main contributions in terms of contents and chapter. 157

List of Figures

2.1. 'Two possible routes for improving deductive reasoning deductive reasoning by WM-training.' 24
2.2 Flowchart of the literature review process.....31
3.1 Flowchart of the literature review process.....52
4.1 Mean scores in Achievements tests (Experiment 1).....86
4.2 Mean scores in achievements tests (Experiment 2). 91
5.1 Mean scores in achievements tests by subgroup. 113

1 General Introduction

Secondary schools have - among other functions - the responsibility to teach students reasoning skills (OECD, 2008). Reasoning skills are needed to prepare students for the demands of the knowledge economy and the complexity of society in general. This has put the use of effective didactical models for teaching reasoning skills on the agenda. In line with education policies in OECD countries, humanities courses in compulsory school curricula have placed a major emphasis on teaching reasoning skills. In addition to civic, ethical, and social-justice literacy, reasoning skills are regarded as an essential aspect of the 21st century skills set, skills which new members of the work force have to acquire to meet the demands of the knowledge economy (Ananiadou and Claro, 2009). In line with this, the development of reasoning skills has become an important aspect of secondary school curricula. This is not entirely new. Even in the early 19th century, the Swiss pedagogue and educational reformer Johann Heinrich Pestalozzi proclaimed that students had to be given the opportunity to construct their own conclusions by following a progressive scheme of observation, intuitive thinking, conceptual thought, and visualization (Isaacson, 2015, p.46).

The reasons for the increased attention and importance attached to reasoning skills lie in the rise of the knowledge economy, in particular, and the increased complexity of society, in general. Accordingly, not only individual but also collective success depends on training these skills (Rotherham and Willingham, 2009). In fact, an OECD paper on 21st century learning states that 'educated workers need a conceptual understanding of complex concepts, and the ability to work with them creatively to generate new ideas, new theories, new products, and new knowledge' (2008, p.1). For many young people, secondary education provides the only opportunity to learn such higher-order skills (e.g. Ariës and Cabus, 2015). The importance of learning reasoning skills also becomes apparent in humanities school tests which increasingly contain reasoning questions. However, teachers regularly observe that adolescent students find it difficult to answer reasoning test questions

properly. This could be caused by an inability to: take into account alternative views; use sources extensively; acquire detailed factual knowledge and a broader frame of reference; judge the past by its own standards; take into account the process of continuity; and understand many substantive concepts - as is argued in the literature (Van Drie and Van Boxtel, 2008; Schug and Birkey, 1983; Thornton and Vukelich, 1988; Voss, Perkins and Segal (eds.), 1991, p.9). This provides the reason for developing an effective training method to improve reasoning skills and reasoning achievements in the humanities education of students in secondary education. The objective of this thesis is to evaluate the effectiveness of such training.

To address the concerns about the development of reasoning skills of adolescents, policymakers and schools are searching for new ways to improve adolescents' reasoning skills. One potentially highly effective strategy consists of learning based on cognitive and brain science (NIHC[1] and NRO[2], 2016). According to the recently published Dutch National Research Agenda on Brain, Cognition and Education (Crone, Verhoeven, Van Aken, Bekkering, van den Broek, 2016), cognitive neuroscience 'can make a significant contribution to […] research and to its successful implementation in educational practice' (p. 4). The education community realizes that this brain-informed approach to learning can help develop new and more effective strategies to improve students' reasoning skills. This belief is rooted in the assumption that neuroscientists have, by analyzing fMRI scans, begun to understand how the brain reasons. Consequently, standardized training batteries have been developed that can alter reasoning performance by focusing on particular areas of the brain that deal with reasoning, such as working memory (WM). But

[1] 'The National Initiative Brain and Cognition (NIHC) is a coordinating body within which specialists carry out scientific research on the brain and its influence on human behavior and society. NIHC is supported and coordinated by the Netherlands Organisation for Scientific Research (NWO)'.

[2] The Netherlands Initiative for Education Research (NRO) contributes to innovation and improvements in education by coordinating and funding education research.'

school-based research on WM-training to improve reasoning is still limited and recent. And school-based WM-training is not yet integrated in secondary education curricula. Consequently, evidence of improved reasoning in education, caused by WM-training, remains poor. In addition, there are some obstacles to the broad implementation of WM-training in secondary school curricula. There is much debate on, for instance, whether training on reasoning skills does actually result in improved reasoning skills in school-based test scores (e.g. Lehman and Nisbett, 1990; Au, Sheehan, Tsai, Duncan, Buschkuehl et al., 2015; Melby-Lervåg, Redick and Hulme, 2016). Furthermore, no school-based study has been conducted for humanities courses, such as history and social studies. There is also no evidence available on whether WM-training favors underachievers when performed in educational settings. And finally, due to the novelty of school-based WM-training, there are no standardized training batteries that specifically focus on humanities reasoning skills for teachers.

This thesis aims to increase our knowledge and insight into effective WM-based strategies that improve reasoning skills in secondary education humanities curricula. Insight into WM-based reasoning training in school-based settings could provide the necessary tools for both teachers and students to improve school-test results in humanities courses.

1.1 Background

Research that examines the role of WM-training to improve academic skills is increasing rapidly. But WM-training is mainly focused on general reasoning skills in laboratory settings. Furthermore, increasingly there are data available on school-based reading and mathematics training, but not in the humanities. In this chapter we describe the specific features and uniqueness of reasoning in secondary education humanities courses and how it can be trained and tested in content-based school tests. We then describe the development of WM concepts

and how WM in its different functions can contribute to reasoning skills in the secondary school context.

1.1.1 Reasoning skills in secondary education: humanities, tests and teaching methods, and working memory

Humanities courses in secondary education in the Netherlands consist of history, geography, economics, social studies, management and organization, and philosophy. Reasoning questions in secondary education humanities courses usually consist of deductive reasoning problems. The problem-solving skills development for these courses is described by the levels that have to be reached by all students in order to graduate. In this regard, SLO (the Centre of Expertise for Curriculum Development) published the National History Exam Program in which problem-solving skills are summarized. According to this document, students should be able to, for example, 'describe relationships between change and continuity in historical processes; identify how elements that come from different historical periods can manifest themselves simultaneously in a particular era by differentiating the continuities of shorter and longer duration (the simultaneousness of the non-simultaneousness); acquire useful sources and select the right information for response; support statements by using argumentation' (SLO, 2007, p.3).

The required problem-solving skills are generally comparable for all humanities courses. Problem-solving skills can be described as the 'process by which facts and themes are arranged to build an interpretative […] case' which then requires analysis, hypothesis, generation, and interpretation (Leinhardt, Stainton, Virj and Odoroffleds, 1994, p.314; Van Drie and Van Boxtel, 2008; College voor Examens, 2012; College voor Examens, 2013). These general problem-solving principles require domain-specific characteristics that are specific to the particular courses. For example, Van Drie and Van Boxtel (2008, p.104) state that 'historical reasoning requires general reasoning skills, but also contains several characteristics that are more specific to this particular domain.

Historical reasoning is not only informed by historical information, domain-specific knowledge, […] but also implies the application of historical heuristics or thinking strategies related to the meta-concepts of history'.

The development of skills is evaluated in both school tests and national exams in which all questions can be categorized in the following three categories: questions that require students to reproduce certain course specific knowledge (knowledge); questions that test a students' understanding in a certain kind of problem (comprehension); and questions that test whether students can solve new problems by using their content-based knowledge (application and analysis). The last two types of questions (reasoning questions) require problem-solving skills on top of content-based reproducible knowledge (College voor Examens, 2012; College voor Examens, 2013). Therefore, achievements on reasoning questions depend on comprehensive domain-specific knowledge, as well as on general and domain-specific thinking strategies. Accordingly, three factors are considered important for students to perform well in school tests in which reasoning questions are required, and all of these factors build upon domain-specific knowledge: (1) domain-specific knowledge of the subjects that are tested by answering reproduction questions and domain-specific reasoning questions; (2) skills to answer certain categories of problem questions on a given new problem, based on domain-specific knowledge; and (3) skills that include application and analyzing skills to solve new problems, based on domain-specific knowledge. These factors become apparent in the KIT (Knowledge, Comprehension, Application questions; Dutch: Kennis, Inzicht, Toepassing) and RTTI (Reproduction, Application from training, Application by transfer, Comprehension questions; Dutch: Reproductie, Toepassing 1, Toepassing 2, Inzicht) test-composition models which form the basis of the test-composition in Dutch secondary education.[3]

[3] E.g. http://www.malmberg.nl/Voortgezet-onderwijs/Methodes/Mens-Maatschappij/aardrijkskunde/KIT-kennis-inzicht-en-toepassing.htm ; http://rtti.nl/

The knowledge and skills tested in the exams are taught to students in all humanities courses. Teachers primarily use direct instruction, course-book questions and (un)official school-tests to train students to acquire problem-solving skills. In practice, teachers generally do not reflect on the reasoning process with (all of the) individual students on the questions. And they expect the skills to be practiced in other (humanities) courses as well. Students do assignments in their work-books during a class or at home. Afterwards, teachers discuss the assignments in the classroom and give the correct answers to students. By using assignments and school tests that contain reasoning questions, teachers expect that students will be trained to acquire problem-solving skills. However, these non-individual and non-self-reflecting training methods cannot ensure this, and there is no evidence of a positive correlation between the number of training years and the development of problem-solving skills by regularly training with these types of questions, other than what can be explained by maturation (Luna, Garver, Urban, Lazar and Sweeney, 2004). In the Netherlands usually 60%-70% of questions in national exams consist of the latter two types of questions, which makes training of problem-solving skills necessary for students who have not mastered reasoning questions.

Reorganizing and manipulating retained information is highly important to solve a reasoning question. Developing a problem-solving strategy and training in cognitive processing capacity can be primarily accredited to WM-functions (Fuster, 2003). According to the neuroscience literature, WM is the activation of cortical cognitive networks (i.e. cognits), with subcortical support, in order to selectively attend to recent information for a prospective decision, choice, or action to solve a problem or to reach a goal. It therefore plays a crucial role in developing problem-solving skills because it underlies several cognitive abilities, including logical reasoning and problem-solving. The prefrontal cortex which includes WM does not mature until late adolescence (Crone,

2009). Also, 'The cortical hubs of the network, including parietal and temporal regions, are all slow to mature'[4] and are sensitive to genetic and environmental factors (Bolger, Mackey, Wang and Grigorenko, 2014; Mackey, Finn, Leonard, Jacoby-Senghor, West et al., 2015). This could explain individual differences in answering reasoning questions in secondary education. In fact, some studies suggest a direct causal relation between the development of WM and school-test results (Klingberg, Forssberg and Westerberg, 2002; Olesen, Westerberg and Klingberg, 2004; St Clair-Thompson and Gathercole, 2006). A positive correlation between cognitive skills and test scores, consistent with studies relating WM to academic performance, has also been found in a study by Amy Finn et al. (2014). Subsequent analyses in this study suggested that: 'School practices that influence standardized achievement tests have limited effects on the development of cognitive skills associated with processing speed, working memory, or fluid reasoning' (Finn, Kraft, West, Leonard, Bish et al., 2014, p.742).

1.1.2 Working memory in the school perspective: definitions, theoretical frameworks, and adolescents

Since Baddeley and Hitch introduced and defined the standard WM model in 1974, there has been an ongoing debate about its definitions and mechanisms. In order to address the dynamics of WM-training to improve content-based reasoning, here we first describe the complexity of WM as a function that underlies reasoning. Consequently, the development of WM concepts and the way in which WM in its different functions can contribute to reasoning skills in the secondary school context are described. Furthermore, the deficits of specific adolescent reasoning skills and difficulties in the development of reasoning skills in school-based settings are explained.

[4] Allyson P. Mackey (2015, November 13), Personal Interview.

Baddeley and Hitch's (1974) standard WM model is usually defined as a brain system that underlies several cognitive abilities, such as reasoning or problem-solving (Klingberg et al., 2002). Evidence, however inconsistent, suggests that WM could be trained to improve reasoning (Klingberg et al., 2002; Olesen et al., 2004; St Clair-Thompson, Stevens, Hunt and Bolder, 2010; Alloway, 2012; Loosli, Buschkuehl, Perrig and Jaeggi, 2012; Shipstead, Redick and Engle, 2012; Melby-Lervåg and Hulme, 2013; Holmes and Gathercole, 2014; Titz and Karbach, 2014). WM is usually described as a system that provides short-term storage and manipulation of information. For instance, Olesen et al. (2004, p.75) address the relationship between WM and reasoning ability when describing WM as the 'ability to retain information during short periods of time. WM capacity is an important factor to determine problem-solving and reasoning ability'. Redick, Shipstead, Wiemers, Melby-Lervåg et al. (2015, p.617) use a broader concept by not defining WM as important for reasoning specifically, but referring to it as the 'system that is responsible for storage of information over short periods of time, whereby the information is used to fulfill some goal-directed activity'. And Holmes, Gathercole and Dunning (2009, p. F1) put WM in a narrower perspective by regarding it as 'the cognitive system that provides temporary storage of information in the course of complex cognitive activities and appears to play a crucial role both in supporting learning and maintaining focused behavior in practical situations'. In his book, 'Cortex and Mind' (2003, p.155), Joaquín Fuster defines WM as attention which is 'focused on the internal representation of a recent event for a pending action'. Klingberg et al. (2002, p.781) define WM capacity as 'the ability to retain and manipulate information during a short period of time'. WM as such, can be described as attention focused on an internal representation. In this view, attention has two components: (1) An inclusionary component consisting basically of the content in focus; and (2) an exclusionary component consisting mainly of the inhibitory control of interference. WM is demonstrably enhanced by concentration and weakened by distractibility. Reasoning ability behaves in a similar but

opposite way inasmuch as it depends on WM. Therefore, the success of addressing a problem-solving question depends on prioritization which could change depending on the context. But, at the same time, reasoning ability requires maintaining flexibility to observe the problem concerned from different perspectives in different stages of the problem-solving process.

New directions in WM-theory emphasize the interconnectivity between WM and other regions of the brain in order to understand *how* WM connects novel problems with information stored in long-term memory (LTM). For instance, Gazzaniga, Ivry, Mangun, and Steven (2009) describe WM as referring to 'transient representations of task-relevant information' which are 'activated to shape behavior in the present [and] requires integrating current perceptual information with stored knowledge (p. 559). WM, in this regard, is used to 'remember instructions, to keep in mind a plan of things to do, and to solve complex problems' (Klingberg, 2009, p.45). Consequently, WM could play an important role in acquiring context-independent procedural knowledge of problem-solving schemes. Context-independent strategies are strategies that tailor the reaction to a unique set of specifications in a given problem. But, the development of problem-solving strategies or schemas imposes a huge cognitive load on a person's cognitive processing capacity (Sweller, 1988). And when it comes to reasoning, adolescents are novices who still are in the process of developing problem-solving strategies. Experts, unlike novices, have acquired generic memories-developed schemas or patterns to solve problems. This 'pattern recognition' is based on learning processes which underlie repeated exposure to similar contexts that originate in matched response strategies. According to Goldberg (2010), a transition occurs from an absence of effective behavior of WM to the formation of effective behavior in reasoning processes. When this transition occurs, reasoning questions can be addressed more effectively and efficiently.

It is not yet fully understood how WM retrieves problem-solving schemas that are stored in LTM. According to cognitive load theory (CLT), the human cognitive system has a limited WM capacity which is able to deal with novel information that is obtained through sensory memory. WM has no known limitations when it deals with information that is retrieved from LTM which, in its turn, alters the characteristics of WM. LTM 'holds cognitive schemas that vary in their degree of complexity and automation. Human expertise comes from knowledge organized by these schemas' (Van Merriënboer and Sweller, 2005, 2010 p.86). Gooding, Mann and Armstrong (2016, p.1.), for instance, indicate that 'as the capacity of WM is finite, information must be processed and stored in LTM for later use. Information in LTM is organized into schemas of increasing complexity, allowing individuals to retrieve a schema for use in WM as a single construct'. Ericsson and Kintsch (1995) suggest that, considering the huge cognitive load, traditional models of WM involving temporary storage, must also include LTM storage. Consequently, they propose a theoretical framework in which cognitive processes, such as schema acquisition, are viewed 'as a sequence of stable states representing end products of processing', and postulate that 'in skilled activities, acquired memory skills allow these end products to be stored in LTM and kept directly accessible by means of retrieval cues in short term memory (STM)' (p. 211). In short: 'In the cortex, as in the rest of the brain, there are no 'systems of memory', but there is the 'memory of systems' ' (Fuster, 2009, p. 2063). In this case, more recent approaches have addressed the relationships between WM and reasoning by focusing on the cooperation of various task specific neuronal networks 'dedicated to the representation and retrieval of individual memory and knowledge', including WM (Fuster, 2009, p. 2048). Fuster's (2003, 2006) Perception-Action Cycle defines a hierarchical model of dynamic development in which attained reasoning strategies are 'reducible to abstractive executive cognits. Of course, the strategies are 'transferrable', by definition; they are made of concrete instantiations and ready to be used in other instantiations. WM helps all deductive reasoning at all

hierarchical levels and is considered to be trainable to help the student reach the higher level (i.e. abstract generalization by deductive reasoning).'[5] In this view, complex behavior such as reasoning requires integration at higher cortical levels of association and prefrontal cortexes.

It is appropriate to include strategies of pattern recognition into school-based reasoning training in order to acquire the relevant cognitive tools for reasoning. This can lead to intuitive reasoning based on tacit knowledge. With regard to the requirements of school-test reasoning questions, training models can best be based on knowledge-based attention training (WM-training) and schema acquisition (Gf, the ability to reason and solve new problems independently of previously acquired knowledge). In fact, studies increasingly indicate that the implementation of subject-oriented content and processes is an important element to enhance the effectiveness of WM-training models due to (more effective) near-transfer effects (Van der Sluis, De Jong and Van der Leij, 2007; Lee, Ng and Ng, 2009). And Paas, Van Merriënboer and Van Gog (2011) indicate that the effectiveness of instruction is determined by the extent to which the characteristics of WM and LTM are taken into account in the design of instruction.

Focusing attention and acquiring schemas are regarded as important skills to address reasoning problems. Adolescents frequently have to overcome difficulties to develop their reasoning skills. Apart from environmental (social and pedagogical) factors in the classroom or at home, such as peer pressure, course didactics, or common distractions, developmental scientists have focused on brain maturation to explain the deficits of adolescent reasoning skills: the ability to exert cognitive control over the adolescent's thoughts and actions increases steadily during adolescence. The development of cognitive control appears to be crucial in cognitive development, along with representation and metacognitive

[5] J. Fuster (2015, August 11 – September 11), Personal Interview.

systems (Chein and Schneider, 2012). Crone and Dahl (2012) state that 'the ability to exert cognitive control over thoughts and actions is of crucial importance to success in most classroom settings – not only for the direct learning of skills such as reading, math and the capacity to reason about abstract ideas but also at the level of behavioral control that supports sitting at a desk, avoiding distractions and doing homework' (p.638). Middle adolescence (13-17 years) is regarded as a period of heightened vulnerability to risk-taking, and a period during which adolescents have problems in the regulation of affect and behavior, while the maturation of the frontal lobes facilitates regulatory competence in late adolescence (18-20 years) (Steinberg, 2005). In fact, several brain development models ascribe the adolescent's more emotional (and less rational) decision making to the 'relatively faster maturation of subcortical affective brain areas in comparison to more slowly maturing frontal cortical brain areas' (Crone and Dahl, 2012, p.637). Furthermore, motivational aspects appear to influence the activation of cognitive control, such as the 'presence of peers, task instructions, strategies and the affective appraisal of the value of priority of performing the task' (Crone and Dahl, 2012, p. 639; Martin and Steinbeck, 2016). Consequently, during middle adolescence young people consistently react more strongly to receiving rewards, but not to the expectation of being rewarded, compared with children and adults. Also, middle adolescence is regarded as a transition period in which more complex cognitive skills, such as mentalization and metacognition, develop.

Training methods usually focus on executive functions (ef), a system that is attributed to WM. Ef are a set of cognitive processes and conceptualized as having four components: the abilities of goal formation; planning; carrying out a goal-directed plan; and effective performance (Jurado and Rosselli, 2007). Training in the following three ef's positively affects reasoning abilities (Süss, Oberauer, Wittman, Wilhelme et al., 2002; Conway, Kane and Engel, 2003; Kramarski and Mevarach, 2003; Cheshire, Ball and Lewis, 2005; St Clair-Thompson and

Gathercole, 2006; Jaeggi, Buschkuehl, Jonides and Perrig, 2008). (1) 'Shifting' between multiple tasks actions or mental sets is related to multi-tasking. It plays an important role in Bloom's classification of six subsequent levels of thinking because multiple skills and tasks are needed to solve a problem (Davidson, Amso, Anderson and Diamond, 2006). Shifting can be trained through various tasks, such as the Objects Shifting task, Making Trails task, or Place Shifting task (Van der Sluis et al., 2007). The correlation between shifting and verbal reasoning and reading is low, according to Van der Sluis. Instead, she insists that shifting can be trained in combination with other ef training to boost the correlation with verbal reasoning, for instance with 'Updating'. (2) 'Updating' encodes newly received information. In addition, it replaces information which is no longer relevant to solve a problem by relevant information in the WM. WM, therefore, becomes a dynamic system and plays an important role in many daily learning and organization processes. There are various tasks available to train the Updating function, such as the Keep Track task, the Letter Memory task, or the Digit Memory task (Van der Sluis et al., 2007). Her research indicates that updating is related to (non-)verbal reasoning, but the correlation with non-verbal reasoning is stronger than with verbal reasoning and arithmetic. Presumably, training of the updating function will less affect verbal reasoning, which is considered to be important in humanities courses. While WM-tasks and updating tasks both require a revision of information, training on WM-tasks can also enhance the processing function of information. Comparable research has also concluded that there is no strong causality between the ability of only storing information and (non-verbal) reasoning, while the ability to both store and process information does enhance (non-verbal) reasoning significantly (Conway et al., 2003). Furthermore, training in updating develops highly specialized skills and is therefore difficult to implement in a general training model for reasoning skills because reasoning questions in education contain many different types of problems that students have to address, for instance texts, graphics, tables and formulae. (3) WM itself provides the brain with an information storage and manipulation

function, and is related to academic progress in reading and math. As such, WM is an important predictor in understanding individual differences in achievement between adolescents (Bull and Scerif, 2001; Lee et al., 2009). Training includes Representation Formation tasks (Lee et al., 2009), or metacognitive training tasks, such as IMPROVE (Kramarski and Mevarach, 2003). Cogmed (a WM and attention training program) and AWMA (Automated WM Assessment) tasks are popular WM capacity training tasks whose usage serves medical, educational and psychological purposes. As far as WM capacity is concerned, training on n-back tasks is observed to lead to small positive effects on fluid intelligence (*Gf*) (Au et al., 2015). Furthermore, evidence suggests that *Gf* and deductive reasoning are positively correlated (Stanovich and West, 2000; Beatty and Vartanian, 2015). And, theoretically, WM could improve deductive reasoning through increases in *Gf* (Beatty and Vartanian, 2015). Also, n-back training has been shown to produce broad transfer effects from training-related changes to untrained tasks and functions (Von Bastian and Oberauer, 2013). According to Salminen, Mártensson, Schubert and Kuhn (2016), this is caused by n-back training to increase 'white matter integrity [in the brain] in pathways that connect brain regions related to WM processes (p.245). Lilienthal, Tamez, Shelton, Myerson and Hale (2013) investigated five possible processes that underlie the effects of (dual) n-back on WM, these being: improvements in executive attention; updating; focus switching; increases in the capacity of the focus of attention and increases in the capacity of short term memory (p.135). 'Executive attention' regulates the contents of the active portion of memory (Engle, 2002, p. 19). 'Updating' requires monitoring and coding of information and appropriately revising the items held in WM by replacing no-longer-relevant information with new, more relevant information (St Clair-Thompson and Gathercole, 2006, p.746). 'Switching' involves switching of attention between multiple tasks (Smith, Geva, Jonides, Miller, Reuter-Lorenz et al., 2001, p.2095). N-back requires storage and updating of information in WM and 'interference resolution'. But there is much debate on whether n-back training affects the

outcomes of far transfer (i.e. performance improvements that are dissimilar to the trained task) measures of deductive reasoning. We therefore include n-back as a training task combined with reasoning-strategy training to provide a rigorous training program to improve reasoning in school-based humanities curricula.

1.2 Aims, outline and research questions

The theoretical background described above provides a brief overview of the complex mechanisms surrounding the dynamics of WM-training in order to improve content-based reasoning. That is the main reason why there is still much discussion about the effectiveness of WM-training for content-based reasoning improvements. Therefore, we hypothesize that WM-training is only effective in educational settings when it is accompanied by reasoning-strategy training in a subject-based manner in the humanities. Evidence for this is provided by the literature review on n-back training and theorized by the research synthesis on WM capacity and reasoning strategy training. Furthermore, we argue that the success of such training depends on the students' motivation. The aim of this dissertation is to evaluate a model which includes content-based WM capacity and reasoning-strategy training, and to evaluate its effectiveness in humanities courses.

To achieve this aim, Chapter 2 of this dissertation reviews the effects of one specific WM-training task, n-back, on deductive reasoning improvements. N-back training might stimulate WM-function and, as a result, improve reasoning. There are no systematic reviews of the literature on the effects of n-back training in school-based settings. We include school-based and non-school-based interventions in our review. Our conclusion is that reasoning test outcomes are ambiguous and tend to be non-significant, which implies that n-back training in its pure form cannot be implemented in school-curricula. We do, however, make recommendations for implementing n-back in school-based training

programs. Therefore, Chapter 2 can inform the debate about WM-training and far transfer effects in primary, secondary and tertiary education, and the first research question we address in this thesis is:

Question 1: Does n-back training improve school-based reasoning achievements?

Chapter 3 of this dissertation provides a synthesis of the literature on WM-capacity- and reasoning-strategies training. It focuses on primary and secondary education and analyzes whether a combined WM capacity and metacognition training is effective in education settings to develop reasoning skills. It differs from previous reviews on school-based WM-training because it specifically focuses on achievement outcomes with students with normal learning behavior as well as specifically focusing on two aspects of WM that are suggested to improve reasoning achievements: WM capacity, and reasoning strategy. We therefore develop a theoretical basis and present a research synthesis that provides the basis for an evidence-based training program. In Chapter 3, we review the literature on WM capacity which is specific for school-based interventions. Furthermore, we review the literature on reasoning strategy training that is based on WM-theories. This analysis provides insight into whether both aspects of reasoning training can indeed enhance reasoning achievements in content-based school tests. Therefore, the second research question we address in this thesis is:

Question 2: Can a combined working-memory capacity and metacognitive training serve as an effective instrument to enhance school-based reasoning achievements?

The effect of the combined training is analyzed in Chapters 4 and 5. Here, the combined training, as well as training in one of the WM-aspects is compared with the control group conditions in history and social studies courses in secondary education.

Chapter 4 evaluates two interventions in the history courses of higher general secondary education (HAVO) and pre-university secondary education (VWO). HAVO students, who pursue the second-highest variant of secondary education in the Netherlands, received a combined training battery in order to analyze its effectiveness in reasoning achievements in school tests. The results are compared with a control group which received a conservative training. In VWO, WM capacity and reasoning strategies are independently trained and compared with control group results in order to analyze the improvements of both training groups independently. This allows us to test two research questions:

Question 3: Does the combined working-memory capacity and metacognition training, based on subject-matter knowledge in history education, have an effect on students' achievements in tests for which reasoning abilities are required?

Question 4: Does working-memory capacity and/or metacognitive training, when independently trained, have an effect on students' achievements in tests for which reasoning abilities are required?

Chapter 5 includes the combined intervention in HAVO social studies courses. The main aim of this intervention is to analyze whether the training duplicates the results of the intervention described in Chapter 4. This is an important question which can give insight into the replicability of the training in all the humanities courses, as well as into possible transfer effects which can be investigated in the future. Therefore, the fifth research question is:

Question 5: Does the combined training have an effect on students' reasoning achievements in school tests in other humanities courses (i.e. social studies)?

Chapter 6 includes the combined intervention to which motivational factors are added in order to investigate the results among less motivated students. The idea is that the success of the combined training is largely based on the individual student's motivation during the training. In this experiment, we therefore compare the school test results between both groups that receive the combined training, where one group is actively being motivated by the teacher. Both HAVO experimental and control groups receive the training in history courses. The main difference is the inclusion of components that are specifically designed to improve attention during the experimental groups' training. The main objective is to analyze whether the addition of motivational components can improve test results even more compared with the test results of the control groups. Therefore, the final research question addressed in this thesis is:

Question 6: Does the combined training affect reasoning achievements in school-test results when motivational factors are added to the training?

Chapter 7 is the concluding chapter in which the combined findings and overall conclusions are presented. In conclusion, it presents policy implications, as well as future research directions.

2 Does n-back training in a school-based context improve reasoning? A literature review[6]

Summary

N-back training might stimulate the working-memory (WM) function and, as a result, improve reasoning. To date, there have been no systematic reviews of the literature on the effects of n-back training in school-based settings. At the same time, school students sometimes experience difficulties in school-tests that are composed of reasoning questions. This chapter reviews the literature on n-back training in relation to reasoning in a school-based context. It focuses on n-back training in school-based designs. School-based WM training, which includes n-back tasks and non-school-based designs, is also analyzed in order to broaden the scope and possible implications and pitfalls of school-based n-back designs. We conclude that the results are ambiguous with a tendency to non-significant effects on reasoning improvements. Future research should better address methodological and didactical considerations before proceeding to implement n-back training in school curricula. These concern: discussing motivational issues; personalizing training by accounting for age; baseline WM capacity, and reasoning skills; and minimizing distractions.

[6] Chapter 2 is based on: Ariës, R.J.F.J., Ghysels, J., Groot, W. & Maassen van den Brink, H. (2017). Does n-back training in a school-based context improve reasoning? A literature review. TIER Working Paper Series 17/08. ISBN: 978-94-003-0124-5.

2.1 Introduction

Working-memory (WM) training – training of the cognitive system that stores and manipulates information for short periods of time – might affect a broad variety of higher-order skills. This comes from the premise that WM is considered to be an 'important predictor of academic achievement' and a 'critical bottleneck underlying higher-order cognitive processes, including controlled attention and reasoning' (Au et al. 2015, p.366). Consequently, there is much interest in whether WM training might improve students' reasoning skills and school-based reasoning test scores. This is because, theoretically, if WM training can improve reasoning skills, a single general training might affect reasoning skills and test scores in all school courses in a positive way. In other words, WM training could improve reasoning skills more effectively and efficiently.

The potential benefit of school-based WM training lies in performance improvements on reasoning questions that are dissimilar to the trained task (i.e. far transfer effects). Although experiments that investigate the far transfer effects of WM training in school-based settings are still scarce, there is a growing number of studies that investigate the transfer effects of WM training to gains in measures of fluid intelligence (*Gf*). In this regard, 'Gf refers to the ability to reason and to solve new problems independently of previous required knowledge' (Jaeggi et al., 2008, p.6829). And, in turn, it is theorized that improvements in reasoning can occur through increases of Gf by WM training (Beatty and Vartanian, 2015; Swanson and McMurran, 2017).

In this chapter, we review the literature on the effects of one specific WM training task, the n-back, on deductive reasoning improvements. We have selected n-back based on the selection criteria used by Au et al. (2015, p. 367): namely, that the n-back task, 'which requires not only the storage and continual updating in WM, but also interference resolution has been used widely in WM training studies that explore transfer to *Gf' (*Au et al., 2015, p. 367). To elucidate on far-transfer results to reasoning, we include studies on school-based and non-school-based interventions in our review. Our conclusion is that

reasoning test outcomes are ambiguous and tend to be non-significant, which implies that n-back training in its pure form should not be implemented in school-curricula. The discussion section provides recommendations for implementing n-back in school-based training programs.

2.1.1 Fluid intelligence and reasoning

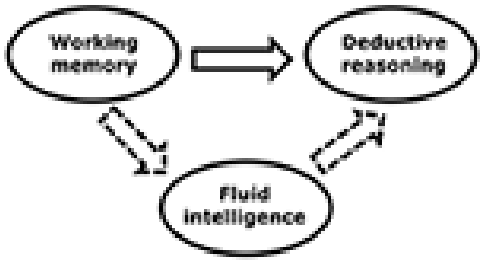
Gf can be defined as a 'complex human ability that allows us to adapt our thinking to new cognitive problems or situations for which we cannot rely on previously acquired knowledge' (Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides et al., 2010). It is considered to play a crucial role in a variety of cognitive tasks, such as reasoning (Jaeggi et al., 2010). Therefore, the evidence suggests that general cognitive ability and deductive reasoning are positively correlated (Stanovich and West, 2000; Beatty and Vartanian, 2015). And, theoretically, WM could improve deductive reasoning through increases in *Gf* (Beatty and Vartanian, 2015; Swanson and McMurran, 2017) (Figure 2.1).

Until recently, *Gf* was viewed to be immutable because of its heritability. But most contemporary scholars now acknowledge the role of heritability, as well as environmental factors and even training, in the development of *Gf* (e.g. Au et al., 2015). Specifically n-back training, which is a commonly used task in WM task-transfer training studies, could lead to positive effects on *Gf* (Au et al., 2015; Jaeggi et al., 2008; Beatty and Vartanian, 2015). Ergo, n-back training could theoretically improve deductive reasoning via its effect on *Gf.* We therefore conduct a systematic review of the literature on the effects of n-back training on the improvements in the students' reasoning skills with a special focus on the learning trajectory through *Gf*.

2.1.2 Effects of n-back training on WM-mechanisms and the focus of attention

N-back requires storage and updating of information in WM and 'interference resolution'. Therefore, more than other WM-training tasks

n-back training could positively affect fluid intelligence (Au et al., 2015). The following description of the n-back task is derived from Au et al.'s (2015, p.367) meta-analysis on improving fluid intelligence with training on WM. The n-back task involves the serial presentation of a stimulus, spaced several seconds apart. The participant must decide whether the current stimulus matches the one displayed *n* trials ago, where *n* is a variable number that can be adjusted up or down to, respectively, increase or decrease cognitive load (see: Merriënboer and Sweller, 2010). In the context of WM-training, efforts have focused on flexibly adapting the task difficulty in accordance with the participants' fluctuating performance levels by increasing and decreasing the level of *n*. The idea is to keep the participant's WM system perpetually engaged at its limit, thereby stimulating an increase in WM function. This translates into more general improvements in tasks that rely on the integrity of WM skills, such as *Gf* (see Figure 2.1).



Note. 'The solid arrow depicts a direct effect. The dashed arrows depict an indirect effect.' Reprinted from 'The prospects of working memory training for improving deductive reasoning,' by E.L. Beatty and O. Vartanian. 2015, *Frontiers in Human Neuroscience, 9*(56). Reprinted with permission from the authors.

Figure 2.1. 'Two possible routes for improving deductive reasoning deductive reasoning by WM-training'

Little is known about which WM components and processes need to be improved to increase the WM function. Lilienthal et al. (2013) investigated five possible processes that underlie the effects of dual n-back on WM, these being: improvements in executive attention;

updating; focus switching; increases in the capacity of the focus of attention and increases in the capacity of short term memory (p.135). 'Executive attention' regulates the contents of the active portion of memory (Engle, 2002, p. 19). 'Updating' requires monitoring and coding of information and appropriately revising the items held in WM by replacing no-longer-relevant information with new, more relevant information (St Clair-Thompson and Gathercole, 2006, p.746). 'Switching' involves switching of attention between multiple tasks (Smith et al., 2001, p.2095). The intervention included an adaptive training group, a non-adaptive training group, and a no-contact control group. Lilienthal et al. concluded that increases in the capacity of the focus of attention underlie the benefits of adaptive dual n-back training. This suggests that increases in *Gf* could be predicted by increases of the capacity of the focus of attention.

2.1.3 The far-transfer debate

There is much debate whether n-back training might affect outcomes of fluid intelligence measures, especially far-transfer effects such as school-based reasoning. Recently, the debate has been defined by two influential review studies, one which supports, and the other which opposes the effect of WM-training on *Gf*.

Au et al. (2015) conducted a meta-analysis of 20 articles on improving fluid intelligence with training on working memory. It focuses on the effects of n-back training on improving *Gf*, and finds a small but significant positive effect. They conclude that even short-term cognitive training can result in a beneficial effect with regard to *Gf* outcome measures within a general population of healthy young adults and regardless of experimenter bias, and active or passive control groups. They propose that the observed improvements are plausible because '*Gf* and n-back draw on overlapping cognitive and neural processes, including shared demands on WM and interference resolution' (p.375). Au et al. conclude that n-back training to increase *Gf* is promising, but do not give insight into educational perspectives.

Melby-Lervåg et al. (2016) have conducted a meta-analysis of WM training in relation to measures of intelligence and other measures of far transfer in which 87 publications are included. The meta-analysis included a broad range of studies, including n-back training, and examines whether study characteristics can explain the variations in outcome measurements (inter alia, the relationships between intermediate-transfer and far-transfer effects). Based on the analysis, the authors question the practical importance of WM-training programs. For instance, no convincing evidence is reported for improvements in measures of far transfer after training compared with a treated control condition. N-back training is also specifically analyzed and shows a significant effect on nonverbal ability. But further examination of those studies with the largest effect sizes revealed several shortcomings, such as: small sample sizes; the use of only one test to measure nonverbal ability; or unexplained pretest-to-posttest decreases for the control group. As a result, they argue that effect sizes are 'too small to be relevant in educational practice' (p.520).

In addition, Redick et al., (2015) published a review on the effects of general WM-training in an educational perspective. Besides a description of criteria on how to evaluate evidence of WM training, such as the use of active or passive control groups, sample sizes or objective measures, Redick et al. provide an overview of studies in which WM is trained in order to analyze an education/achievement-related transfer. The studies include Cogmed training, updating, complex span, jungle memory, strategy WM complex span and memory booster training tasks. They conclude that WM training produces only limited effects in intermediate transfer, and no effects for achievement-based reading and arithmetic outcomes.

Redick et al. (2015) did not specifically focus on n-back training. The results of n-back training on academic achievements were therefore not isolated in the analyses. An overview that focuses on the n-back task to far-transfer effects specifically in an educational perspective is still lacking. Along with the idea that n-back training could possibly mediate the link between WM and deductive reasoning, this warrants a review

that can inform the debate about WM training and far-transfer effects specifically in primary, secondary and, tertiary education. The current review provides insight on this topic by reviewing the literature on (non-) school-based interventions with children and young adults and far-transfer results in reasoning. The method and search strategy is discussed next, in Section 2.2. Then, Section 2.3 presents the results, before the final Section 2.4 draws some conclusions.

2.2 Method and search strategy

Concerning the analysis of WM-training in educational settings, Redick et al. (2015) proposed a best-practices framework for strong evidence (Table 2.1, categories 1-7). Redick et al. considered this to provide the ideal criteria for the evaluation of the efficacy in achievement-related tests and measures. It serves as a 'best practices guide in terms of study design and data interpretation' (Redick et al., 2015, p.3-4). The proposed best practices framework is partly built upon the tradition of causal interpretative effect research (e.g. the minimum of participants to retain statistical power, and the use of valid measurements). On the other hand, parts of the proposed criteria focus specifically on WM training in an educational perspective (e.g. the use of active and passive control groups, causal relationships between near- and far-transfer results). It therefore provides useful guidance to collect and analyze the evidence for far-transfer effects. Considering the specifics of WM-training in educational settings compared with WM training in general, we address the definitions of the criteria and alter them where needed to better suit the goals and dimensions of education-based training and of our review.

Redick et al. proposed that the nature of the control group affects causal interpretations, a passive control group being one in which participants do not receive WM-training and have no contact with the experimenters. Active control groups are defined as practicing an alternative type of task, unrelated to WM. Redick et al. include only those

studies that have active control groups, because passive control group studies would generate many other variables, such as amount of experimenter/computer contact, or motivation, which can complicate the analysis of causal relationships. But many education-based training tasks rely on WM to a greater or lesser extent that can affect WM measurements, such as reasoning training. Also, education-based research designs often make use of the same trainer for experimental and control groups. Furthermore, experimental and control groups are usually located in the same school, so that subjects of both experimental and control groups are probably aware of an experiment taking place. Therefore, excluding a passive control group in favor of active control groups may not 'eliminate alternative explanations', as Redick et al. (2015, p. 620) mention in their article. We also point out that studies on education-based WM-training to affect academic outcomes usually compare school-based outcomes of WM-training compared with standard school training in order to gain more insight into the added value of WM-training to academic development. Based on these considerations, first, we include research designs that make use of active or passive control groups in our review and address effect differences.

Second, we include studies that contain data for a minimum of 20 subjects (to ensure sufficient statistical power). Redick et al. (2015) consider this sample size an absolute minimum, while the use of larger sample sizes provides stronger and accurate evidence.

Third, in line with Redick et al.'s criterion for the use of objective measures (valid measurements), research that reports non-objective academic measure and achievement tests – such as teachers' opinions - are excluded from our review. Redick et al. also express the concern that responses on questionnaires may reflect placebo effects because those doing the rating are aware of the participant being in the experimental or control condition. This may also be the case when studies make use of open-ended questions in achievement tests. We therefore add a category in our analysis that provides insight into the rating process in order to

determine it is blind or not (Table 2.1, category 3a) and include both types of studies in our review.

Fourth, it is important to find causal relationships between WM-training and far transfer to academic and achievement outcomes. According to Redick et al., causal relationships can be demonstrated by tests in which near-transfer effects to other WM tasks have been measured. This means that academic and achievement pretesting and posttesting should be accompanied by tests in WM tasks other than, in the case of the current review, n-back. In other words, when WM training generates far-transfer effects, but did not generate near transfer effects, then, according Redick et al., causal relationships can not be directly demonstrated. But, considering the main goal of our review, which is to analyze whether n-back affects academic achievement, we chose to include articles that lack WM tests or have WM tests that are identical or very similar to the training programs.

Fifth, the strength of evidence can be interpreted by looking at whether or not 'the pattern of transfer results is sensible' (Redick et al., p.621). This means that: (1) training and control groups have similar pretest scores; and (2) decreasing performance values from pretest to posttest is an issue, because this could mean that the effect is primarily driven by 'the control group's declining performance instead of the training group's improving performance'. Considering the nature of school-tests, many of them being produced by teachers in order to test different book chapters and/or subjects, we propose that this type of school-testing is not suitable for examining within-group effects on the posttest if the pretest is used as a covariate. Instead, we opt that the main criterion should be that experimental and control groups receive identical pretests and identical posttests as far as school tests are concerned. If testing takes place through school-tests, we only include studies in which both types of groups receive identical tests.

Sixth, in line with Redick et al., we provide insight on follow-up tests in transfer. Redick et al. rightly acknowledge the debate on the duration of transfer after training, to which no definitive conclusion has yet been

reached. On the one hand, there are reasons to think that transfer does not persist after training, and, on the other hand, there is reason to expect that transfer to achievement measurements only shows up some time after the training has finished.

Seventh, stronger evidence of the transfer ability of the training is accomplished by the use of multiple measures, instead of based on one test only. Criteria six and seven are considered to be important, but do not count as requirements to be included in the review.

Table 2.1 Redick et al. (2015). 'Criteria for strong evidence of working memory training efficacy in [an] educationally relevant context'.

1. Use of active control group
2. Use of large sample sizes in each training and control group
3. Use of objective measures

3a. Use of blinded rating process

4. Evidence for positive transfer results to working memory
5. Transfer results follow a sensible pattern
6. (Follow-up transfer assessment)
7. (Multiple measures of each construct)

Note. 3a is not included in Redick et al.'s original list.

We include (school-based) interventions that use n-back training, or a combination of n-back and other WM-training tasks, to affect (educational) problem-solving or reasoning achievements. We have searched for published peer-reviewed articles and unpublished material (e.g. dissertations, from 2007 to 2016), since school-based n-back training has only been conducted in the last few years. ERIC, Psycinfo, PubMed and Google Scholar databases have been consulted in the search for published material. We have also searched in the Google Scholar database for unpublished material (see Figure 2.2 for flowchart).

We have used the following keywords in the search: n-back (training), working memory training, *or* cognitive training in combination with school achievements, *or* (primary/secondary) education, *or* problem-solving, *or* reasoning. This yielded 68 articles. Based on a first selection of

the literature by excluding all interventions in pre-school or adult (not being 'university') education, we extracted 11 articles. This means that the mean age of the restricted group is between 6 and 23.

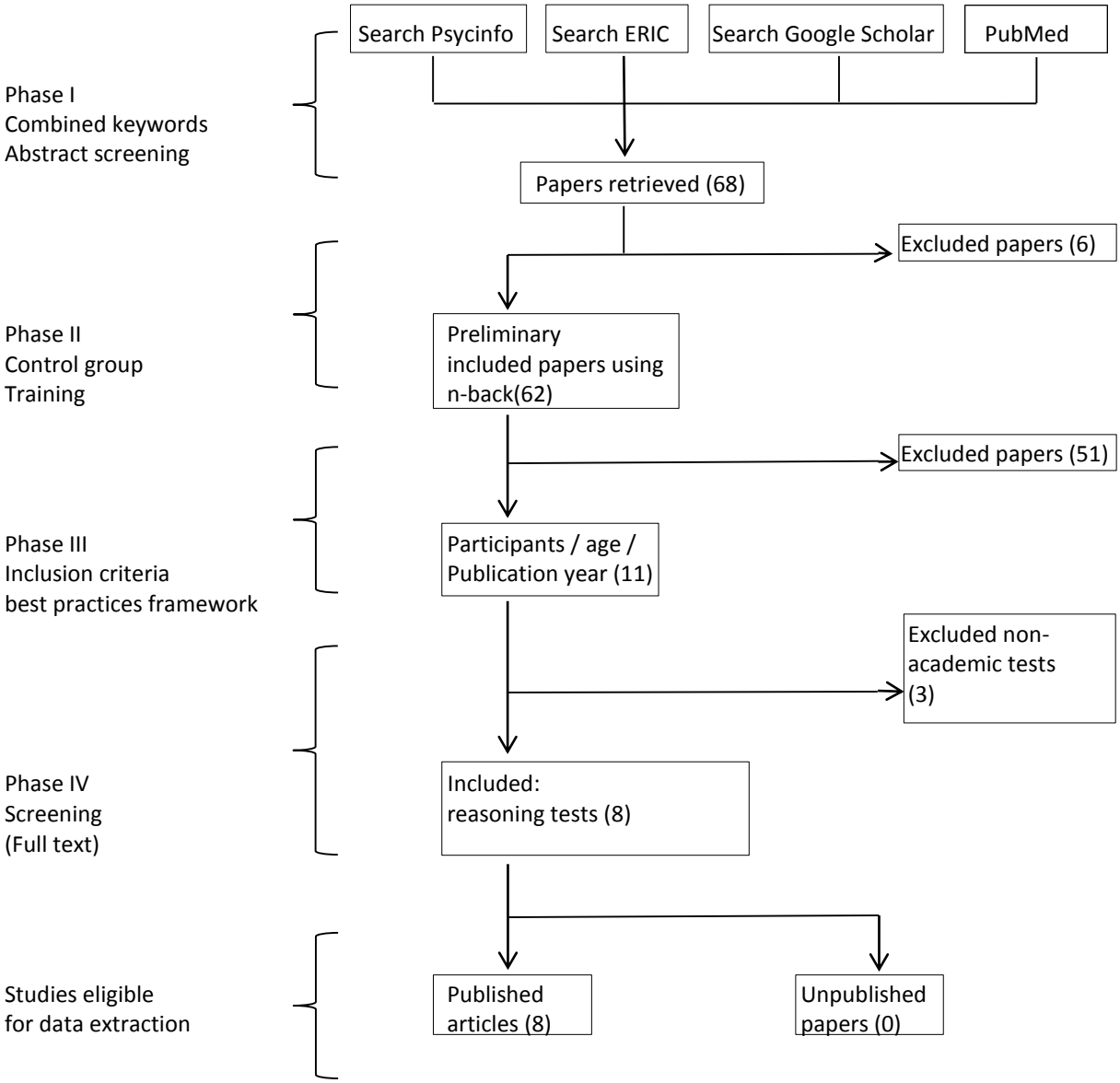


Figure 2.2 Flowchart of the literature review process

Our inclusion criteria are as follows. The main characteristic is that interventions must include some form of n-back and have far-transfer measures to (school-based) reasoning, use a control group, and include a deductive reasoning outcome measure. The age of the participants of the restricted group is between 6 and 23 because we focus on n-back training

in the young population. We have not excluded articles that combine n-back with another training task. However, we have elucidated in Table 2.3 whether an n-back task or a combination of tasks is used in the intervention.

Table 2.2 Evaluation of the criteria per intervention and far transfer results on reasoning

References	Control group	Sample size	Objective Measurement	Blind rating	Transfer to WM	Sensible	Far transfer results
Jaeggi (2010) Exp 2	Passive	89	Yes	n.a.	Yes	Yes	Yes
Jaeggi (2011)	Active	60	Yes	n.a.	No	Yes	No/Yes
Colom (2013)	Passive	60	Yes	Yes	No	Yes	No
Redick (2013)	Active / Passive	73	Yes	Yes	Yes	Yes	No
Thompson (2013)	Active	58	Yes	Yes	Yes	Yes	No
Chapter 4 Exp 1	Active	91	Yes	No	No	Yes	Yes
Chapter 5	Active	81	Yes	No	No	Yes	Yes
Minear (2016)	Active / Passive	116	Yes	Yes	Yes	Yes	No

Notes. Control group: use of an active or passive control group; Sample size: minimum of 20 subjects; Measurement: use of achievement tests or academic measurements; Blind rating: rater awareness of participant being in a control or experimental condition; Transfer to WM: Use of WM tests to analyze near transfer effects to WM; Sensible: use of identical pretests and posttests in groups; Far transfer results: evidence of WM-training to school-based reasoning. n.a.: information not available. Reasoning or Gf tests depicted in the 'effect size' column when more tests were used.

Next, in line with our altered best practices framework for strong evidence of WM-training in educational settings analysis, described on the previous page, we have made the final selection (Table 2.2). This means that we have included interventions that have made use of either passive or active control groups, contain data for at least 20 subjects, and report objective academic measure and achievement tests. Furthermore, groups should have been tested by the use of identical tests. The final selection included 8 articles and 0 unpublished papers, including the experiments described in Chapters 4 and 5 of this dissertation. In Table

2.3 we have listed information about the samples, design, duration, and tasks and measurements used for each study.

2.3 Results

In order to elucidate when far-transfer effects do or do not occur, we describe the results of our review in different categories (Table 2.4). First we provide information about the effect sizes per study, followed by n-back training in school-based settings, and non-school-based interventions. Then, we report the effects of the training to far transfer for primary, secondary, and tertiary education. After that, we analyze the far-transfer effects of the intervention groups when compared with either the passive or the active control groups. Lastly, we describe the effects on reasoning for interventions that train solely on n-back, or for interventions in which n-back was part of the training battery.

All interventions used a pretest-intervention-posttest design in which objective academic measurements are used. All but one intervention (Jaeggi et al., 2011) used randomized experimental and control groups. The average sample size was 67 participants, with Thompson, Waskom, Garel, Cardenas-Iniguez, Reynolds et al. (2013) having the smallest sample size (n=58), and Minear, Brasher, F., Brandt Guerrero, Brasher, M., Moore et al. (2016) having the largest (n=116). Most participants were trained in a 4-week training program, with Minear et al. (2016) having the lowest training duration, and Colom, Román, Abad, Chun Shih, Privado et al. (2013) having the highest. All interventions used identical pretests and posttests to measure for far transfer effects. Raven's Progressive Matrices test, a test for Gf that uses matrix reasoning, is the most commonly used reasoning test. Only 4 of 8 studies have checked for near-transfer effects to WM by using a WM task in the test that is different from the task that was used in training. The mean age of the participants was 18. The oldest group of participants had a mean age of 22, and the youngest a mean age of 8, which means that interventions were implemented from primary to tertiary education. We found four

Table 2.3 Overview of interventions assessing the impact of n-back training on reasoning

References	Sample	Design	Duration	Tasks	Reasoning test
Jaeggi (2010) Exp 2	mean age 19 university	20 sessions, control group matching intervention, pretest posttest	4 weeks	n-back	RAPM, BOMAT
Jaeggi (2011)	mean age 9 primary education	20 sessions, pseudorandom control group intervention pretest - posttest	4 weeks	n-back	RSPM, TONI
Colom (2013)	mean age 18 university	20 sessions, control group intervention pretest - posttest	10 weeks	n-back	RAPM, DAT, PMA
Redick (2013)	mean age 21 university	20 sessions, control group intervention pretest - posttest	5 weeks	n-back	i.a. RAPM, RSPM, Cattell Culture-Fair, Paper Folding, Letter Sets, Number Series
Thompson (2013)	mean age 22	20 sessions, control group intervention pretest - posttest	3-5 weeks	n-back	RAPM, WASI, WAIS
Ariës (2015) Exp 1	15-16 years high school	randomized passive control group intervention, pretest - posttest	6 weeks	Content-based n-back, Odd One Out, reasoning strategy	Reasoning in history school tests
Ariës (2016)	15-16 years high school	randomized passive control group intervention, pretest - posttest	4 weeks	Content-based n-back, Odd One Out, reasoning strategy	Reasoning in humanities school tests
Minear (2016)	20 years university	5 sessions, randomized active and passive control group, pretest - posttest	1 week	n-back	ETS, RAPM,

Table 2.4 Significant effects of far transfer to reasoning: additional categories.

References	Effect size (Cohen's d)	School-based	Non school- based	1 educ	2 educ	3 educ	Passive con	Active con	N-back	N-back +
Jaeggi (2010) E 2	0.10		Yes			Yes	Yes		Yes	
Jaeggi (2011)	0.80 / 0.55		No/Yes	No/Yes				No/Yes	No/ Yes	
Colom (2013)	RAPM 0.35, DAT_AR 0.07, PMA_R 0.16		No			No	No		No	
Redick (2013) Passive control	RAPM 0.08, RSPM -0.32, Catell -0.03, Paper 0.26, Letter 0.11, Number 0.04, Inferences -0.27, Analogies -0.06		No			No	No		No	
Redick (2013) Active control	RAPM 0.003, RSPM -0.15, Catell -0.06, Paper -0.14, Letter 0.09, Number 0.18, Inferences 0.13, Analogies -0.35		No			No		No	No	
Thompson (2013)	RAPM 0.29, WASI/WAIS 0.32		No			No		No	No	
Ariës (2015) E 1	2.20	Yes			Yes			Yes		Yes
Ariës (2016)	1.04	Yes			Yes			Yes		Yes
Minear (2016)	n.a.		No			No	No	No	No	

Notes. Far transfer of intervention to reasoning is analyzed per category. Yes: transfer to reasoning. No: no transfer to reasoning. School-based: school-based interventions. Non-school-based: non-school-based interventions. 1/2/3 educ: intervention in a group that belongs to one of the levels of education. 1 educ: primary education. 2 educ: secondary education. 3 educ: tertiary education. N-back: solely n-back in intervention. N-back +: n-back and another training method in the intervention. Passive / active con: passive or active control groups. E 1/2: experiment in the study.

significant and five non-significant results concerning far-transfer effects to reasoning (*Mdn d* = 0.11*, range* -0.35 – 2.20*).*

2.3.1 School-based interventions

Reasoning measurements are usually conducted using school tests that include content-based reasoning questions. Interventions within schools are primarily conducted in groups containing 20 or more students which were simultaneously trained by the use of n-back and other training tasks, while the tests were administered individually.

Two interventions were school-based. Both studies found significant results (*range of Cohen's d*=1.04 – 2.20). This means that training was implemented in (secondary) school curricula. Both interventions made use of n-back training in combination with another training type (reasoning strategy). When WM training is combined with reasoning strategy training, then significant improvements occur in reasoning test results. For instance, later in Chapter 4 (Experiment 1) we trained the experimental group (n=36) with a training battery that included content-based Odd One Out (a task suitable for training reasoning abilities), n-back, and reasoning strategy during the history course in secondary education. The experimental group improved significantly on WM capacity measurements compared with the control groups. Also, the experimental group achieved significantly better scores compared with the control group test scores 3 weeks after the start of the intervention and during post-testing. And these differences are shown to be durable, given that the experimental group performed significantly better even 16 weeks after the end of the intervention. The intervention was replicated in humanities education (Chapter 5). As for the reasoning test scores, posttest scores improved significantly for one experimental group compared with the control group, but not for the other. This difference remained durable at least for one of the two experimental groups. In conclusion, no far-transfer occurs when solely n-back is trained. When n-back training is combined with other training

tasks, then the results appear to be ambiguous with respect to far-transfer to reasoning.

2.3.2 Non-school-based interventions

The second group includes studies of participants who do not participate in school-based training, or use general reasoning tests to measure far transfer to reasoning, rather than school-tests. Frequently-used tests for reasoning include the matrix reasoning tests. Six non-school-based interventions were included in the review. All interventions used a variant of n-back (single or dual n-back) as the sole training task. Dual n-back is a variant on the n-back task in which 'dual' means that visual and oral stimuli were simultaneously presented to the participant. Four out of six interventions observed no far transfer to reasoning posttests (*Mdn d*= 0.09*, range* -0.32 – 0.80). For instance, in Colom et al.'s study (2013) 28 female students participated in the experimental group which trained on three variants of n-back in 20 sessions. While the training group results demonstrated training improvements in the n-back test, the results showed that, relative to a passive control group, there was no significant transfer to reasoning. In Redick et al. (2013) the experimental group (n=24) was trained by the use of dual n-back. The experimental group's results demonstrated no positive far-transfer effects to any of the tests. Thompson et al. (2013) trained dual n-back.

One study (Jaeggi et al., 2010, Exp 2) found significant far-transfer effects to reasoning. 46 university students were trained on an n-back or dual n-back task during the course of four weeks. 43 students were assigned to the control group. The results yielded significant improvements for the trained groups for each matrix reasoning task.

Jaeggi et al. (2011) obtained ambiguous results for far-transfer effects to reasoning with spatial n-back training. Analysis revealed a highly significant interaction between training and measures of Gf compared with low training gain subgroup scores. This significant positive correlation between improvement on the training task and improvement to Gf suggested that the greater the training gain, the greater the

transfer. Whole group test results demonstrated no transfer effects of the training to measures of *Gf*. To conclude, in the majority of studies, far transfer effects to measurements of *Gf* did not occur.

2.3.3 Age

In this section, we analyze the far-transfer results to reasoning in three groups. The first group focuses on children in primary education. The second group of studies reports results on adolescents in secondary education. The third group of studies addresses far-transfer results for young adults in tertiary education.

Far-transfer results remain ambiguous when controlled for age. One study (Jaeggi et al., 2011) intervened in a sample of primary school children. One study reported significant results (*range of Cohen's d* = 0.55 – 0.80*).* Jaeggi et al.'s n-back training did not affect far-transfer results to matrix reasoning in general, but did observe significant improvements in low-level students. Two interventions (Chapters 4 and 5) trained secondary school students. Both studies reported significant results (*range of Cohen's d* = 1.04 – 2.20). Far transfer only occurs when WM in combination with reasoning strategy is trained. Five interventions were conducted in a sample of participants who attended university. One of those studies (Jaeggi et al., 2010) observed far transfer to reasoning, while three studies (Colom et al., 2013; Redick et al., 2013; Minear et al., 2016) did not (*Mdn d =* 0.16*, range =* -0.32 – 0.80). None of the studies controlled for motivation during training and testing.

2.3.4 Passive vs. active control groups

Passive control groups are groups in which participants receive no (WM) training and have no contact with the experimenters. When experimental group test results were compared with passive control group posttest scores, far transfer to reasoning was observed in one of four studies *(d* = 0.10)(i.e. Jaeggi et. al, 2010, Exp 2). In this study, the groups were matched based on age, gender, pretest performance in n-back and Bochumer Matrizentest (BOMAT). The experimental groups

were trained for 4 weeks with a total of 20 sessions. The results yielded significant improvements for the trained groups for each matrix reasoning task, while the control group showed a marginally significant re-test effect on the BOMAT, but not on the Raven's Matrices (RAPM).

Active control groups are defined as practicing an alternative type of task, unrelated to WM, or a low-level WM task. When active control groups were used, significant improvements occurred for the experimental groups in two out of six studies (*Mdn d* = 1.04, *range* = 0.55 – 2.20). These studies involved combinations of n-back training with other training methods, such as reasoning strategy training (Chapter 4, Exp 1; Chapter 5). Considering the nature of what are called 'active control groups', which are groups that have been trained in low-dose WM levels in contrast to passive control groups, one might expect that the WM levels of these groups should improve. We therefore consider it much more difficult to find significant results when experimental groups and active control groups, contrary to passive control groups, are compared because of the active control group's maintenance of WM during the experiment, as shown by the research results.

2.3.5 Solely n-back training vs. n-back as part of the training

The last section provides information about the role of n-back tasks in the interventions. First, interventions are described that solely use n-back. The results of these tests are then compared with passive or active control group test results for reasoning.

The results of solely n-back training are usually compared with passive control group test results (5 out of 8 interventions). In one of five interventions, the training resulted in significant transfer results to reasoning compared with control group test results (*d* = 0.10)(Jaeggi et al., 2010). Compared with active control group test results, the intervention did not result in significant posttest score differences in reasoning. Therefore, the main finding is that when solely n-back or WM training was used, no far-transfer to reasoning was observed. This observation supports the review results of Redick et al. (2015) and Melby-

Lervåg et al. (2016), who also did not find significant transfer from WM training to reasoning, stating that there is no advantage of n-back training to academic achievements.

N-back as part of a broader training design usually occurs in practice-based settings. We found two studies containing n-back as part of the training and both studies generated significant results (*range of Cohen's d*= 1.04 – 2.20). In Chapters 4 and 5, we added another WM task (Odd One Out) and a reasoning strategy task to the training. When only WM training (n-back and other types of WM training) was performed then we find no far transfer results to reasoning in the posttests compared with the active control groups. But we did find far-transfer results to reasoning when reasoning strategy training is included in the training. Accordingly, one could reasonably conclude that far-transfer results caused by the latter training design are mainly attributed to the reasoning strategy training and not to n-back training specifically or WM training in general. We elaborate on this in the discussion and conclusion section.

2.4 Discussion and conclusion

This review contributes to the literature on the far-transfer effects of n-back training to reasoning, and addresses its implications for school-based training. We focused on two types of training practices. First, we review school-based n-back training and its implications for school-based reasoning achievements. There are only four interventions that have implemented n-back training into school-curricula. Three of the four n-back interventions were trained in combination with other WM-tasks or in combination with reasoning strategy tasks. We therefore also included in our review other studies that address far transfer between n-back training and reasoning within school age groups, but which are not incorporated in school curricula. By including the non-school-based type of studies, we are able to analyze if n-back can affect general reasoning achievements. If the results then suggest that these general reasoning achievements are improved and can be replicated, they can be a

prerequisite for school-based types of n-back training. If not, then the accountability of school-based n-back training to improve reasoning decreases. Many studies lack a critical analysis of the mechanisms that cause reasoning achievements through WM- or n-back training, such as skill internalization, other than WM, executive attention, updating and focus-switching (see also: Lilienthal et al., 2013, p.135). In this Discussion and conclusion section, we therefore specifically focus on outcomes in problem-solving by comparing the characteristics of the studies for (non-) school-based interventions, age categories, various types of n-back, and active or passive control groups.

In school-based settings, no transfer effects were measured when solely n-back was trained (see Table 2.4 for effect sizes). But when WM training is combined with reasoning strategy training, then significant improvements did occur in reasoning test results. Therefore, one might question the role of n-back in this transfer to reasoning and whether reasoning strategy training is primarily responsible for the transfer results. In this regard, it might be the case that WM (improvement) functions as a subsystem that supports improvements in reasoning through reasoning strategy training. The dynamics of n-back and reasoning strategy training should therefore be further investigated. In Experiment 2 of Chapter 4 in this dissertation, we lift a corner of the veil by comparing n-back and Odd One Out training groups with active (reasoning strategy) and active (low-dose training) control groups. Reasoning-strategy training was observed to be far more rewarding than WM or low-dose training, the latter two groups' reasoning posttest scores not being significantly different from each other.

Another explanation for the absence of transfer effects of n-back training can be the nature of the samples that are included in the interventions. For instance, five of seven studies in the young adult group used samples of university students. One might consider that this rather homogeneous group already has above average WM-levels, which has been theorized to limit WM-training benefits. Namely, according to compensation account theories, students who are already functioning at

high or optimal levels will benefit less from cognitive interventions, because high cognitive baseline levels leave less room for improvements (e.g. Titz and Karbach, 2014, p.861). The cognitive baseline levels of these samples are comparable to the sample in Chapter 4, Experiment 2 (students of preparatory academic education), in which no far transfer was observed compared with active control groups. In fact, Jaeggi et al. (2011) have demonstrated important individual differences that determine training and transfer. They found that transfer only occurred when children considerably improved on the training task, giving notice of possible far-transfer gains with groups of participants with below-average WM-levels. In this case, Jaeggi et al. (2011) argued for a more individual approach in determining what training regimes and conditions result in the best transfer effects for different WM-level groups.

While interventions in secondary and tertiary education tend to be implemented in homogeneous groups with above average WM-levels, groups in primary education tend to be very heterogeneous. Not only do groups consist of children with diverse levels of general cognitive ability, but also of children whose trajectory of development of reasoning strategies is also quite individual. Therefore, differences in test scores based on questions that presuppose on procedural-based strategies and memory-based strategies could be explained by the mechanism that children of these age groups increasingly shift from 'more procedural-based to more memory-based strategies' (Titz and Karbach, 2014, p.860).

Motivation might be an additional key to understand the training to far-transfer tasks, because motivational aspects appear to influence the activation of cognitive control (Crone and Dahl, 2012, p.639). As such, WM-training is to a considerable extent based on the training of top-down attention ability, in terms of WM being attention focused on an internal representation.

Attention has two components: (1) An inclusionary component consisting basically in the content in focus, and (2) an exclusionary component consisting mainly on the inhibitory control of interference. Much of the variability evident in the literature seems to stem from the imponderable control and confounding of those two components. WM is

demonstrably enhanced by concentration and weakened by distractibility. Reasoning ability behaves in the same opposite ways inasmuch as it depends on WM.
(J. Fuster (2015, June 10), Personal interview)

We have found no evidence of intrinsic motivational aspects integrated in the study: rewards for continuous improvement during training were shown to be absent during interventions; the majority of non-school-based participants were given a fixed monetary reward every week or at the end of the study regardless of improvements; the majority of participants arguably already had high levels of WM and reasoning skills. We propose that future research should investigate the role of motivation in WM-training to far transfer.

Motivation could also play a role when experimental-group posttest results are compared with either passive or active control group test results. Passive control groups are considered to be pre- and post-tested without receiving any form of training during the intervention period. Active control groups received a form of training that relates to reasoning. Active control groups usually train on low-dose WM tasks. For instance, when dual n-back is trained by the experimental group (high-dose n-back), then control groups received single n-back (low-dose) or updating tasks. This difference could theoretically lead to larger differences in reasoning achievements when experimental group test scores are compared with passive control group test results, as well as with active control group test results, due to the absence or presence of a training type that could affect reasoning achievements. Consequently, larger training differences should theoretically be observed between experimental group and passive control group interventions. However, this is not the case. Two of the six passive control group interventions observed far transfer to reasoning posttest results. Two to three out of eight active control group interventions observed far transfer to reasoning posttest results. But when we focus on studies that only train n-back, we observe transfer in two of the six passive control group interventions. We observe transfer of nil to one in four active control

group interventions. Future interventions, therefore, might reevaluate the value of an active vs. a passive control group in research designs.

So what implications for reasoning skills does n-back training have when implemented in school-based curricula? The answer to that is ambiguous. The effects tend to be non-significant. This corresponds with previous reviews (Au et al., 2015; Melby-Lervåg et al., 2016; Redick et al., 2015). First of all, we did not find any school-based intervention that solely trains n-back and tests for reasoning. So, evidently, school-based research of n-back in relation to reasoning remains a 'tabula rasa', and should be further explored. But contradictory results of n-back to reasoning subtests of intelligence (e.g. matrix reasoning) should not hold back school-based n-back reasoning, while the two types of reasoning are not interchangeable. While matrix reasoning tests assess Gf (i.e. solving novel problems independent of past knowledge), school-based reasoning tends to make use of crystallized intelligence as well: as the majority of school curricula are based on practical and theoretical knowledge, school-based reasoning tests integrate knowledge and reasoning into content-based reasoning questions which students are not able to answer without using short and long-term memory. This could mean that the chances that standard n-back training will improve school-based reasoning achievements decrease even further. Future research should therefore determine what training designs result in transfer to school-based reasoning achievements.

Motivational aspects need to be addressed before school-based n-back training can be properly assessed, since WM is demonstrably weakened by distractibility and enhanced by attention (e.g. Kane and Engle, 2002; Lilienthal et al., 2013). In school practice, this means investigating who can benefit from training. When motivation improves the focus of attention and, in its turn, predicts near- and far-transfer effects, future research should determine what motivational aspects should be included in the training to serve the needs of specific groups. Different motivational aspects could be determined by accounting for

age, motivation, learning disabilities, baseline WM capacity, and baseline reasoning skills. Furthermore, underachieving could be caused by distractions. Therefore, it has to be considered how distractions can be minimized and included in a potentially effective methodology. The possible extent of these motivational issues makes school-based n-back training a challenging field of research.

Since n-back training tends to generate non-significant effects to improvements in reasoning, we review the literature on the effects of WM-training combined with metacognitive training (or: reasoning strategy training) in Chapter 3.

3 Is working-memory capacity and metacognitive training effective in enhancing school-based reasoning achievements? A synthesis of the research[7,8]

Summary

Secondary school-pupils frequently underachieve in tests which require problem-solving skills. Training of working memory (WM) may improve problem-solving skills. We review primary and secondary school-based interventions to evaluate the capability of WM-training to enhance problem-solving skills. Two aspects of WM appear to affect reasoning skills. WM capacity plays a role in short-term storing and in manipulating information. Meta-cognitive WM concerns the storage of acquired knowledge of problem-solving skills in long-term memory. We conclude that both aspects could improve reasoning skills. Metacognition causes students to achieve significantly better compared with non-trained groups regardless of training type and intensity. Few studies are found that focus on school-based interventions. Furthermore, in line with earlier reviews, many studies lack a critical analysis of the mechanisms by which WM causes reasoning achievements in classroom environments. The question whether transfer effects to domain-specific reasoning is expected remains a topic for further research.

[7] Chapter 3 is based on: Ariës, R.J., Ghysels, J., Groot, W., Maassen van den Brink, H. (2016). Is working memory training effective in enhancing school based reasoning achievements? A synthesis of the research. TIER working paper series 17/03. ISBN: 978-94-003-0118-4.

[8] We gratefully acknowledge the contributions of Joaquín Fuster, UCLA Semel Institute for Neuroscience & Human Behavior, and Allyson P. Mackey, MIT McGovern Institute for Brain Research, who supported and gave extensive feedback on the description of the mechanisms and implications of the training methods.

3.1 Introduction

In Chapter 2 of this dissertation we found that n-back training tends to generate non-significant effects to improvements in problem-solving. Now, in Chapter 3, we therefore conduct a review of studies on WM capacity and metacognition. It specifically focuses on school-based WM training and achievement outcomes with students with normal learning behavior, as well as on two aspects of WM that are suggested to improve reasoning achievements.

Working memory (WM) training, the training of the brain function that reorganizes and manipulates retained information, may enhance a broad variety of academic skills (Au et al., 2015). In recent years, research on WM has provided evidence that WM training could improve problem-solving skills by enhancing the ability to provide solid analyses on, and answers to novel problems (e.g. reasoning questions) (Conway et al., 2003; Kramarski and Mevarach, 2003; Cheshire et al., 2005; Jaeggi et al., 2008) and improve test results (Agin, 2001; St. Clair-Thompson et al., 2010). The evidence on this is mixed, as other (review) studies have reported no significant improvements caused by WM training (e.g. Redic et al., 2015). Also, the excitement about the prospects for *far transfer*, which are 'performance benefits in outcome measures that are contextually, structurally and superficially dissimilar to the trained task' (Beatty and Vartanian, 2015, p. 1), has resulted in studies that investigate the enhancement of reasoning with novel information by trained participants. This is based on the view that WM is the main brain function that is activated when reasoning takes place, and on its (prospected) ability to reason and solve new problems, independent of any past knowledge. But few WM training strategies have been specifically developed and evaluated to improve problem-solving skills and reasoning test results in primary and secondary education (Kramarski and Mevarach, 2003; Mevarach and Kramarski, 2003; Jolles, De Groot, Van Benthem, Dekkers, De Glopper et al., 2006; Alloway and Gathercole, 2009; De Jong, Van Gog, Jenks, Manlove, Van Hell et al., 2009; Lee et al., 2009; Redick et al., 2015).

WM plays a crucial role in developing problem-solving skills because it underlies several cognitive abilities, including logical reasoning and problem-solving. In this regard, WM is the activation of cortical cognitive networks (cognits), with subcortical support, to selectively attend to recent information for a prospective decision, choice, or action to solve a problem or to reach a goal. Two aspects of WM affect the process of problem-solving: (1) WM temporarily stores and manipulates information during complex cognitive activities, such as reasoning. Holmes et al. (2009) define WM as the cognitive system that provides temporary storage and manipulation of information in the course of complex cognitive activities. According to Olesen et al. (2004), the amount of information that a person can retain is an important factor for problem-solving: WM needs to change its content constantly and rapidly because problem-solving abilities often require processing different types of information, swiftly switching from one to another cognitive task, and simultaneously solving diverse problems (Goldberg, 2010). The capacity of WM could therefore empower reasoning abilities (Sweller, 1988; Bull and Scerif, 2001; Klingberg et al., 2002; Holmes et al., 2009). Adaptive training of WM or related executive functions can lead to substantial and sustained enhancement of (initially poor) WM in children. (2) Prescriptive knowledge, the generic memories of effective ways to solve higher-level cognitive problems, is stored in WM (Gazzaniga et al., 2009; Klingberg et al., 2009; Goldberg, 2010). Experts, unlike novices, possess generic memories-developed schemas or patterns to solve problems (Sweller, 1988). This 'pattern recognition' is based on learning processes which underlie repeated exposure to similar contexts that originate in matched response strategies. According to Goldberg (2010), a transition occurs from the absence of effective behavior of WM to the formation of effective behavior in reasoning processes in order to address reasoning questions more effectively and efficiently when meta-cognitive components are added to training strategies. Therefore, meta-cognitive training causes WM to form blueprints, and to plan for rational analysis and analytical methods (Gazzaniga et al., 2009; Goldberg, 2010).

Despite the lack of consensus about the effect of WM-training, it is concerning that many of the existing studies and reviews (e.g. Shipstead, Redick and Engle, 2012; Melby-Lervåg and Hulme, 2013; Titz and Karbach, 2014) on WM and reasoning cannot be replicated in compulsory education because of its specifics and deficits. Van der Sluis et al. (2007), for example, argue that in most studies involving children 'the executive performance in samples of normal children is compared to that of children in clinical samples, for example children with learning difficulties' (p.430) or ADHD, and there are correlational studies (e.g. Bull, Johnston and Roy, 1999; Bull and Scerif, 2001; Redick et al., 2015). Furthermore, many experimental studies primarily use Fluid or General Intelligence tests; Shipstead et al. (2012), for example, argue that the training and testing of basic cognitive functions is a far cry from the training of actual school-based reasoning skills. In a more recent review of the effects of WM-training on a broad range of academic skills, Redick et al. (2015) support this claim by concluding that WM-training alone produces 'no advantage for academic or achievement-based reading and arithmetic outcomes' (p.1). Nearly all studies included in their review have used samples of students with learning difficulties, which could affect learning outcomes compared with normal learning behavior. Furthermore, studies increasingly indicate that the implementation of subject-oriented content and processes is key to enhancing WM-training models (Vos, 2001; Van der Sluis et al., 2007; Lee et al., 2009). Other, more recent studies suggest that training on WM tasks correlates with (short-term) improvements in academic achievements in reasoning (e.g. Au et al., 2015; Stephenson and Halpern, 2013). Moreover, many experimental and correlational studies have tested children in their early years and in late adolescence (18-21 years' old), and are thus not focused on primary and/or secondary education. Early and middle adolescence is a period of high vulnerability to reward orientation, risk-taking, and problems in the regulation of affect and behavior, and cannot be compared to late adolescence (e.g. regulatory competence) according to Steinberg (2006). Our synthesis of the literature distinguishes itself from these previous reviews on school-

based WM-training by specifically focusing on achievement outcomes with students who have normal learning behavior, as well as on two aspects of WM that are suggested to improve reasoning achievements.

This review addresses these issues by reviewing studies that focus on primary and secondary education in order to analyze whether WM-training is effective in education settings to develop problem-solving skills. The search strategy on which this review is based is focused on school-based interventions, as well as on school tests

This chapter proceeds as follows. First the search strategy is discussed in section 3.2. The following section (3.3) presents the findings. The last section (3.4) discusses and concludes.

3.2 Method and search strategy

This chapter focusses on school-based WM capacity interventions as well as on (metacognitive) reasoning strategy interventions. We only include articles that directly link WM training interventions to the children's educational problem-solving achievements. The review uses peer-reviewed journal articles from 1990 to 2015. Journal articles have been selected via databases ERIC (Educational Resources Information Center) and Psycinfo (see Figure 3.1 for the flowchart). Google Scholar has been consulted to search for more articles.

The following keywords were used as key search values: 'working memory (capacity)', '(primary/secondary) education', 'reasoning strategies', 'metacognition', 'school-based'. The combination of at least three of these keywords yielded 145 articles. We found that many articles did not address the school-based and content-based approach that we were looking for. Therefore, we added the following search terms to address the school-based approach more specifically: 'adolescence', 'young adults', 'executive function training', 'reasoning skills', 'critical thinking', 'cognitive learning', 'brain-based learning', 'academic achievement', 'compulsory education', 'content-based'. The search strategy yielded 73 articles. We extracted 51 articles based on the

following inclusion criteria: English or Dutch language, participants in randomized groups between 5 and 20 years old, publication year 1990-2015.

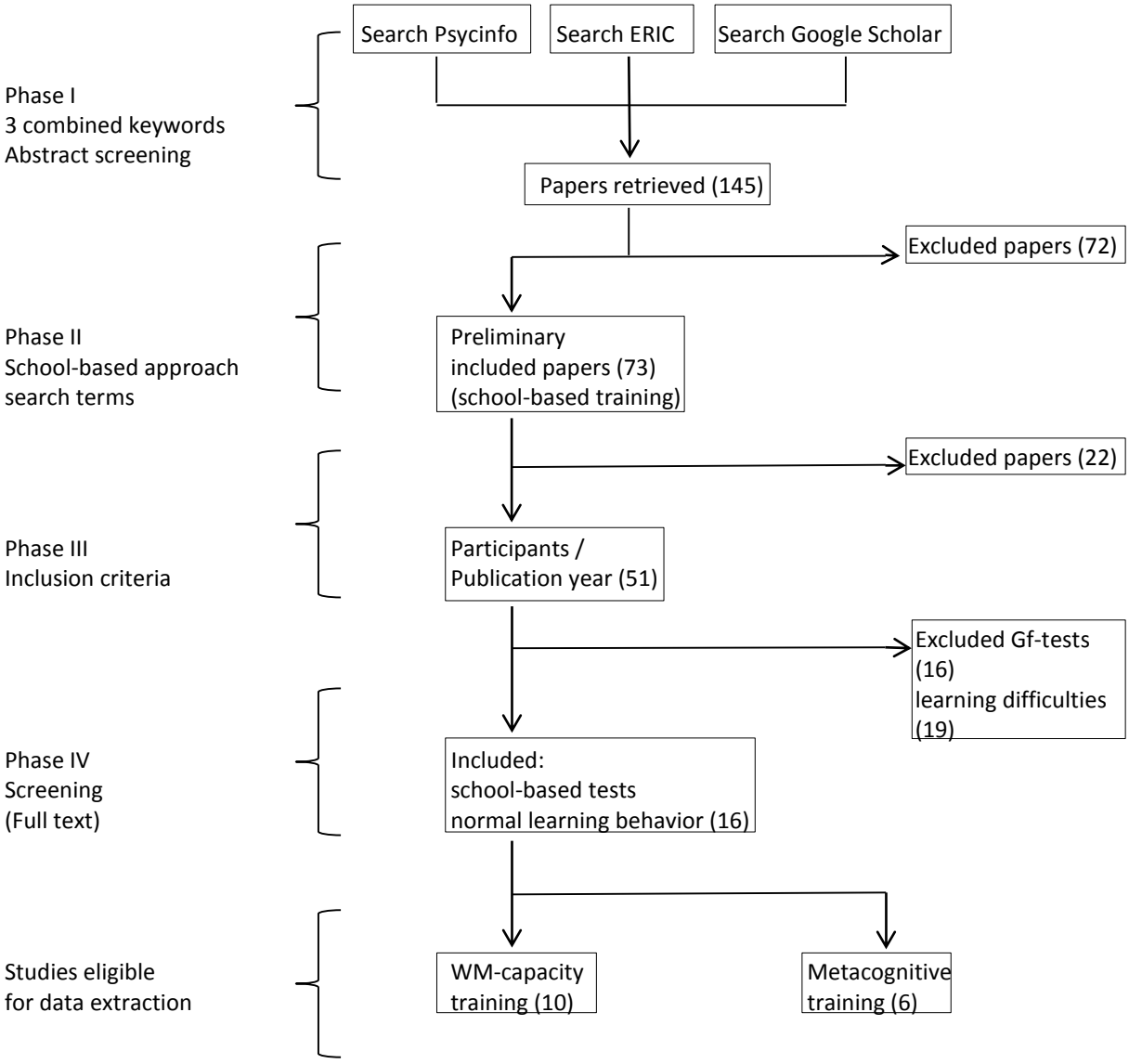


Figure 3.1 Flowchart of the literature review process

The greater part of the literature describes school-based interventions, but did not use school tests to measure improvements in reasoning skills. In fact, they generally use general intelligence tests or tests for general reasoning. Also, much of the literature consists of correlational studies which do not allow for causal inferences. Although WM capacity and general intelligence are generally acknowledged as

strong predictors of academic achievement, we exclude studies that use these general intelligence tests (16 articles). This is due to the absence of strong correlations between fluid intelligence training tasks (which is the capacity to reason and solve new problems, independent of any past knowledge) and school-based reasoning, as *far transfer* of WM training to content-based reasoning abilities seems very limited (e.g. Titz and Karbach, 2014). Correlational studies and tests between normal children and children with learning difficulties are also not included in this review (19 articles), as we focus on training and normal learning behavior. We have found a total of 16 articles (10 articles concerning WM capacity training, 6 articles concerning metacognitive training) eligible for inclusion in the systematic review (Tables 3.1 and 3.2). The selected articles have been published in peer-reviewed journals, and used (semi-)randomized groups of children aged 5 to 17.

3.3 Results

The current consensus in neuroscientific research is that developing a problem-solving strategy and the training of the cognitive processing capacity could be primarily accredited to WM functions. WM, the cognitive system in the prefrontal cortex of the brain, is argued to play a crucial role in developing problem-solving skills because it underlies several cognitive abilities, including logical reasoning and problem-solving (Vos, 2001). As described in the Introduction to this section, two aspects of WM affect the process of reasoning: (1) WM temporarily stores and manipulates information during complex cognitive activities, such as reasoning (WM capacity). (2) Prescriptive knowledge, the generic memories of effective ways to solve higher-level cognitive problems, is stored by WM (Metacognition). We therefore categorized each function in either WM capacity or Metacognition.

3.3.1 Working-Memory Capacity

The selected studies that address WM as a training intervention to specifically enhance reasoning skills in school-based settings are

described in Table 3.1. Many different training tasks have been used in the studies with training periods that highly differ in duration. Almost all interventions have been conducted in (non-)compulsory education that precedes secondary education, ranging from preschool to middle school. Furthermore, in almost all interventions participants were tested for verbal or mathematical reasoning.

Holmes et al. studied the improvements in mathematics and reading performance with training batteries in two studies. The first study (Holmes et al., 2009) included twenty-two 8 to 11 year-old students in primary education and tested for mathematical ability. In a 7-week training period (20-35 sessions) a randomized experimental group had to perform the tasks of the AWMA training battery (see Alloway and Gathercole, 2008). Testing was conducted prior to and directly after the training period and again 6 months after the end of the training on standardized tests for word reading and mathematical reasoning which, according to Holmes et al., would predict academic achievement. The analysis was conducted by using a series of MANOVA's and ANCOVA's. The authors reported significant improvements on the posttests compared with a control group (20 participants) who completed the non-adaptive version. But, as Titz and Karbach (2014) had already noted: 'they did not include the control group into the follow-up measurements' (p.859) which could attribute the improvements solely to maturation over the course of 6 months. Furthermore, the sample size was very small. The second study (Holmes and Gathercole, 2014) included 22 (Trial 1) and 50 (Trial 2) students in primary education. Experimental groups, specifically selected because of low-academic performance, were trained with Cogmed Working Memory Training (CWMT) batteries in 20-25 sessions (in 20 training days) and compared with a control group (50 children) who received no WM training. Pretests and posttests contained national standard mathematics and English tests. Grade 5 and Grade 6 participants showed significant improvements in math test scores compared with control group test scores using MANOVA's. In English test scores, only Grade 6 students improved significantly compared with the control

group. These results indicate that transfer between training on standard WM tasks and school-test results may occur, specifically in the case of low-achievers. The experimental group was compared with a passive control group, which might explain the differences in test results.

St. Clair-Thompson et al. (2010) used memory strategy training (Memory booster) to explore the relationship between WM and children's scholastic attainment. Subgroups of 254 students aged 5-8 in primary education took standardized tests in reading, arithmetic and mathematics. One class from each of the five selected schools was assigned to the experimental condition. The other class was assigned to the control group. Experimental group pretest and posttest data (117 children) were compared with control group test results (137 children) after an 8-week intervention period, and again 5 months later. The control group received no Memory Booster training. No improvements on both posttests were observed in the standardized tests (ANOVA's). St. Clair-Thompson et al. argue that generalizing strategies from one domain to another is not applicable to children in these age categories. However, the use of domain-specific content in WM-training might be able to address this.

Alloway (2012) used an interactive training method based on the Jungle Memory and tested for improvements on vocabulary, spelling, and arithmetic. Jungle Memory is an interactive computer program and contains three games that include spatial WM training. The control group received low-dose targeted educational support. At the beginning and end of an 8-week training period the randomized experimental (8 students) and control (7 students) groups (total of 15 participants, mean age 13) were tested. The standardized results were compared using t-tests. The training showed improvements in arithmetical abilities, but these were not statistically significant. They did find significant improvements in vocabulary and math post-training which they attribute to the interactive training regime. But only one math test was administered, and the study used a small sample size which, in general, does not provide strong evidence of effect sizes.

Loosly et al., (2012) used a WM span task as a primary training task in a 2-week randomized field study. The study investigated whether a brief intervention would result in near- and far-transfer effects. The span task was performed by an experimental group of 20 students aged 9-11 in primary education and aimed to improve reading skills. Pretests and posttests were conducted to compare the test results with control group data (40 children) containing standardized tests for reading abilities. One class was assigned to the experimental group condition, while selected children from four other classes formed the control group. The children in the control group were matched with children in the experimental group. The control group did not take part in any intervention. The intervention group received training over a period of 2 weeks during the first school lesson in the morning. The analyses (t-tests and MANOVA) showed that span task training improved text and word reading significantly, but not pseudo-word reading. The study made use of only a small sample size, and the experimental group results were compared with passive control group results, which might explain the differences in the test results because of the novelty of the span test.

Karbach, Strobach and Schubert (2014) compared pretest and posttest data of an experimental (14 students) and a control group (14 students) who were allocated by random assignment. In 14 sessions the students in the experimental group, aged between 7 and 9 and attending primary education, performed on a battery of adaptive Braintwister tasks. The control group received a non-adaptive low-level version of the same tasks and materials. Before and after the training, they took standardized reading and mathematics tests to study the transfer of WM training to these test scores by using ANOVA on standardized scores. Significant improvements compared with pretest and control group data were found in reading, but not in math. In this regard, Titz and Karbach (2014) pointed to the mechanism that children of these age groups increasingly shift from ‘more procedural-based to more memory-based strategies’ (p.860). This forms a strong argument for differences in test scores based on new mathematical skills questions (procedural-based strategies) and

Table 3.1 Overview of studies assessing the impact of WM-capacity training on school reasoning skills.

References	Sample	Design	Tasks	Measurements	Effects of training
Alloway (2012)	15 students, 13 years, middle school	8-week randomized control group intervention, pretest posttest	Jungle Memory	Vocabulary, spelling, arithmetic	Significant improvement in vocabulary and math post-training, not spelling; Improvement in arithmetic
Ariës (2015)	63 students, 15-16 years, high school	6-week randomized passive control group intervention, pretest posttest	Content-based non-adaptive n-back, odd one out	Reasoning in history school tests	No significant improvements in historical reasoning
Holmes (2009)	42 students, 8-11 years, primary education	7 week randomized passive control group intervention, pretest posttest	AWMA training battery	Mathematical ability	No significant improvements in math reasoning, verbal IQ or word reading
Holmes (2014)	132 students, 8-9 years and 9-11 years, primary education	20-25 sessions, control group intervention, pretest posttest	CWMT/AWMA training batteries	Mathematics, English	Significant improvements in mathematical ability and English
Karbach (2014)	28 students, 7-9 years primary education	14 sessions, randomized control group intervention, pre-post test	Braintwister WM-training battery	Mathematics, reading	Significant improvements in reading, not math
Kroesbergen (2014)	51 student, 5 years, preschool	4 week randomized passive control group intervention, pretest posttest	AWMA battery subtests	Early numeracy skills, general and domain-specific WM-training	Significant improvements in domain-specific, but not domain general training groups
Loosly (2012)	40 students, 9-11 years, primary education	2-week randomized passive control group intervention, pretest posttest	WM span task	Reading, nonverbal intelligence	Significant improvement in word reading and text reading, not in Gf and pseudo-word reading
Nevo (2014)	97 students, 8.5 years mean age, primary education	12 sessions semi-randomized control group intervention, pre-post test	WMP/RAP	Reading	Significant improvements on word fluency and pseudo-word accuracy, not reading
Söderqvist (2015)	42 students, 9-11 years, primary education	5-week passive control group intervention, pretest posttest	CWMT	Reading comprehension, math performance	Significant improvements in reading comprehension and math performance
St Clair-Thompson (2010)	254 students, 5-8 years, primary education	8-week semi randomized passive control group intervention, pretest posttest	Memory booster	Reading, arithmetic, mathematics	No improvements

reading skills (memory-based strategies). These strategy differences can also be explained by expert-novice theories, which state that novices' problem-solving skills differ from experts' problem-solving skills, which can be explained by the experts' acquirement of schemas by which problems are solved. In this case, Sweller (1988, p.259) argues that 'differences in memory of problem states, strategies used and categories into which problems are placed can all be explained by assuming that experts have acquired schemas which play a crucial role in the way they approach and solve problems'.

Kroesbergen, Van 't Noordeinde and Kolkman (2014) performed a battery of AWMA subtests in a 4-week training period (eight 30-minute training sessions) on an experimental group of 5-year old children in preschool. A total of 51 children were selected based on low achievements on two standardized math tests. The children were then randomly assigned to one of three conditions. The control group did not receive extra training. The intervention group received either a domain-general or a domain-specific WM-training. By using pretest and posttest data in comparison to control group data, Kroesbergen et al. found significant improvements in early numeracy by using domain-specific training tasks using ANOVA. Although transfer between the training and numeracy tests was observed, it is commonly believed that early numeracy skills do not predict secondary school reasoning skills. Furthermore, no significant improvements were found between the domain-general and the control condition on the early numeracy test. Test-result differences could be explained by the extra training time which the experimental groups received. Furthermore, the small sample size makes it difficult to generalize training outcomes.

Nevo and Breznitz (2014) performed a combined training of WM and reading acceleration on students (mean age 8.5 years) to improve reading skills. Three study groups received a different combination of the training programs, and their data were compared with control group pretest and posttest results. The training consisted of 12 training sessions, each lasting 24 minutes, in three training conditions per class

(each with 27 children) which were compared with control group data (20 children) with no training in a series of ANOVA's. All training programs significantly improved reading skills, such as word and pseudo-word fluency, but not reading. Nevo and Breznitz conclude that the combined use of a short WM program and a long reading acceleration program is the most effective way to improve reading abilities that are related to scholastic achievement. But the sample size is small, and no characteristics of the participants were analyzed to explain the test differences.

Later, in Chapter 4 of this dissertation, we present the effects of content-based non-adaptive WM-training on reasoning achievements in a secondary school. In that chapter, two subtests of the AWMA WM battery in a 6-week training pretest - posttest design (25 minutes per week) were used. Two classes were assigned to an experimental or control group condition. The participants in the experimental group (26 students) had improved reasoning achievements compared with their pretest achievements, and the passive control group (16 students) in school-test results, but not significantly. This could be attributed to the non-adaptive form of the tasks. Improvements could therefore solely be attributed to more familiarity with historical content among the participants of the experimental group compared with those in the control group.

Söderqvist and Bergman Nutley (2015) investigated the long-term effects of the Cogmed Working Memory Training (CWMT) on math and reasoning with 9-10 year-old children. Nationally standardized tests on math and reading were used to test for improvements. 20 students in a classroom received the CWMT, while the control group (n = 22) received education as usual. After 5 weeks of training (20 training blocks) in a classroom setting under teacher guidance, the primary school students were tested. Baseline results and long term results (2 years later) were compared with passive control group results using linear regression techniques. Significant positive results were achieved on long-term attainments, but not on immediate posttest results. Again, the sample

sizes in this study were small. Also, the long-term positive results of the experimental groups were not attributed to training.

3.3.2 Metacognition

Critical thinking is related to meta-cognition, which Kuhn and Dean Jr. (2004) define as awareness and management of one's own thought processes, or 'thinking about thinking'. One can learn how to think critically by reflecting on critical thinking which makes it increasingly more effective (Kuhn, 1999; Vos, 2001; Kuhn and Dean Jr., 2004). The same qualities can be attributed to the structure of analysis concerning reasoning questions. Meta-cognitive functions are mostly executed by the WM function 'executive control'. Consequently, WM involves executing meta-cognitive processes. Reflecting on one's own thinking process also contributes to an increased interest in the objectives of certain activities, such as argumentation (Kuhn and Dean Jr., 2004). A meta-cognitive process can serve as a long-term learning strategy to develop reasoning skills, because the awareness of the relevance of learning activities enhances the long-term memorization and re-use of these learning activities. Kuhn and Dean Jr. (2004) define this ability as 'internalization'. The selected studies on meta-cognitive training specifically related to reasoning in education are described in Table 3.2.

Four different training tasks have been used in the training studies within training periods that greatly differ in duration. Almost all interventions (excluding Cardelle-Elawar, 1995) were conducted in secondary education. Furthermore, all interventions, except the intervention of Chapter 4 in this dissertation, aimed to improve mathematical reasoning.

Cardelle-Elewar (1995) used Mayer's model and sets of metacognitive training batteries to address the effects of metacognitive instruction on the mathematics achievements of low achievers. Mayer's model includes 'four types of processes or knowledge that are required to solve mathematics problems: translation, integration, planning and monitoring, and solution execution' (Cardelle-Elewar, 1995. p. 82).

Table 3.2 Overview of studies which assess the impact of metacognitive training on school reasoning skills

References	Sample	Design	Tasks	Type of academic achievement	Effects of training
Ariës (2015)	63 students 15-17 years secondary education	6-week randomized control group intervention, pretest - posttest	IMPROVE	Reasoning in history school tests	Significant improvements in historical reasoning
Cardelle-Elewar (1995)	489 students 8-14 years primary and secondary education	1-year randomized control group intervention, pretest - posttest	Mayer's model, sets of metacognitive training batteries	Mathematical reasoning	Significant improvements in mathematical reasoning independent of grade level
Kramarski (2003)	384 students 13 mean age secondary education	2-week randomized control group intervention, pretest - posttest	3 sets of self-addressed meta-cognitive questions	Mathematical reasoning	Significant improvements in cooperative settings in mathematical reasoning
Maqsud (1998)	40 students 15-17 years secondary education	10-week randomized control group intervention, pretest - posttest	Basic problem solving strategies	Mathematics achievement	Significant improvements in mathematic achievements
Mevarach (1997)	265 students 12-13 years junior high school	Longitudinal intervention study semi-randomized control group, pretest - posttest	IMPROVE	Algebra, mathematical reasoning	Significant improvements in mathematical reasoning
Mevarach (2003)	122 students 14 mean age secondary education	4-week randomized control group intervention, pretest - posttest	IMPROVE	Mathematical reasoning	Significant improvements in cooperative settings in mathematical reasoning

489 students (aged 8-14), most of Hispanic origin and low socioeconomic status, were randomly assigned to experimental and control classes, and the students were given a full years' individual training. The control group received no training. Math achievement was measured by 'specially designed criterion-referenced tests consisting of 20 problems for each

grade level' (p.91). Test results were analyzed by t-tests and ANOVA. Pretest and posttest results between and within groups showed significant improvements on learning outcomes in all experimental groups. This indicates the benefits of implementing metacognitive instruction in regular classrooms where a substantial share of students are low-achievers. Because the control group received no extra training, the experimental group received 17 extra training hours.

Maqsud (1998) performed an experiment to improve mathematics achievement and attitudes towards the mathematics of low achievers by using a set of basic metacognitive problem-solving strategies. 40 students (aged 15-17) were randomly assigned to experimental (n=20) and control groups. The metacognitive strategies were individually trained on four mathematics topics in a 10-week training period. The control group was trained by using the conventional teaching method. Math achievements were tested by four short math achievement tests that were constructed by the teachers. By comparing pretest and posttest results, Maqsud concluded that experimental group mathematics achievements improved significantly compared to pretest and control group data. However, it is not known if all classes received the same tests.

The IMPROVE method has been developed and implemented in two studies by Mevarach and Kramarski on algebra and mathematical reasoning. The first study (1997) aimed to design IMPROVE and investigate its effects on mathematics achievement. Three classes implemented IMPROVE, while the non-treatment control group consisted of five classes. A 36-item test, covering rational numbers, identification, operation, order, laws of math operations, was administered. In a full academic year field study, they performed individual training and testing on 99 students, aged 12-13, of junior high school and compared the results with control group data (n=148). Experimental group students improved significantly in mathematical reasoning compared with the control groups. The second study (2003) used the IMPROVE method in both cooperative and individual training setting and compared pretest and posttest results with control groups who were instructed with

worked-out examples of reasoning structures. Control group data were collected based on cooperative and individual settings. The results indicate that students who were exposed to metacognitive training outperformed both control groups. Furthermore, metacognitive cooperative training outperformed metacognitive individual training. The study indicated that lower achievers also gained more reasoning skills with the metacognitive training than with worked-out examples. Furthermore, the way groups are composed and group interaction is structured was important for the training effects. The effects were largest for small groups of 4 participants consisting of an overachiever, middle achiever, underachiever.

Kramarski and Mevarach (2003) studied the effects of 4 instructional methods on mathematical reasoning and meta-cognitive knowledge: cooperative learning including meta-cognitive training (COOP+META); individual learning including meta-cognitive training (IND+META); cooperative learning excluding meta-cognitive training (COOP); and individual learning excluding meta-cognitive training (IND). A group of 384 students, with an average age of 13, were randomly selected from 4 secondary schools, which, in turn, were randomly selected from 15 secondary schools in a single district in Israel. The schools were comparable in size and socioeconomic status. Twelve certified female teachers with a minimum of 5 years of experience taught during the training period. Students were trained in mathematical problem-solving in all classes 5 times a week during a period of 2 weeks. The meta-cognitive training was based on 3 sets of meta-cognitive questions: comprehension questions; strategic questions; connection questions. All students were tested before, during, and after the training period. COOP+META used small heterogeneous groups that had to respond to the 3 sets of meta-cognitive questions. IND+META used the same sets of meta-cognitive questions, but students had to respond to them individually. COOP used small heterogeneous groups that did not respond to math problems by using the meta-cognitive questions. IND used individual response without meta-cognitive questions. Kramarski

and Mevarach found significant differences in mathematical reasoning. COOP+META scored better on posttests than IND+META which, in turn, scored better on posttests than COOP and IND. There were no significant differences between COOP and IND in achievements on posttests. Significant differences were found in fluency in reasoning and flexible reasoning. Meta-cognitive groups achieved significantly better than non-meta-cognitive groups on a transfer task. But there were no significant differences between the meta-cognitive groups and the non-meta-cognitive groups.

In Chapter 4 Experiment 2 of this dissertation, we studied the effects of a 6-week metacognitive history training on reasoning achievements in secondary school history tests. The participants were trained in answering reasoning questions from old history exams by using the IMPROVE method in a 6-week training pretest - posttest design (50 minutes per week). The IMPROVE method consists of a set of metacognitive questions which participants had to address cooperatively. Pretest and posttest data (school tests) were compared with a passive (traditional schooling) and an active (WM-capacity training) control group. Participants in the experimental group had significantly improved reasoning achievements compared with their pretest achievements and with both control groups after 3 weeks of training.

3.4 Discussion and conclusion

This review contributes to the literature on enhancing problem solving-skills by focusing on whether a combined WM-based training method of WM capacity and metacognition is effective and can be implemented in primary and secondary education for students to develop problem solving skills. In this 'Discussion and conclusion' section, we therefore specifically focus on outcomes in problem-solving.

The WM-capacity learning strategy implies the enhancement of WM which is argued to correlate with improved reasoning abilities. The research reviewed above shows that training tasks and training intensity

strongly differ between studies. The participants' ages ranged from 5 to 16 years, and they had attended primary or secondary education from preschool to secondary school. Almost all studies included training of mathematics abilities and/or reading, and all but one study (Chapter 4 of this dissertation) included general (non-content-based) training tasks. 50% of the interventions in the studies that we included in this review resulted in significant improvements in reasoning achievement, regardless of the type and intensity of training or age. Significant improvements occurred more often in reading compared with math tests.

In the Introduction of this chapter we pointed out that the transfer to reasoning achievement is likely to be more effective when participants are trained by using content-based interventions due to near-transfer effects. Near-transfer effects occur when the content of the training batteries is similar, but not identical, to test-content. In this regard, it seems rather peculiar that the study described in Chapter 4, Experiment 2 of this dissertation is one of the few studies that did not result in significant improvements in reasoning achievements. This can be attributed to the absence of an adaptive element in the training tasks, which is considered a core element in WM-training, and to the cognitive levels of the participants in this study who were students in preparatory academic education (the highest track in Dutch secondary education), and therefore already had above average WM levels. The latter argument is in line with compensation account theory, about which Titz and Karbach (2014) state that 'high-performing individuals will benefit less from cognitive interventions, because they are already functioning at the optimal level, which leaves less room for improvement' (p. 861). One study on reading comprehension used content-based training batteries in which near-transfer effects could occur. In the Nevo and Breznitz (2014) study, near transfer effects occurred in vocabulary, but not in reading. Two studies concerning mathematics used content-based training batteries in which near-transfer effects could occur. Alloway (2012) found positive results in mathematical reasoning, but had a small sample size. In

the Kroesbergen et al. (2014) study, content-based training resulted in significant improvements in early numeracy.

All other studies used general training tasks, which could provide evidence of *far transfer* of standardized training to reasoning achievements in school tests. Far-transfer effects occur when the training content is not similar to the tests, yet improvements are found that can be attributed to training. However, the results that we found in studies that contained general training batteries prove to be more ambiguous. With regard to reading, four studies found positive relationships with vocabulary (Holmes et al., 2014; Loosly et al., 2012), spelling (Loosly et al., 2012) and reading comprehension (Holmes and Gathercole, 2014; Karbach et al., 2014; Loosly et al., 2012; Söderqvist and Bergman Nutley, 2015). These results were compared with passive control groups. Two studies (Holmes et al., 2009; St Clair-Thompson et al., 2010) found no results in reading comprehension. Alloway (2012) did find a significant improvement in vocabulary, but not in spelling. The ambiguous pattern of results in the Alloway (2012) study can most likely be attributed to the small sample size. With regard to mathematics, far-transfer effects between general training and tests are less likely to occur. This could be attributed to the specifics of reasoning in mathematics, in which students reason at an abstract level, which can not be compared with reading comprehension or reasoning in the humanities, in which students mainly deal with concrete topics. Two studies (Holmes and Gathercole, 2014; Söderqvist and Bergman Nutley, 2015) found positive relations with math reasoning. Three studies (Holmes et al., 2009; Karbach et al., 2014, St. Clair-Thompson et al., 2010) did not find significant improvements on math reasoning or early numeracy (Kroesbergen et al., 2014).

Most participants attended primary education but not secondary education during the intervention. Therefore, further research is needed to find out whether the existing training designs can be copied to secondary education settings in which course-specific heuristics and meta-concepts are more important than in primary education. The articles included in this review by Alloway (2012), Holmes et al. (2009),

Holmes and Gathercole (2014), and Kroesbergen et al. (2014) all included samples of children with low performance, which could suggest that the results can not be generalized to all children. However, we could not determine whether the results of these studies differed from those using samples of children with normal learning behavior.

Reading comprehension skills play a significant role in the humanities (e.g. history) in interpreting and synthesizing historical sources to build an interpretative case, in which the historical sources frequently contain written documents or, to a lesser extent, data in the form of tables and graphs. In this regard, the reviewed studies could be of value in enhancing reasoning skills in humanities courses.

The meta-cognitive learning strategy trains the management and control over one's own thinking and internalization. The awareness of the relevance of learning activities results in long term memory storage whereby learning activities can constantly be re-used. This is specifically applicable to achievements in general reasoning, but also applies to fluent reasoning, flexible reasoning, and the transfer of reasoning. The reviewed articles show that students achieve significantly better on reasoning tests when meta-cognitive skills are trained compared with students who have had no meta-cognitive training. The findings that are reviewed in the previous paragraphs show that training tasks and training intensity strongly differed. The participants' ages ranged from 8 to 17, and they had attended compulsory education from primary to secondary school. All studies but one (i.e. the study in Chapter 4 of this dissertation) included training on mathematics abilities and used general (non-content-based) training tasks. All interventions reviewed in this chapter resulted in significant improvements in reasoning achievements, regardless of the type and intensity of training or age.

Two instructional designs on meta-cognition have been reviewed: the method in which small heterogeneous groups work together in a meta-cognitive process and the method in which meta-cognition is trained individually. Heterogeneous groups that are trained in meta-

cognition turn out to achieve better results, compared with individual training methods on general reasoning achievements, fluent reasoning, and flexible reasoning. But this significant difference cannot be applied to the transfer of reasoning between different school courses without more research on this specific topic.

The review further suggests that five issues need to be addressed in meta-cognitive training methods for problem-solving structures to be internalized in long-term memory.

1) Participants need to solve problems by using a training method which includes both self-reflection and reflection proposed by group members.
2) As stated in the introduction of this chapter, the evidence suggests that problem-solving structures need to be trained by using content-based knowledge on reasoning questions (e.g. Van der Sluis et al., 2007).
3) Problem-solving skills can best be internalized by training in heterogeneous groups, containing underachievers, overachievers, and average achievers in problem-solving skills. The group is responsible for reflecting on a participants' problem-solving process.
4) Skills are best internalized when a participant follows a step-by-step plan and verbally presents the reasoning process to the other group members.
5) Training needs to be consistently repeated.

When implementing the IMPROVE method, Mayer's model, or subsets of metacognitive training batteries in a cooperative form, and addressing content-based reasoning questions, and repeated (weekly) training, then all criteria are accounted for.

In line with Leinhardt et al. (1994) we suggest that content-based knowledge may be crucial to acquire problem-solving skills because of the alternative views, detailed factual knowledge, broader frames of reference, and other concepts that are frequently used in compulsory

education courses. This contradicts statements suggesting that transfer will occur to content-based problem-solving skills when students are only trained on standardized tests. In fact, studies indicate that the implementation of subject-oriented content and processes is an important element in enhancing the effectiveness of ef and WM-training models (Vos, 2001; Van der Sluis et al., 2007; Lee et al., 2009). More research needs to be conducted to link standardized WM tests to content-based knowledge in which students are trained. Furthermore, the implications of content-based training methods, contrary to standardized training methods, should be studied.

The students participating in the studies include primary school and secondary school aged children. Few WM capacity studies have been conducted on an important part of the target group of the review: adolescents. Most of the research on this matter has been conducted on children and adults. This means that no firm conclusions can be drawn about the training of cognitive reasoning abilities for adolescents. Because research on children and adults proved to have similar results, a tentative conclusion is that similar training methods can enhance adolescents' reasoning abilities. More research on the adolescents' brain in relation to WM-based training methods needs to be conducted in order to provide more insight on this matter.

Another practical problem lies in the development of the adolescent brain, specifically the working level. The training model has to account for differences in working levels between different students, and the use of a longer training period implies that the model should grow as the brain matures. The question is: How? Should a model be developed per level, class, or students' age? How long should students be trained in order to become effective? When is training no longer effective? Can individual training be practically integrated in curricula? These are practical and financial issues that could be obstacles to implementing these training methods into educational practice.

Practical restrictions to implement general WM training in school curricula make it difficult to integrate general and content-based

cognitive training methods in school practice. For instance, to what extent are students motivated when being trained by a general method to enhance content-based reasoning? Can teachers be convinced that they should train students in general reasoning methods, instead of using only content-based interventions? Therefore, research has to be conducted on how to implement content-based problem-solving training in WM training models. With regard to concentration, training methods require that students focus their attention on an internal representation. Attention has two components: (1) an inclusionary component consisting of the content in focus; and (2) an exclusionary component consisting of the inhibitory control of interference. Since WM is demonstrably enhanced by concentration and weakened by distractibility, underachieving could easily be caused by distractions. Therefore, one should consider minimizing distractions as a requirement for an effective methodology. Furthermore, many studies lack a thorough critical analysis of the mechanisms by which WM causes reasoning achievements. For instance, training advantages may be caused by automated skills, other than WM, that have reduced the load on WM. Or training might improve concentration or improve the motivation of the training group. These possibilities are rarely considered but undermine the effects of WM-training on WM capacity enhancements. Future research needs to be conducted on the added value of course-specific training of WM functions compared with domain-general training in order to enhance problem solving skills.

4 Improving reasoning skills in secondary history education by working-memory training[9]

Summary

Secondary school pupils underachieve in tests in which reasoning abilities are required. Working-memory (WM) training may improve reasoning abilities. In this study, we use a WM-based training program based on historical content to enhance reasoning abilities in history courses. In the first experiment, a combined intervention of WM capacity and reasoning strategies is trained and compared with control group data in 'HAVO' (medium-track secondary education in the Netherlands). WM capacity and reasoning strategies of the experimental group improve significantly after 3 weeks of training. Students achieve significantly better in school tests in which reasoning abilities are tested. The gain in reasoning abilities is also significant even 16 weeks after the training program is completed. In the second experiment, WM capacity and reasoning strategies are trained independently and compared with control group data in 'VWO' (upper-track secondary education). Training of WM capacity did not improve achievements in reasoning tests significantly, but training of reasoning strategies did. The results of both experiments cannot be compared because of cognitive group differences between Experiments 1 and 2 (in terms of educational tracks). Nevertheless, these results demonstrate that WM-based learning strategies to improve reasoning abilities can easily be integrated in history curricula. Furthermore, a 6 week training period can significantly improve the reasoning abilities of students in secondary education. The study used content-based, non-adaptive training methods that are based on standardized tests. Cognitive WM test scores indicate a ceiling effect, which can be explained by the absence of the adaptive element in the training.

[9] Chapter 4 is based on: Ariës, R.J., Groot, W. and Maassen van den Brink, H. (2015). Improving reasoning skills in secondary history education by working memory training. *British Educational Research Journal 41*(2), 210-228.

4.1 Introduction

School-tests and exams in history courses frequently contain a large number of reasoning questions. Students not only have to reproduce theoretical knowledge in these tests, but also need to demonstrate their ability to write reasoned arguments (Perfetti, Britt, Rouet et al., 1994). Leinhardt et al. (1994, p.314) define historical reasoning as 'the process by which central facts (about events and structures) and concepts (themes) are arranged to build an interpretative historical case'. This requires analyses, synthesis, hypothesis, generalization, and interpretation (Van Drie and Van Boxtel, 2008). Students in secondary education usually have poor reasoning abilities. Students are frequently unable to: take into account alternative views; use sources extensively; acquire detailed factual knowledge, and a broader frame of reference; judge the past by its own standards; take into account the process of continuity; and understand many substantive concepts (Van Drie and Van Boxtel, 2008). Therefore, students frequently underachieve in school reasoning tests in which a score of 56% is required to pass the test.

Working memory (WM), the cognitive system in the prefrontal cortex of the brain, plays a crucial role in developing reasoning skills because it underlies several cognitive abilities, including logical reasoning and problem solving (Klingberg et al., 2002). WM needs to change its content constantly and rapidly, because problem-solving abilities often require processing different types of information, swiftly switching from one to another cognitive task, and simultaneously solving diverse problems (Goldberg, 2010). As we have already pointed out in Chapters 2 and 3, the capacity of WM therefore correlates with reasoning abilities, and children with poor WM typically make poor academic progress in higher cognitive skills, such as reasoning (Alloway, Gathercole, Kirkwood et al., 2009; Holmes et al., 2009; Süss, Oberauer, Wittmann et al., 2002). Two functions of WM affect this process. (1) The cognitive function of working memory stores and manipulates information during complex cognitive activities, such as reasoning (Fuster, 2005; Goldberg, 2010; Klingberg, 2009). According to Olesen et al. (2004), reasoning abilities are

improved when more information is being stored and manipulated in WM. (2) The meta-cognitive function of WM regulates the pattern recognition of reasoning structures (Gazzaniga et al., 2009). Pattern recognition is based on learning processes which underlie repeated exposure to similar contexts that originate in matched response strategies. According to Goldberg (2010), a transition occurs from an absence of effective behavior of WM to the formation of effective behavior in reasoning processes in order to address reasoning questions more effectively and efficiently. Meta-cognitive training causes WM to form blueprints and to plan for rational analysis and analytical methods. Furthermore, WM contains memories of previous leads to the solution of problems, which can be used and altered to solve new reasoning problems.

Neuropsychological research provides evidence that training of WM improves reasoning abilities (Cheshire et al., 2005; Conway et al., 2003; Jaeggi et al., 2008; Kramarski and Mevarech, 2003). The short-term storage and manipulation of information in WM can be improved by training of the cognitive WM function. Pattern recognition of reasoning structures can be internalized when the meta-cognitive function of WM is being trained. These training methods, when combined, have never been tested in academic studies (Cheshire et al., 2005; Conway et al., 2003; Fuster, 2005; Holmes et al., 2009; Kramarski and Mevarech, 2003; Olesen et al., 2004). Furthermore, a combination of cognitive and meta-cognitive WM-based learning strategies for the enhancement of reasoning skills has never been developed for secondary education, except for mathematical reasoning (De Jong et al., 2009; Jolles, 2010; Kramarski and Mevarech, 2003; Mevarech and Kramarski, 2003; Van der Sluis et al., 2007). Abstract mathematical reasoning skills cannot be compared with contextual reasoning skills that are used in history courses (Perfetti et al., 1994; Torney-Purta (ed.), 1994).

In Chapters 2 and 3, we reviewed the literature on WM capacity and metacognitive training. Now, in chapter 4 we investigate whether combined cognitive and meta-cognitive WM training, based on subject-

matter knowledge in history education, results in an effect on students' achievements in tests for which reasoning abilities are required. We conduct two individual experiments that are based on WM training tasks. First, we use a complete WM-based learning strategy which includes cognitive and meta-cognitive WM tasks. The aim of the training intervention is to investigate the effect of the complete training on achievement in reasoning tests and its durability after completion of the training. In the second experiment, we investigate whether cognitive or meta-cognitive WM training, when independently trained, affects achievement in reasoning tests. The interventions are implemented in the existing history curriculum in which the subject-matter approach of the intervention contributes to a specific and concrete training that may result in an improvement of reasoning achievements (Van der Sluis et al., 2007; Vos, 2001).

The experimental groups performed a battery of cognitive (Automated Working Memory Assessment) and/or meta-cognitive (IMPROVE) tasks after pretesting on reasoning abilities (Alloway et al., 2008; Mevarech and Kramarski, 1997). These tests were modifications of the original tests by Alloway et al. (2009) and Mevarach and Kramarski (2003). After the training they were post-tested to evaluate improvements in cognitive and meta-cognitive WM and reasoning abilities. The data of the experimental groups were then compared with the data of the control groups that were trained by using 'low-dose' WM tasks, such as evaluating reasoning questions using whole-class teaching methods. We describe both experiments in more detail below.

4.2 Experiment 1: Complete training and durable effects

4.2.1 Methods

4.2.1.1 Participants

92 students (38% male) from 4 heterogeneous history classes in higher secondary general education (HAVO) from a school in the Netherlands participated in the study. The school had already allocated students to a class based on a chosen set of courses and school

schedules. 2 classes were used as a control group, and 2 other classes were taken as the experimental group. Experienced teachers had been appointed to the classes prior to the research project. The two male teachers and one female teacher had all been teaching history to this level for over 10 years, and all teachers have Master's degrees in history. All teachers used a conservative instruction method in most of their courses, which contains an introduction to new subjects and the use of dialog to answer questions (more information about this method is given later in the section 'Organization': see 4.2.1.5). The classes were randomly assigned to the control or experimental group. The experimental groups were taught by the same teacher. Therefore, part of the effect could be attributed to 'teacher effect'. The second and third class, each of which was taught by, respectively, the second and third teacher, were appointed to the control group.

The control group and the experimental group both had 2 pupils who were dyslectic, which is important to take into account because of the verbal elements in interventions and tests. The mean age (15.84; SD = .75), and the number of participants that had repeated the class, came from a lower level, or came from a higher level in both groups was approximately the same. The male/female ratio was also approximately the same in both groups, and all students lived in the urban area of Sittard-Geleen. All participants attended the same level of higher secondary general education. These are characteristics of homogeneity according to Paus (2005).

4.2.1.2 Data collection

We used two pretests, an intermediate test, and a posttest to measure the combined effects of cognitive and meta-cognitive WM training for the experimental group. The results were compared with the control group data to control for a natural increase of reasoning skills. Both components of the training were trained in the same intervention in the experimental group. Therefore no independent measurements on

cognitive or meta-cognitive training in relation to reasoning skills were possible.

Pretests were used to measure the homogeneity of the experimental and control groups. Two pretests for meta-cognitive skills and achievements in reasoning questions were analyzed and compared in order to reduce the chances of a type-I error. The intermediate tests and posttests were used for testing short-term effects. Both posttests, conducted, respectively, 8 and 16 weeks after the training ended, were used for testing the durability of the effects.

We conducted tests on two tasks, the n-back and the Odd One Out tasks, in order to measure cognitive WM function levels (Holmes et al., 2009; Jaeggi et al., 2008). The experimental group was trained on both tasks, which were altered in subject-related versions of task standards, during the intervention period. The subject-related tasks used historical nouns and images to measure both the verbal and the visual components of cognitive WM, which are important for historical reasoning. In the n-back task (see Appendix A), the participants saw a series of 20 historical nouns that were successively presented in a PowerPoint presentation at the rate of 3 seconds per stimulus. The task was to remember the sequence in which the nouns were presented. Most of the nouns that appeared were colored in black, but when a green colored noun appeared on the screen, the participants had to remember the previous noun and write it on an answer form within 5 seconds. When participants only have to remember the previous word, then the n-back is called a '1-back task'. When they do not need to remember the previous word, but the preliminary word, it is called a '2-back task'. When a word and a historical picture are shown simultaneously, it is called a 'dual n-back task'.

According to Jaeggi et al. (2008), this form of training engages processes required for the management of different tasks, engages executive processes required for each task, and discourages the development of task-specific strategies and the engagement of automatic

processes because of the variation of n, the inclusion of two different classes of stimuli, and repeatedly changed sequences. Furthermore, these tasks test the storage capacity and manipulation of information in WM, as new information has to be stored and updated in WM.

For the Odd One Out task (see Appendix B), four historical words or pictures were successively shown on a video screen in a PowerPoint presentation. Each word or picture was shown for 3 seconds. Three words or pictures in each series were related to each other and one word or picture was the 'odd one out'. Participants had to remember all the words or pictures within a sequence. After all words or pictures were shown, the participants had to write down the word or picture that was the odd one out on an answer form. This task tests the storage capacity and manipulation of information in the WM because words or pictures have to be stored and processed simultaneously there. The test included two series of four words and two series of four pictures. The participants could score a maximum of 49 points (1 point per good answer) on the cognitive test.

Both the experimental and the control groups were also tested for meta-cognitive WM function levels. The use of meta-cognitive reasoning skills was tested by the analyses of answers on reasoning questions in official school tests that were based on Dutch examination standards. The tests and examination models were compiled and audited by three history teachers, and contained a number of theoretical expertise questions and ten reasoning questions (Appendix C). The individual answers of participants on reasoning questions were analyzed with respect to three criteria for meta-cognitive reasoning skills: contextualization, use of sources, and argumentation (Van Drie and Van Boxtel, 2008). The participants' use of these criteria, but not the correct responses to the reasoning questions, were rated by the researcher. Participants could score a maximum of 1 point per criterion per reasoning question, with a maximum of 30 points per test.

Participants' answers on reasoning questions were analyzed according to historical content criteria based on the examination model

used to test achievement levels on reasoning. The same tests as described in the tests of meta-cognitive WM function levels were used for this test, but only correct responses to the reasoning questions were rated by the researcher. The participants could score a maximum of 1 point per criterion per reasoning question, with a maximum of 30 points per test.

4.2.1.3 Intervention

The experimental group was trained on the n-back and the Odd One Out tasks. The same two cognitive WM functions tasks were used for tests on cognitive WM function levels. However, the content differed. The tasks that were trained contained 1-back, 2-back, dual 2-back, two verbal and two visual Odd One Out. Training took place in one history lesson per week over a 5-week period and the duration of the training per lesson was 30 minutes. All tasks contained historical nouns and/or images from the historical content of the course that was taught during the experiment, and in its content, is as such, equivalent to the conservative instruction method. However, the pupils were not tested on substantive knowledge in the tests, but were tested on reasoning strategies and WM capacity.

The standardized IMPROVE-method (Appendix D) was used for meta-cognitive WM training to internalize the structure of reasoning (Mevarech & Kramarski, 2003). This method uses questions that are based on understanding the reasoning hypothesis, relating the skill that is needed to solve the problem with the skills that are already used to solve a previous problem, applying strategies for problem-solving, and reflecting on the reasoning process. These tasks were trained in heterogeneous groups that contained 3 or 4 participants: one low-, one or two middle-, one high-achieving student.

4 reasoning questions (Appendix D) that are all at a higher level of difficulty than school test questions were successively presented on a video screen, and all the participants in the group had to analyze the problems individually. Every problem had to be addressed by one

participant in the group using the IMPROVE-questionnaire. Communication in the problem-solving process between group members plays a crucial role in internalizing the reasoning structure that is trained by using the questionnaire (Mevarech & Kramarski, 2003). The other individuals in the group could give directions when the participant who addressed the problem could not answer one of the questions in the questionnaire. When all the questions in the questionnaire had been answered, a new participant then addressed a new problem. The teachers' role was to participate in each group for the duration of 2 minutes per problem. The tasks were trained in one 50-minute history lesson per week over a 5-week period.

4.2.1.4 Design

This experimental study employed pretests and posttests to compare the effects of the WM-based intervention with conservative instruction methods in order to determine relationships. A nested design was used, as the research was conducted in a single school type in which participants are 'nested' in multiple classes.

The experiment was a between-subjects design in four 4^{th} -Grade classes of Higher General Secondary Education in the Netherlands. The two classes of the experimental group were trained using the cognitive and meta-cognitive tasks to improve reasoning abilities. The two classes of the control group were trained by the method of conservative instruction, which means that the teacher introduced new subjects in new paragraphs and the participants answered (reasoning) questions from the textbook that were verified by interaction between the teacher and all of the students.

Intermediate test and posttest scores from the experimental group and from the control group were analyzed independently by using paired t-tests to measure the effects of the intervention on cognitive WM function levels, meta-cognitive WM-function levels, and reasoning achievements. Furthermore, the test scores of the experimental and the control groups were compared using independent t-tests, as the

Kolmogorov-Smirnov test showed that the distribution in both the experimental, $D(36) = 0.12$, $p > .05$, and the control groups, $D(55) = 0.10$, $p > .05$, suggested no deviation from normality (Field, 2009).

4.2.1.5 Organization

The experimental and the control groups studied the same subjects from the same history textbook during the same time span of three 50-minute lessons per week, prior to and during the training. The content and structure of the lessons prior to the intervention was the same for both groups. The teacher introduced new subjects in new paragraphs, and the participants answered (reasoning) questions from the textbook, and these verified by the interaction between the teacher and all the students. This method of teaching is defined in this research as the 'conservative instruction method'.

5 weeks after the academic year started, all the groups were pretested on WM capacity. Prior to taking the test, the test was introduced to the participants. The topics included historical words from the studied subject. In the next lesson, the first pretest to measure meta-cognitive strategies and reasoning achievements took place. This official school test was, like all other official school tests, compiled by the researcher using the content of the textbook. 2 weeks prior to testing, the school tests were evaluated by the control group teachers. In week 6, the second pretest to measure meta-cognitive strategies and reasoning achievements took place. The reasoning questions in the second pretest were different from the first pretest but the same subject was tested in both pretests. The heterogeneous cooperative groups in the experimental group were compiled after analyses of all pretests to train the meta-cognitive and cognitive WM function. These heterogeneous groups all consisted of one low-, one or two middle-, and one high-achiever.

A 6-week intervention period, including intermediate tests, followed after pretesting in which both groups studied the same subjects in the same time period of three lessons per week. The control groups

studied these subjects using the conservative instruction method. The experimental groups studied these subjects in one-week lesson cycles by using the WM-based method: an introduction to the subject was given by the teacher in the first lesson, and the participants answered the reasoning questions from the textbook. Meta-cognitive reasoning skills were trained in the second lesson. The training method was introduced by the teacher in the first 5 minutes. Reasoning skills were trained in the next 40 minutes (10 minutes per reasoning question) by using the IMPROVE-method (Mevarech and Kramarski, 2003). Process evaluation took place during the last 5 minutes. After the training was introduced during the first 5 minutes, cognitive WM was trained in the third lesson for 25 minutes (Jaeggi et al., 2008). The answers that participants gave to the textbook reasoning questions in the first lesson were evaluated in the remaining 20 minutes.

Intermediate tests were conducted on both groups after 3 weeks of training and the first posttests were conducted immediately after the training ended. The second and third posttests were conducted 8 and 16 weeks after the first posttests.

4.2.2 Results

Pretest analysis accounted for the homogeneity of both groups (Table 4.1). The variances do not differ significantly between the experimental and the control groups, except for the cognitive WM function test. Pretest scores in cognitive WM function show a significantly lower average score in the control group than in the experimental group $t(83.97) = -4.83$, $p<.05$. These pretests were not official school tests, and the groups were tested on different days in the week and during different hours of the day. The test scores, therefore, depend strongly on the motivation of the participants. In addition, when the researcher informed all participating teachers about the organization of the cognitive WM test, it turned out that one teacher did not fully understand the organizational implications of the test. Therefore, the teacher acknowledged that his

group was annoyed, and that the testing had to be postponed for a while, resulting in a shorter time span in which testing could take place. Because this could explain the significant differences between group results, we re-ran the test, but now we split the control group into the two original classes (Control Group 1 and 2) (Table 4.1). These analyses show that cognitive pretest scores between Control Group 1 and the experimental group do not differ significantly. The cognitive pretest scores between the experimental group and Control Group 2, the group that is tested in a shorter time span, do differ significantly. Therefore, we excluded the cognitive test results of Control Group 2, and used the Control Group 1 cognitive test results for further analyses on cognitive test scores.

No significant differences between both groups in their pretest scores for reasoning achievements were measured $t_1(58.99) = .72$, $p>.05$; $t_2(89) = 1.28$, $p>.05$. Both groups scored similarly in both meta-cognitive pretesting and reasoning achievements. Therefore, the analyses of the intermediate tests and the posttests are based on equality between both groups.

The test scores in the second pretest for meta-cognitive strategies ($p<.001$, $r = .38$) and reasoning achievements ($p<.001$, $r = .65$) improved for both groups compared with the first pretests. The improvements can be attributed to the natural learning process and more acquaintance with the tests. The second pretests were used in the subsequent analyses of training effects within and between both groups to account for type-I errors.

Two-way analyses of variance were performed on cognitive WM function tests, meta-cognitive strategies tests, and reasoning achievement tests (Table 4.2). The analyses show statistically significant group differences for all cognitive WM function, meta-cognitive strategies, and reasoning achievements tests between the experimental and the control group. These analyses indicate that the gain in reasoning achievements can be attributed to the training.

Table 4.1 Independent t-test results for pretest scores per group (Experiment 1)

	Control group		Experimental group		Difference and effect		
	Mean	SE	Mean	SE	Mean diff	*p*	*r*
Cognition pretest	29.06	.92	34.83	.75	-5.77	<.05*	.47
Control Group 1	*33.04*	*1.41*	*34.83*	*.75*	*-1.79*	*.23*	*.15*
Control Group 2	*25.52*	*1.28*	*34.83*	*.75*	*-9.31*	*<.001**	*.65*
Meta-cognition pretest 1	8.40	.34	9.46	.50	-1.05	.08	.18
Meta-cognition pretest 2	10.02	.37	10.42	.56	-.40	.54	.06
Reasoning achievements pretest 1	8.85	.36	8.34	.59	.50	.47	.09
Reasoning achievements pretest 2	11.76	.44	10.78	.66	.66	.20	.13

Notes. Mean diff = mean difference; * = p < .05.

Table 4.2A ANOVA results and descriptive statistics for cognition intermediate tests and posttests between groups (Experiment 1)

	Group	Mean score	SD	df	F
Intermediate test	Exp	42.66	6.407	1	27.967*
	Con	33.04	7.480		
Posttest 1	Exp	42.17	5.289	1	45.375*
	Con	30.54	7.989		
Posttest 2	Exp	42.89	5.251	1	41.206*
	Con	30.92	9.050		
Posttest 3	Exp	40.20	4.934	1	31.759*
	Con	29.29	9.809		

Notes. * = significant; F = F-ratio for each univariate ANOVA 'Group'; Sig. = significance of values for each univariate ANOVA 'Group' <.001; Exp = experimental group; Con = control group.

Table 4.2B ANOVA results and descriptive statistics for strategy intermediate tests and posttests between groups (Experiment 1)

	Group	Mean score	SD	df	F
Intermediate test	Exp	16.49	3.906	1	95.553*
	Con	8.72	3.235		
Posttest 1	Exp	16.06	2.807	1	178.112*
	Con	7.02	3.172		
Posttest 2	Exp	18.89	3.612	1	52.417*
	Con	12.70	3.895		
Posttest 3	Exp	12.63	4.030	1	13.182*
	Con	9.32	3.851		

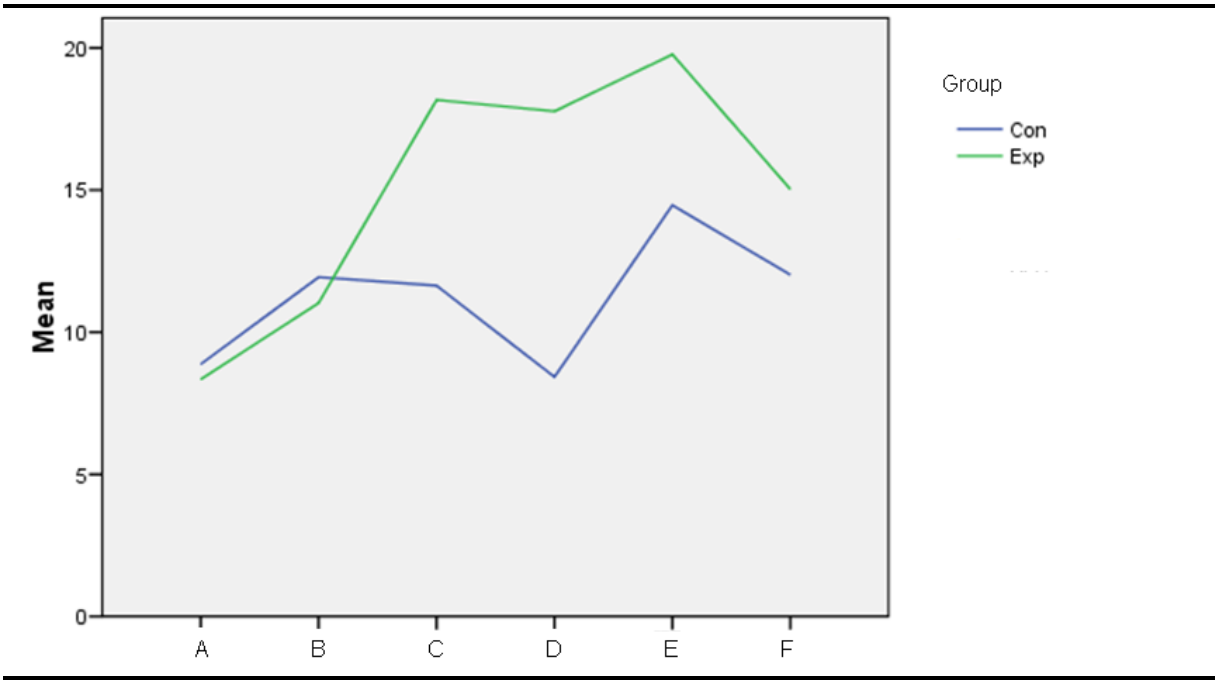
Table 4.2C ANOVA results and descriptive statistics for achievements intermediate tests and posttests between groups (Experiment 1)

	Group	Mean score	SD	df	F
Intermediate test	Exp	18.17	3.468	1	94.464*
	Con	11.64	3.253		
Posttest 1	Exp	17.77	3.465	1	181.920*
	Con	8.43	3.676		
Posttest 2	Exp	19.77	4.278	1	39.310*
	Con	14.47	3.961		
Posttest 3	Exp	15.03	4.084	1	15.917*
	Con	12.02	4.271		

The analyses of test scores during the intervention show a significant effect of the intervention on reasoning achievements in both intermediate tests and posttests for the experimental group (Table 4.2C). Both cognitive and meta-cognitive test scores show significant effects for the experimental group, but not for the control group. We therefore conclude that the combined cognitive and meta-cognitive WM training improves reasoning achievements in school tests significantly because of improvements in WM capacity and the internalization of reasoning structures (Figure 4.1). This effect is significant even 16 weeks after the training ended. 8 weeks after the training ended, the effect size is similar to the effect size that is measured during training, but it declines 16 weeks after the end of training. The control group average test scores are lower for the intermediate tests than for the pretests, but not significantly higher for the posttests when compared with their pretest scores.

The test-score results for both the experimental and the control groups show that the intervention has a significant effect of the intervention for the experimental group compared with the control group which was trained according to the conservative intervention method. This significant effect is shown in the cognitive and meta-cognitive intermediate tests and posttests of WM, and in the same tests for reasoning achievements. Furthermore, the experimental group even

performs significantly better 16 weeks after the training ended compared with the control group.



Notes. Con = control group; Exp = experimental group; A = pretest 1; B = pretest 2; C = intermediate test; D = posttest 1; E = posttest 2; F = posttest 3.

Figure 4.1 Mean scores in achievements tests (Experiment 1)

4.3 Experiment 2: The effects of the components of working-memory training on short-term test results

4.3.1 Method

4.3.1.2 Population

63 students (46% male) in three heterogeneous history classes at Preparatory Academic Education (VWO) school level participated in the study. The school had already allocated students to a class based on a chosen set of courses and school schedules. One class was used as a control group and two other classes were used as intervention groups. Three experienced teachers had been appointed to the classes prior to the research project. To avoid the situation that one or more teachers did not understand the instructions on the organization of the cognitive WM test, as in Experiment 1, all teachers received relevant training prior to testing. The mean age (15.63; SD = .60) and the number of participants

who repeated the class, who came from a lower level, or who came from a higher level was approximately the same in both groups. The male/female ratio was also approximately the same in both groups, and all students lived in the urban area of Sittard-Geleen. All participants attended the same level of preparatory scientific education. All these are characteristics of homogeneity according to Paus (2005).

4.3.1.3 Data collection

We used a pretest, an intermediate test and a posttest to measure the effects of cognitive WM training and meta-cognitive WM training. The results were compared with the control group data to check for a natural increase in reasoning skills. Therefore, the effects of both independently trained interventions on reasoning achievements were analyzed. Pretests were used to measure the homogeneity of the experimental groups compared with the control group. Intermediate testing and posttesting was used for testing short-term effects.

We conducted tests on two tasks, the n-back and the Odd One Out, in order to measure cognitive WM function levels in all groups: the experimental-group WM capacity; the experimental group meta-cognitive strategies (MC); and the control group using the conservative method. All tasks were the same as described in Experiment 1 and expanded by a 3-back task. Both tasks were altered in subject-related versions of task standards which were only trained by the experimental group WM. The participants could score a maximum of 49 points (1 point per good answer) on the cognitive test. All groups were tested for meta-cognitive WM function levels as well, using schools tests and criteria that are described in Intervention 1. Only the experimental group MC was trained in the intervention period, using the same tasks as described in Experiment 1.

4.3.1.4 Design and organization

This experiment used pretests and posttests in order to compare the effects of both WM-based interventions with conservative instruction

methods in order to determine causal relationships. Three classes were included in the experiment, and all three classes were the subject of a separate experiment. The experiment was a between-subjects design in three 4th –Grade classes of Preparatory Academic Education in the Netherlands. One class, the experimental-group WM, was trained using the cognitive training method in one of three history lessons per week. One class, the experimental group MC, was trained using the meta-cognitive training method in one of three history lessons per week. Both groups were trained with the use of conservative training methods in two of three lessons per week as well. One class, the control group, was trained with the use of conservative training methods in all three lessons per week as described in Experiment 1. The content and structure of the lessons prior to the training were the same as described in Experiment 1. This was followed by the pretest to measure working memory capacity and pretests for meta-cognitive strategies and reasoning achievements. The tests were compiled as described in Experiment 1. Intermediate testing and post-testing took place, respectively, 6 and 12 weeks after the start of the intervention.

A Kolmogorov-Smirnov test was used to measure for group variances in pretest and posttest scores for reasoning achievements. The outcome was normal for the experimental group WM, $D(26) = 0.11$, $p > .05$, the experimental group MC, $D(19) = 0.12$, $p > .05$, and the control group $D(17) = 0.12$, $p > .05$. We then used analyses of variance to measure group differences on cognitive WM functions, meta-cognitive WM functions, and reasoning achievements in intermediate tests and posttests.

4.3.2 Results

Pretests are used to control for homogeneity in all groups (Table 4.3). We therefore used analyses of variance on reasoning achievement tests with the factor group (WM, MC, and control), and the posttest scores as the dependent variable (Table 4.4). The analyses revealed that the cognitive training method increased posttest scores significantly

when compared with the control group and the MC training group. The MC training significantly increased posttest scores when compared with the control group. The analyses further showed statistically significant group differences between the MC and the WM groups in all tests, but not between the MC and the control groups, and between the WM and the control groups. Furthermore, statistically significant group differences were revealed after controlling for age, gender and prior education level subsequent to the end of the training period ($F(2,55) = 10.53$, $p < .001$, *partial* $\eta^2 = .28$). Planned contrasts revealed that the MC training method significantly increased posttest scores compared with the conservative training method ($t(55) = -4.58$, $p < .001$, $r = .53$). Gender ($F(1,55) = 7.13$, $p = .01$, $r = .34$), and the pretest scores ($F(1,55) = 9.28$, $p < .01$, $r = .38$) were significantly related to the posttest scores.

Table 4.3 ANOVA results for pretests between groups (Experiment 2)

	Group	Mean score	SD	F	Sig.
Cognition	Con	41.53	5.410	.349	.707
	WM	42.73	4.201		
	MC	42.65	5.373		
Strategy	Con	16.29	6.070	.127	.881
	WM	17.15	5.159		
	MC	16.68	5.608		
Achievements	Con	18.41	6.175	.055	.947
	WM	18.92	5.238		
	MC	18.53	4.869		

Notes. Con = control group; WM = experimental group cognitive training; MC = experimental group meta-cognitive training; F = F-ratio between groups; Sig. = significance of values between groups.

The variances of the between-groups scores do not differ significantly in all pretests. The mean test scores of the experimental group WM, which trains cognitive memory function, improved during the training when compared with the control group test scores, and are significant for the cognitive intermediate tests and posttests. This difference is caused by improvements in the storage and manipulation of information in the WM, and is consistent with prior research on

improving WM capacity with WM function tasks (Süss et al., 2002; Jaeggi et al., 2008; Holmes et al., 2009).

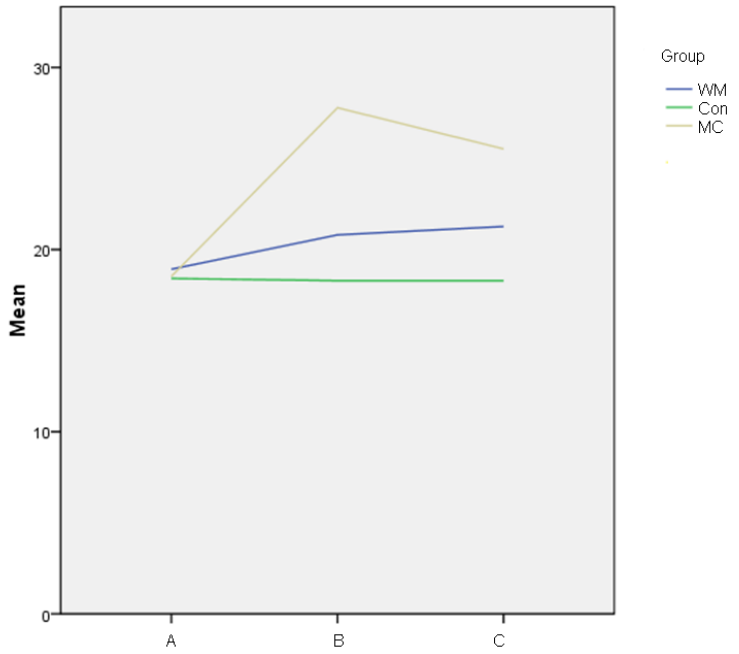
Table 4.4 Multiple comparisons and mean differences in test scores by instruction type controlling for group (Experiment 2)

Test	Comparison	Mean Difference	s.e.	Bonferroni Adjusted 95% CI
Cognition Intermediate Test	Con vs. WM	-6.28*	1.23	-9.32, -3.24
	Con vs. MC	-1.18	1.32	-4.43, 2.07
	WM vs. MC	5.10*	1.19	2.16, 8.04
Cognition Posttest	Con vs. WM	-8.42*	1.96	-13.25, -3.59
	Con vs. MC	-1.87	2.10	-7.04, 3.30
	WM vs. MC	6.55*	1.90	1.88, 11.22
Strategy Intermediate Test	Con vs. WM	-2.32	1.20	-5.28, .63
	Con vs. MC	-8.72*	1.28	-11.88, -5.56
	WM vs. MC	-6.39*	1.16	-9.25, -3.54
Strategy Posttest	Con vs. WM	-2.84	1.60	-6.77, 1.09
	Con vs. MC	-9.61*	1.71	-13.81, -5.40
	WM vs. MC	-6.77*	1.54	-10.57, -2.96
Achievements Intermediate test	Con vs. WM	-2.51	1.34	-5.81, .78
	Con vs. MC	-9.50*	1.43	-13.02, -5.97
	WM vs. MC	-6.98*	1.29	-10.17, -3.80
Achievements Posttest	Con vs. WM	-2.98	1.55	-6.80, .85
	Con vs. MC	-7.23*	1.66	-11.33, -3.14
	WM vs. MC	-4.26*	1.50	-7.96, -.55

Notes. * $p < .05$, where p-values are adjusted using the Bonferroni method; Con = control group; WM = experimental group cognitive training; MC = experimental group meta-cognitive training.

Compared with the control group test scores, reasoning achievements have improved after the cognitive training, but not significantly. Therefore, we can conclude that the cognitive WM training improves WM capacity, but this improvement is not transferred into

significant improvements in meta-cognitive reasoning strategies and reasoning achievements. This conclusion is not consistent with prior research on the transfer between WM capacity and reasoning skills (Klingberg et al., 2002; Conway et al., 2003; Jaeggi et al., 2008).



Notes. Con = control group; WM = experimental group cognitive training; MC = experimental group meta-cognitive training; A = pretest; B = intermediate test; C = posttest.

Figure 4.2 Mean scores in achievements tests (Experiment 2)

The test scores of the experimental group MC, which trains meta-cognitive WM function, improved significantly in the intermediate tests and the posttests for meta-cognitive reasoning strategies and reasoning achievements when compared with the control group test scores. This improvement is caused by the internalization of reasoning pattern structures, through which adapted response strategies can be used to solve reasoning problems (Goldberg 2010). Pattern recognition is trained by cooperative meta-cognitive WM training, in which communication between group members is crucial for the internalization of response strategies (Kramarski and Mevarach, 2003; Mevarach and Kramarski,

2003). This training method does not affect WM capacity, which shows that no transfer takes place between meta-cognitive training of response strategies and improvements in cognitive WM tests. Both WM-based training methods, when trained separately, contribute to higher achievements in reasoning tests. Meta-cognitive training improves test scores significantly, while cognitive training does not improve test scores significantly. A significant difference remains when experimental-group WM test scores are compared with experimental-group MC test scores on reasoning achievements (Figure 4.2).

4.4 Discussion and conclusion

Training in WM capacity and reasoning skills improves test scores. The control groups in both experiments show no significant improvement in posttest scores for WM, reasoning strategies, and reasoning achievements. In both experiments, the training of the experimental groups leads to improved test results as a consequence of better WM capacity and the internalization of reasoning structures. This improvement is observed in the intermediate tests only a few weeks after the beginning of the training. This fast improvement is a confirmation of conclusions from comparable research in which WM capacity and meta-cognitive reasoning strategies are trained (Holmes et al., 2009; Mevarech and Kramarski, 2003; Olesen et al., 2004). Furthermore, when compared with the pretest results in Experiment 1, the control-group test scores show a significant decline in meta-cognitive posttest results - a decline that remained in Experiment 2. This can be attributed to the fact that the control group had not been trained in reasoning skills differently to their pretest training, while the historical content being tested was far more detailed. Content-based reasoning skills training therefore affects long-term memory storage of reasoning patterns and flexibility in reasoning that precipitate the better test scores of the experimental group (Kane and Engle, 2002; Kramarski and Mevarech, 2003). While test score differences in Experiment 1 and in Experiment 2 seem to develop in the same way, the groups of students in Experiment 1 can not be compared

with the groups of students in Experiment 2. The participants in Experiment 1 all participated in the lower-level Higher Secondary Education, while the participants in Experiment 2 all participated in the higher-level Preparatory Academic Education. This could also explain why the Experiment 2 group had a much higher pretest performance than the Experiment 1 group. According to Alloway (2009), there is a correlation between achievements in reasoning and WM capacity. The test-score differences between the experimental and the control groups remain significant in Experiment 1 even 16 weeks after the training ended. This implies that reasoning structures are internalized in the WM and that WM capacity has improved. Therefore, a complete training has a durable effect on reasoning skills that causes higher achievements in school tests that contain reasoning questions. Significant improvements in reasoning achievements are also obtained when only the meta-cognitive WM function is trained in Experiment 2. No significant improvements are observed when only cognitive WM functions are trained. This can be attributed to the nature of training via PowerPoint presentations, which did not train WM functions in an adaptive form. The adaptive element in the training is considered to be highly important for the continuous improvement of the WM capacity of participants. The absence of the adaptive element in WM-training and testing can also explain why, after performing significantly better in the intermediate tests, the intervention groups in both experiments do not perform significantly better in intermediate test scores compared with their posttest scores. Consequently, the cognitive WM intermediate test and posttest scores suggest that there is a ceiling effect with the mean scores of both experimental groups being very close to their maximum scores. This can also be explained by the absence of the adaptive element. This could provide the basis for some discussion on the wider implementation of the experiment. The lack of knowledge on how to write computer programs to improve WM capacity makes it difficult for teachers to implement this training in secondary school training programs. Although the WM tests which have been used are adapted to history content and are presented

as a trial of content-relevant WM training in secondary education, these tests are less reliable in testing and training WM capacity due to their content (Goldberg, 2010; Swanson, Cooney and Brock, 1993). There is no research available to account for the validity of content-based WM capacity training. Furthermore, we used the same type of tests for training as those used to assess training improvements. It therefore remains unclear whether the benefits of the training are limited to the trained tasks, or if they transfer to non-trained tasks as well.

The standardized tests, which have been adapted to historical content in this research, have been used in neuropsychological research for many years to test improvements in WM capacity and reasoning skills. These tests have not yet been translated to content-based training methods for secondary education. This state of affairs can be attributed to a lack of research on adolescent reasoning skills and WM capacity, and to a lack of standardized criteria for content-based usage in secondary education (De Jong et al., 2009; Jolles, 2010). The current study has used a content-based training method that is based on standardized tests to improve adolescents' reasoning achievements. No comparable data are available to compare the outcomes of this (combined) research with other WM-based research on reasoning improvements. Therefore, the role of the semantic contents of these training methods in encouraging the engagement of long-term memory more than other more neutral stimuli would do is a topic for further review.

Meta-cognitive tests to measure reasoning skills also include knowledge questions. It is possible that these knowledge-based questions have influenced reasoning-skills test scores. Knowledge-based questions were included in the tests because the tests were official school tests that require both knowledge and reasoning questions. The level of difficulty of the reasoning questions can also differ between questions and between tests. This is important in the measurement of training duration, but not important in test-score analyses between experimental and control groups because both groups are tested by the same tests. However, not

enough data are available to analyze the effect of training on higher- and lower-achievers or dyslectic students.

Although we used different criteria to measure the participants' performance, the meta-cognitive scores and the achievement scores are derived from the same test. We observed an increase in performance on the cognitive and meta-cognitive tests for the experimental groups, and consequently an increase of achievement in school tests. This supports the evidence for a correlation between WM capacity and reasoning abilities. But test scores in reasoning achievement were not significantly increased by WM capacity training in Experiment 2. We therefore conclude that the WM capacity test scores do not predict reasoning achievement scores, which can be explained by the nature of the cognitive training methods and tests in which the participants have not been individually trained in an adaptive way.

5 Combined working-memory capacity and reasoning-strategy training improves reasoning skills in secondary social studies education: evidence from an experimental study[10,11]

Summary

The evidence suggests that working memory (WM) abilities and WM training correlate with deductive reasoning achievements. In this chapter, a combined WM capacity- and WM reasoning strategy training is incorporated in secondary school social studies curricula to investigate its effects on reasoning achievements. Four secondary classes in three schools in the Netherlands participated in the present study with a total of 81 students in higher general secondary education. The WM capacity and reasoning achievements of the experimental groups improved significantly after 4 training weeks compared with the control group results. The gain in reasoning abilities is demonstrated in both experimental subgroups, while the control group results did not improve. The study supports the notion that deductive reasoning gains can be achieved through a combination of WM-capacity- and reasoning-strategy training in a content-based context.

[10] Chapter 5 is based on: Ariës, R.J., Ghysels, J., Groot, W. & Maassen van den Brink, H. (2016). Combined working memory capacity and reasoning strategy training improves reasoning skills in secondary social studies education: evidence from an experimental study. *Thinking Skills and Creativity, 22,* 233-246.

[11] We gratefully acknowledge the contributions of Joaquín Fuster, UCLA Semel Institute for Neuroscience & Human Behavior, and Allyson P. Mackey, MIT McGovern Institute for Brain Research, who supported and gave extensive feedback on the description of the mechanisms and implications of the training methods.

5.1 Introduction

Reasoning in secondary social studies education requires higher thinking skills, which students often find difficult to master and implement in social studies school tests (Newmann, 1987, p.383; Baildon and Sim, 2009). Frequently, these tests contain a large number of reasoning questions (Sluijsmans, 2013). Neuropsychological research has provided evidence that the training of working memory (WM) improves skills in both math and reading comprehension (Conway et al., 2003; Cheshire et al., 2005; Holmes et al., 2009; Karbach et al., 2014; Nevo and Breznitz, 2014). Little research has been conducted to show whether WM-training improves reasoning abilities in humanities education at secondary schools (e.g. history, geography, social studies, and economics). In Chapters 2 and 3, we presented systematic reviews of the literature on school-based WM training, and concluded that optimal WM training consists of both a short-term memory and a long-term memory component. In Chapter 4 we conducted an experiment of combined WM capacity- and reasoning-strategy training, and concluded that the training resulted in an effect on reasoning test outcomes in history courses. Chapter 5 adds to the literature by presenting the results of an evaluation of the two-component WM-training method, including WM capacity and reasoning strategy, implemented over a 6-week training period in secondary school social studies education. The main aim of this intervention is to analyze whether the training duplicates the results of the intervention described in Chapter 4.

Reasoning in the humanities in secondary education is the deductive process of reasoning from statements to reach a conclusion. While inductive reasoning becomes slightly more relevant in reasoning in pre-university education, deductive reasoning is the dominant form of reasoning in the humanities in the lower and medium tracks of Dutch secondary education (SLO, 2011, p.53). Deductive reasoning starts with premises, and yields a logically necessary conclusion that is not explicit in the premises. Deductive reasoning involves three phases: (1) the premise processing phase; (2) the premise integration phase; and (3) the

validation phase (Fangmeier, Knauff, Ruff and Sloutsky, 2006). Two theories have dominated the cognitive literature on deductive reasoning. Goel (2007) states that 'the major issue of contention between the two theories is whether deduction is underwritten by a system of (linguistic) rules sensitive to the logical form of the evidence (mental logic theory), or whether a visuospatial representation of the argument is constructed and evaluated (mental model theory)' (p.435). Based on Neuroimaging studies, Goel concludes that: 'cognitive neuroscience data point away from a unitary system for logical reasoning and towards a fractioned system dynamically reconfigured in response to specific task and environmental cues' (p.435). Reasoning in history education is defined by Leinhardt et al. (1994, p. 134) as 'the process by which central facts and concepts are arranged to build a historical case'. According to Van Drie and Van Boxtel (2008), this requires analysis, synthesis, hypothesis, generalization and interpretation of questions, sources, and retained knowledge. On the basis of a review of the empirical literature, Van Drie and Van Boxtel (2008) concluded that the main reasons for students to underachieve are that they are frequently unable to: (1) take into account alternative views, and use sources extensively; (2) acquire detailed factual knowledge and a broader frame of reference; (3) judge the past by its own standards; (4) take into account the process of continuity; and (5) understand many substantive concepts. Though research on the relevance of reasoning for social studies courses has not been found, we may expect that the analyses of both Leinhardt et al. (1994) and Van Drie and Van Boxtel (2008) on reasoning in history education can inform the results of the social studies courses because of the similarities between history and social studies education (SLO, 2011). Both history and social studies are courses of the humanities. Students have to: (1) develop both general and domain-specific thinking strategies that are used in all of the humanities education; (2) address reasoning questions that meet the criteria of reasoning as defined by Van Drie and Van Boxtel (2008); and (3) address similar subjects. For instance, the main subjects of the social studies course in Dutch secondary education are: 1. Parliamentary

Democracy; 2. the Dutch Constitutional State; 3. Multicultural Societies; and 4. the Welfare State (College voor Examens [Commission for Exam Standards], 2013). The same subjects are addressed in secondary history courses, but in a historical context (College voor Examens [Commission for Exam Standards], 2012). The social studies course addresses the four main subjects in a contemporary context. In line with the generality of its subjects, the social studies course addresses historical, geographical, economic, philosophical, and history of art subjects, all being secondary education courses with which it is closely related.

WM, 'the cognitive system that provides temporary storage of information in the course of complex cognitive activities which appears to play a crucial role both in supporting learning and maintaining focused behavior in practical situations', monitors higher cognitive processes in the brain, and plays a crucial role in developing reasoning skills (Holmes et al., 2009, p.F9). It underlies several cognitive abilities, including logical reasoning and problem-solving (Klingberg et al., 2009). WM is associated with higher cognitive abilities and underlies the frequent inability of children with poor WM to make progress in higher cognitive skills, such as reasoning (Alloway and Gathercole, 2009; Holmes and Gathercole, 2014; Süss et al., 2002). WM needs to change its content constantly and rapidly in order to effectively address a problem state. Therefore, cognitive load seems to play a crucial role in higher cognitive abilities. In this regard, Sweller (1988) states that the capacity of WM could empower reasoning abilities. Also, Sweller (1988) states that novices, in contrast to experts, do not yet possess generic memories-developed schemas or patterns to solve problems, and therefore have to bear a higher cognitive load. So, this process is affected by the two functions of the WM. First, the cognitive function, which stores and manipulates information during reasoning tasks, depends largely on the capacity of information that can be stored in WM (Fuster, 2003; Goldberg, 2010). When more information can be stored and, subsequently, manipulated, reasoning abilities are improved (Olesen et al., 2004). The short-term storage and manipulation of information can be improved by training the cognitive WM function

(Conway et al., 2003; Jaeggi et al., 2008). Second, pattern recognition of reasoning structures is regulated by the meta-cognitive function of WM (Gazzaniga et al., 2009). When frequently exposed to similar reasoning contexts, a student can develop and internalize matched response strategies (Gold, Berman, Randolph, Goldberg and Weinberger, 1996). In this view, WM will form blueprints, and will plan for rational analysis and analytical methods more effectively when trained via reasoning-strategy training. This also causes WM to develop memories of previous patterns that successfully lead to solutions of problems (Goldberg, 2010). Consequently, this can be used to solve new reasoning problems, making reasoning processes more efficient and more effective (Sweller, 1988; Kramarski and Mevarach, 2003; Cheshire et al., 2005). As such, we regard WM as being part of short-term memory and extending to long-term memory as conceptualized by Cowan (e.g. 1998, 2008). Cowan does not view WM as a separate system apart from short- or long-term memory, but representations in WM are a subset of long-term memory representations. Therefore, WM consists of activated long-term memory representations and also of focusing attention, which resides in short-term memory and has a limited capacity.

A school-based training method which contains domain-specific content of both cognitive WM and reasoning strategies integrates several general and domain-specific reasoning processes. Therefore, combined training may better address course-specific reasoning problems and thereby improve the efficiency of the training. In the study in Chapter 4 on secondary history courses, which integrated general and domain-specific reasoning processes, it was shown that significant improvements in reasoning were caused by a combined WM training. This confirmed conclusions from comparable research in which students were trained to develop WM capacity and reasoning strategies in order to benefit their reasoning (Mevarach and Kramarski, 2003; Olesen et al., 2004; Holmes et al., 2009). Furthermore, it showed that reasoning structures were internalized and consequently improved achievements in reasoning tasks and strategies.

In this chapter we investigate whether cognitive WM- and reasoning-strategy training, based on subject-matter knowledge in the social studies course of the secondary education school curriculum, has an effect on students' achievements in tests for which reasoning abilities are required. In three secondary schools in the Netherlands, we conducted an experiment that is based on WM-training tasks. The learning strategy includes domain-specific cognitive WM and reasoning strategy.

The aim of the training intervention is to investigate: (1) the effect of the complete training on achievement in reasoning tests; (2) its durability after completion of the training; and (3) the effects of age, gender, pretest reasoning achievement scores and WM pretest scores on post-test achievements by conducting t-tests and using regression techniques. The intervention is implemented in the existing social studies course curriculum. The subject-matter approach of the intervention is based upon a specific training method that, according to earlier research, is expected to result in an improvement of reasoning achievements (Van der Sluis et al., 2007; Lee et al., 2009; and see also Chapters 2 and 3 of this thesis).

We find that the experimental groups performed better on a battery of cognitive tasks (n-back and Odd One Out) and reasoning strategy (IMPROVE) tasks after pretesting on reasoning abilities (Davidson et al., 2006; Jolles et al., 2006). The IMPROVE tasks were modifications of the original tasks by Alloway and Gathercole (2009) and Mevarach and Kramarski (2003). After the training the students were post tested to evaluate improvements in their cognitive WM and reasoning abilities. The data of the experimental groups were then compared with the data of the control groups that were trained by using 'low-dose' working memory tasks, such as evaluating reasoning questions using whole-class teaching methods.

5.2 Methods

5.2.1 Participants

81 students (51 women, 30 men, *Mage* = 15.8 years, age range: 15-18 years), from four heterogeneous social studies classes in the higher secondary general education medium track (HAVO) from three schools in the Netherlands, participated in the study. The four classes were the only classes in the participating schools that were being taught social studies in Grade 4. The students were not randomly assigned to the groups by the researcher at the beginning of the study. They were allocated to the classes by the school at the beginning of the academic year exclusively based on the students' schedules, and not based on their personal characteristics or IQ-measures. Because social studies is an extracurricular course, all students followed the course by their own choice. The control group (28 women, 14 men, *Mage* = 15.83 years, age range: 15-18 years) and the experimental group (23 women, 16 men, *Mage* = 15.77 years, age range: 15-17 years) both contained 1 or 2 pupils who were dyslectic, which can influence test outcomes owing to their limitations in verbal impression and/or expression. The mean age and the number of participants that repeated the class or came from a lower or higher track was approximately the same for both groups, as was the male/female ratio. All students lived in or near the urban areas of Maastricht, Heerlen and Geleen, which are three neighboring urban areas in the most southern part of the Netherlands that are approximately 18 kilometers from each other.

The control group consisted of two classes (one Maastricht and one Geleen class) (Table 5.1). Two other classes (one Heerlen and one Geleen class) formed the experimental group. Three experienced and certified teachers (all male) had been appointed to the classes prior to the intervention by the schools. In the Heerlen group, both an experienced teacher and an intern provided the intervention. All teachers used a conventional instruction method in their courses, which contained an introduction to new subjects and the use of whole class dialog and discussion to answer questions. The classes were randomly assigned to

the control or to the experimental group. Only the Geleen teacher taught in both the control and the experimental groups. Therefore, part of the effect in both the Heerlen and Maastricht groups may be attributed to 'teacher effect', 'school effect' or other context effects.

Table 5.1 Research design

	Group name	Teachers' experience
Control group	Maastricht	Experienced
	Geleen	Experienced
Experimental group	Geleen	Experienced
	Heerlen	Experienced and intern

5.2.2 Intervention

In this chapter, we have duplicated the combined training method from Chapter 4 and study whether the training method that had been tested in the history courses could also be implemented in secondary education social studies courses. The following description of the tasks is derived from the study in Chapter 4 on improving reasoning skills in secondary history education by WM training.

The experimental groups were trained in two cognitive tasks, the n-back and the Odd One Out. In the n-back task, the participants saw a series of 20 domain-specific nouns, successively presented in a PowerPoint presentation at the rate of 3 seconds per stimulus. The task was to remember the sequence in which the nouns were presented. Most of the nouns that appeared were presented in black, but when a noun colored green appeared on the screen, the participants had to remember the previous noun and write it on an answer form within 5 seconds. The tasks were presented in 1-back, 2-back, and dual 2-back batteries (Appendix A), which were derived from the standardized Automated Working Memory Assessment test battery (Alloway and Gathercole, 2008). While Jaeggi et al. (2010) consider that the n-back is not a useful measure of individual differences of WM due to its insufficient reliability, other studies have provided evidence of n-back to lead to positive effects

on fluid intelligence (Au et al., 2015) while fluid intelligence could mediate the link between n-back training and deductive reasoning (Beatty and Vartanian, 2015). Also, n-back training has been shown to produce broad transfer effects from training-related changes to untrained tasks and functions (Von Bastian and Oberauer, 2013). According to Salminen, Mártensson, Schubert and Kuhn (2016), this is caused by n-back training to increase 'white matter integrity [in the brain] in pathways that connect brain regions related to WM processes (p.245).

The Odd One Out task presented four domain-specific nouns or images on a video screen in a PowerPoint presentation lasting 12 seconds. In the sequence, one noun or image was the 'odd one out', which participants had to write down on an answer form when all nouns or images were shown (Appendix B).

The content of these training tasks, which were content-based social studies subjects, was the same as in the intermediate tests and the posttests. The pretests contained non-trained historical content for better baseline measurements. Furthermore, the dimensions of the cognitive training did not differ from the tests which contained 1-back, 2-back, dual 2-back, three verbal and three visual Odd One Out tasks during a section (30 minutes) of one lesson (50 minutes) per week. The content of the tasks was equivalent to the subjects concerned (Mass Media and Criminal Law) and to the conventional instruction method that was used in the control group lessons.

We used the standardized IMPROVE-method (Appendix D) for reasoning-strategy training (Mevarech and Kramarski, 2003), which is used as a format to develop reasoning strategies. The training was performed in heterogeneous groups that contained three or four participants: one low-, one or two middle-, and one high-achieving student. Four open-ended reasoning questions were presented in turn, with 10 minutes to answer each question. The question had to be addressed by one of the groups' participants by using the IMPROVE-questionnaire in order to analyze the reasoning process. Directions by the other members of the group were given when the participant could not

answer one of the questions of the questionnaire. After 10 minutes, a new participant addressed a new problem. The teacher participated in each group's reasoning process for 2 minutes per problem in the same way that the other members of the group participated. Teacher participation in collaborative learning is considered very important in developing students' critical thinking skills in liberal studies (Fung and Howe, 2012, 2014). The tasks were trained in one 50-minute social studies lesson per week over a 6-week period.

5.2.3 Data Collection

The combined effects of cognitive WM and reasoning strategy training were measured by a pretest, an intermediate test and two posttests in the intervention and control groups. The pretests measured the homogeneity of the experimental and control groups. One intermediate test and two posttests measured short- and long-term effects at the end of the training period and 8 weeks after the first posttest to test for effect durability.

Tests to measure cognitive WM function levels were conducted on two tasks, the n-back and Odd One Out tasks (Jaeggi et al., 2008, Holmes et al., 2009). The tasks were specifically developed to engage the processes required for the management of different tasks, engage the executive processes required for each task, and discourage the development of task-specific strategies and the engagement of automatic processes. Furthermore, both tasks test the storage capacity and manipulation of information in WM. During the intervention period, the experimental groups were trained with altered, domain-specific, versions of the task standards, and nouns and verbs were used to train both the verbal and the visual components of cognitive WM. The tests were conducted on the same tasks, but the content differed from the content that was used in the training tasks: historical nouns and images were used to prevent students of both groups becoming accustomed to the nouns and images that were shown in the intervention period, which could positively affect the test results. The participants could score a

maximum of 58 points (1 point per good answer) on the cognitive test. In this test, the participants were not tested for substantive knowledge, but for WM capacity.

To test for the acceleration of reasoning skills, all groups were tested by analyzing the answers on open-ended reasoning questions in official school tests. The tests were compiled by the three social studies teachers, in consultation with their school social studies colleagues, and contained about 10 knowledge questions and about 10 reasoning questions. In each school, the tests contained questions that were compiled and analyzed by the same teacher based on contemporary developments, as well as questions that were extracted from the social studies teachers' guidebook (note that all classes used the same book). As an additional source of homogeneity, it should be noted that the school tests refer to the common Dutch graduation exam standards. While each school developed the tests independently from the other schools, the tests could provide for differential effects when the groups are compared with each other. These differential effects could be caused by: (1) the subjects that differed between groups; (2) the composition of the test in which both knowledge and reasoning questions were included; and (3) the level of difficulty of sources on which a large number of the reasoning questions were based. The participant's reasoning achievements were assessed by the teachers' analyses of the correct responses to the reasoning questions compared with the examination model, with a maximum of 30 points per test. All answers were analyzed based on three criteria for reasoning skills: contextualization - the ability to put the specific reasoning question in a (broader) context; use of sources - the ability to analyze an attached source which provides additional information; argumentation - the ability to provide arguments (see Appendix C) (see Van Drie and Van Boxtel, 2008). Teacher training in advance of the experiment also included the awareness of these categories, which is likely to have contributed to more homogeneous examination assessment models between tests.

5.2.4 Design of the experiment

This experimental study compares the effects of the WM-based intervention with conventional instruction methods. The experiment employed pretests, intermediate tests, and posttests to determine training improvements. Research was conducted in four 4th Grade classes of Higher General Secondary Education in the Netherlands. The participants were nested in multiple classes in three different schools. Students of two classes were assigned to the experimental condition and those of the other two classes to the control condition.

The two classes of the experimental group were trained by using the cognitive and reasoning-strategy tasks to improve reasoning abilities. The two classes of the control group were trained by methods of conventional instruction. In the latter, the teacher introduced new subjects and the participants answered (reasoning) questions from the textbook that were verified by interaction between the teacher and all of the students.

To measure the effects of the intervention on cognitive WM function levels and reasoning achievements, intermediate-test and posttest scores from the experimental group and from the control group were analyzed independently by using paired t-tests. Reasoning achievement test scores of the experimental and the control groups were also compared by using independent t-tests.

5.2.5 Procedure

The various classes in the experiment studied different subjects from the same social studies textbook. The Geleen school, which included an experimental and a control group, studied Criminal Law, while the two other schools studied Mass Media. All the classes spent three 50-minute lessons a week on social studies prior to and during the training. The structure of the lessons prior to the intervention was the same for all groups. All teachers used the aforementioned conventional instruction method as the dominant method of teaching. 2 weeks prior to the first pretests, the intervention groups' teachers were trained in 60 minutes on

the mechanisms of the training by providing them with the necessary information to conduct all training tasks. The researcher monitored the teachers' skills by evaluating the teachers' skills every 2 weeks by telephone. All groups were pretested on WM capacity 8 weeks after the start of the academic year. Prior to taking the test, the teacher explained the reason, contents, and structure of the test. The topics included historical nouns and general images that were not related to the topic concerned, in order to get a baseline measurement. The pretest to measure reasoning strategies and achievements took place in the next lesson by means of a regular school test. Like all other official school tests, the latter test was compiled by the teachers and based on the content of the textbook. The school tests were evaluated by the teachers. After the evaluation, heterogeneous cooperative groups in the experimental groups were trained in reasoning strategies. All groups consisted of one low-, one or two middle-, and one high-achiever. The students were allocated to the groups by the researcher on the basis of their test results.

A 6-week intervention period followed after pretesting, in which both groups studied the subjects in the same time span in three lessons per week (total: 150 minutes). The control groups studied the subjects by using the conventional instruction method. The experimental groups studied these subjects in 1-week lesson cycles by using the WM-based method: an introduction to the subject was given by the teacher. After the cognitive WM training was introduced during the first 5 minutes, training took place in the second lesson for 25 minutes (Jaeggi et al., 2008). In the remaining 20 minutes the textbook reasoning questions that were discussed in the first lesson were evaluated. In the third lesson, training of reasoning skills took place by using the IMPROVE-method (Mevarech and Kramarski, 2003). After a short introduction by the teacher, reasoning skills were trained in the next 40 minutes (with 10 minutes per reasoning question). Process evaluation took place during the last 5 minutes in which the questions from the students concerning the reasoning process were discussed by the teacher and students by

means of Socratic debate. Only the results of those participants who were fully assessed and missed less than two interventions (96%) were included in the analyses.

The first posttests were conducted 1 week after the training ended. The second posttest was conducted 8 weeks after the first posttest.

5.3 Results

The Maastricht school did not have an experimental group. The Heerlen school did not have a control group. The Geleen school contained both an experimental and a control group. All the groups used the same cognition pretest with the same historical content, which differed from the social studies subjects, and was compiled specifically to exclude foreknowledge as a distinctive predictor for achievements.

Table 5.2 Independent t-test results for pretest scores: for the complete sample and per group

		Cognition pretest		Reasoning achievements pretest	
	N	Mean	SE	Mean	SE
Control group total	42	39.12	.83	15.71	.82
Con. Maastricht	*24*	*39.67*	*1.17*	*15.25*	*.85*
Con. Geleen	*18*	*38.39*	*1.18*	*16.33*	*1.56*
Experimental group total	39	38.94	.93	14.97	.83
Exp. Geleen	*21*	*38.05*	*1.35*	*15.00*	*1.21*
Exp. Heerlen	*18*	*40.20*	*1.16*	*14.94*	*1.14*

Notes. No significant differences between groups were observed. Mean diff = mean difference; Con = control group; Exp = experimental group.

In the following text, we first analyze the comparability between groups by using t-tests on the pretests (Table 5.2). Then, we use

Independent sample t-tests[12] on all intermediate tests and posttests to demonstrate group differences. Because the small sample size can lead to a type II error, additional analysis was conducted to determine whether student characteristics could explain the effects of the intervention on posttest reasoning achievement scores. Also, according to Murnane and Willet (2010, p.103), multivariate regression increases the precision of the estimate of the potential effect of the intervention. We lastly evaluate these predictor variables by means of a multiple regression analysis in which the pretest scores are included as control variables.

T-tests on the pretests show that there is no significant difference observed between groups with regard to mean scores. Therefore, it can not be ruled out statistically that the groups belong to the same population.

There was no significant difference between the average pretest scores of the control and the experimental groups in reasoning achievements. While both group's test scores showed equal variances, we analyzed the intermediate and (both of) the posttest scores based on the equality between both groups.

Table 5.3A Independent t-test results and descriptive statistics for cognition intermediate tests and posttests between groups

	Group	Mean score	SD	df	t
Intermediate test	Exp	41.68	5.330	74	-4.11*
	Con	36.86	4.887		
	Exp Geleen	*41.67*	*6.118*	*61*	*-3.38**
	Exp Heerlen	*41.69*	*3.966*	*53*	*-3.245**
Posttest 1	Exp	42.06	3.929	73	-4.39*
	Con	38.17	3.728		
	Exp Geleen	*43.00*	*3.578*	*61*	*-4.92**
	Exp Heerlen	*40.42*	*4.122*	*52*	*-1.802*

[12] The Kolmogorov-Smirnov test showed that the distribution in both the experimental, $D(39) = 0.14$, $p > .05$, and the control groups, $D(42) = 0.12$, $p > .05$, suggested no deviation from normality in the regular school pretest scores. Also, Levene's test for equality of variances showed no significant values $F(1, 79) = 0.006$, $p = 0.94$.

Notes. t = t-ratio for each 'Group'; * = significance of values for each 'Group' <.05; Exp = experimental group; Con = control group.

Table 5.3B Independent t-test results and descriptive statistics for reasoning achievements in intermediate tests and posttests between groups

	Group	Mean score	SD	df	t
Intermediate test	Exp	17.69	4.319	79	-2.21*
	Con	15.69	3.805		
	Exp Geleen	*18.14*	*3.336*	*61*	*-2.28**
	Exp Heerlen	*17.17*	*4.328*	*58*	*-1.21*
Posttest 1	Exp	19.77	4.556	79	-3.44*
	Con	16.19	4.779		
	Exp Geleen	*21.52*	*4.226*	*61*	*-4.33**
	Exp Heerlen	*17.72*	*4.142*	*58*	*-1.18*
Posttest 2	Exp	20.31	4.079	79	-3.80*
	Con	16.55	4.774		
	Exp Geleen	*22.14*	3.610	61	-4.730*
	Exp Heerlen	18.17	3.585	58	-1.289

Notes. t = t-ratio for each 'Group'; * = significance of values for each 'Group' <.05; Exp = experimental group; Con = control group.

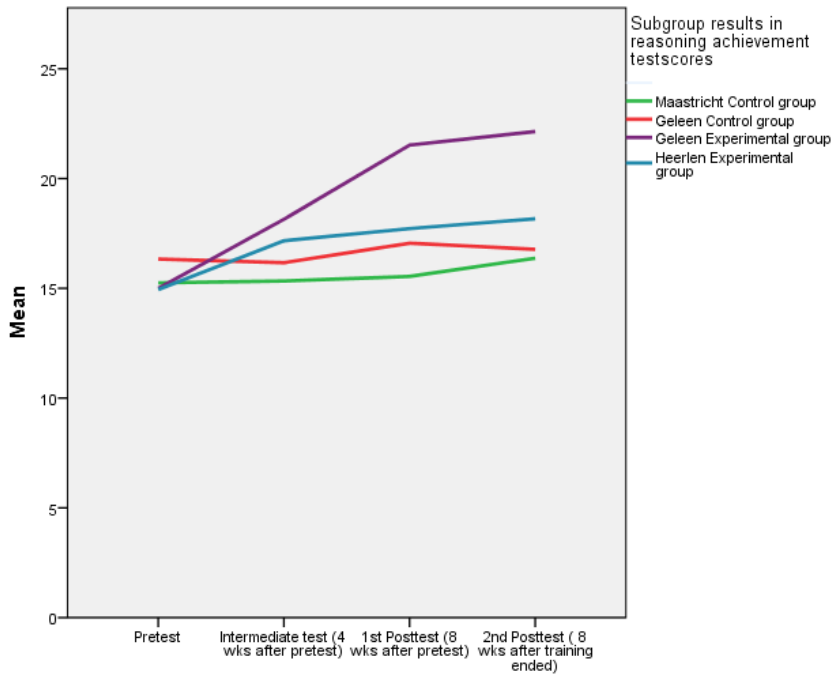
Assuming on the basis of the above-mentioned analyses of the pretests, that the control and experimental groups were 'equal in expectation' regarding their learning process at the start of the experimental period, we investigate the effect of the experiment, first, by a simple comparison of the outcomes of the intermediate and posttests. Table 5.3 shows the results of the relevant t-tests for the cognitive outcomes (panel A) and for reasoning (panel B). The analyses demonstrate that group differences are statistically significant for all intermediate tests and posttests between the control and the experimental groups.

For more detailed information, we then compared each experimental *sub*group (Geleen 1, Heerlen) with the control group's *general* scores in order to investigate each experimental group´s achievements[13]. With regard to the intermediate tests and posttests that measure cognitive levels (Table 5.3A), the Geleen experimental group's

[13] Specific analyses of the control subgroups' compared with the experimental groups' test scores are not pertinent to the present discussion, since the control groups' general test scores showed no significant variations in the intermediate tests and posttests.

test scores show significant effects compared with those of the control group. The effect of the Geleen group remains significant when compared with the Geleen control group in both cognitive intermediate scores (not shown in the table): $t(37) = -2.37$, $p < .05$, $r = .36$ and posttest $t(37) = -4.51$, $p < .001$, $r = .60$. While the Heerlen experimental group´s test scores also demonstrate statistically significant effects in the intermediate test, its posttest scores do not significantly differ compared with the control group.

With regard to short-term improvements in reasoning achievements (Table 5.3B, Figure 5.1), achievements in the intermediate tests and the first posttests demonstrate statistically significant effects for the Geleen experimental subgroup, but not for the Heerlen experimental group compared with the general control groups´ results.



Note: Wks. = weeks

Figure 5.1 Mean scores in achievements tests by subgroup

The effect of the Geleen experimental group does not remain significant when compared with the Geleen control group in the intermediate achievement test ($t(37) = -1.77$, $p = .09$, $r = .28$). Significant

effects are found in the first posttest ($t(37) = -3.02$, $p < .01$, $r = .44$) when the Geleen experimental and control groups are compared.

The non-significant improvements of the Heerlen experimental group might be attributed to the relatively low pretest and intermediate test scores, as well as to the first posttest scores of the Heerlen group compared with the general control group's scores. Both categories of test results could result in non-significant enhancement of reasoning achievements in the Heerlen group compared with the control group, but, on the other hand, could result in a significant improvement when compared with its own pretest results. Note that the Heerlen intermediate test $t(17) = -1.54$, $p = .14$, $r = 0.35$ and first posttest $t(17) = -2.00$. $p = 0.06$, $r = 0.44$ results have also not statistically significantly increased when compared with this groups' pretest results in a paired samples t-test. During the experiment, the Heerlen group was not taught by the experienced teacher but by an intern teacher who trained the students during the experiment. The pretest results could therefore be attributed to the experienced teacher, while the intermediate and the posttest results could be attributed to the intern. In this regard, experienced teacher vs. intern effects could affect the outcomes of the Heerlen group's test scores because of the teacher's role of managing and directing the students' reasoning processes (e.g. Kane, Rockoff & Staiger, 2008).

Additional analysis was conducted to determine whether student characteristics could (in part) explain the effects of the intervention on posttest reasoning achievement scores. In this regard, age and gender differences and the scores of both pretests are regarded as commonly used predictor variables. For instance, WM limitations are characterized by large differences of 'marked age-related trends' (Sander, Lindenberger and Werkle-Bergner, 2012, p.2007; Engle, 2002; Gathercole et al., 2004). Adolescence represents the final phase of brain growth. It therefore represents a 'major transition that takes place over most of the second decade of human life' (Paus, 2005, p.60), in which WM matures into a system that accommodates complex cognitive activities, such as

reasoning (Holmes et al, 2009). Therefore, maturation is argued to be a predictor variable for improved reasoning skills. But we expect no significant differences between age groups because of the minor age differences and the fact that older students in general repeated a class or came from a lower level.

Two different mechanisms have been postulated to explain differences in training improvements, the magnification account (i.e. the Matthew effect) and the compensation account. In this respect, Titz and Karbach (2014) state that the magnification account 'assumes that individuals who are already performing very well will benefit most from cognitive interventions' (p.861), owing to their more efficient cognitive resources that enable these students to acquire and implement new strategies and abilities. The compensation account 'assumes that high-performing individuals will benefit less from cognitive interventions, because they are already functioning at the optimal level, which leaves less room for improvement' (p.861). This can be caused by high pretest results, due to which there is less scope for high performers to improve significantly better in intermediate test or posttest results. But it can also be caused by the individual's higher cognitive levels which leave less room for improvement. The former correlation can be tested by analyses of the reasoning achievement test results of school tests. The latter correlation can be analyzed by looking at WM levels. We therefore use two specifications of test results (WM-capacity and reasoning pretests) to investigate the correlation between pretest and reasoning posttest results.

Another matter concerns the cognitive and emotional developments which have been known to differ between boys and girls (Nagy, Jacklin and Martin, 1999). For example, Kimura (1996), states that:

'A solid body of research, carried out primarily in North America and Western Europe, has established that men, on average, excel on spatial tasks (particularly those tapping ability to imaginably rotate a figure), perception of the vertical and horizontal, mathematical reasoning, and spatio-motor targeting ability. Women, on average, excel on tasks of verbal fluency (where words must be generated with constraints on the letters they contain),

perceptual speed (in which rapid pattern-identity matches are made), verbal and item memory, and some fine motor skills' (p. 259).

We expect to see gender differences in improvements in WM capacity and reasoning because the WM training tasks only contain elements of verbal fluency, perceptual speed, and verbal memory. Consequently, both girls and boys will improve more on reasoning posttests because of two opposing mechanisms. In this regard, male students could improve more because the specific characteristics of their WM are underdeveloped compared with those of female students' WM, which leaves more room for male improvement. On the other hand, female students are expected to improve more because the training taps into cognitive resources that are already more developed (e.g. verbal fluency, perceptual speed, and verbal and item memory, of which the training consists).

We evaluated these predictor variables through a multiple regression analysis in which five models were tested (tables 5.4A and 5.4B). In Model 2 the pretest scores of both pretests were included as dummy variables, and in Model 3 as continuous variables in order to analyze whether the correlations between pretest scores and posttest scores were linear. Magnification and compensation account theories require non-linear correlations. The dummy variables in Model 2 are categorized in the same manner as the group formation during IMPROVE-training (low, middle, high performers; boundaries 33 1/3%, 66 2/3%). Models 4 and 5 include interrelated predictor variables intervention*gender and intervention*reasoning pretest scores, respectively. Model 6 is included as an addition to Model 2, and shows test interactions between the intervention condition and the reasoning pretests of low, middle, and high performers to check for possible compensation or magnification accounts. Table 5.4A shows comparisons of all intervention and control groups. In Table 5.4B only the Geleen control and experimental groups are compared. Table 5.4A shows that

the intervention is a significant predictor of posttest improvements (Models 2 and 3). Furthermore, $R^2 = 0.25$ (and 0.23) indicates that roughly 25% of the variance is accounted for by the variance of all explanatory variables combined. The results also show that there is a statistically significant correlation between the reasoning achievements pretest score ($p < .01$) and the prediction of posttest improvements. Model 2 shows that having a low score in the achievement pretest is a significant negative predictor (b = -3.42, $p < .05$), while students with an initially high test value do not seem to differ significantly from the reference group, that is, students with medium pretest scores on reasoning achievement. Model 3 shows the (continuous) pretest reasoning score as a significant predictor for posttest improvements (b = 0.27, $p < .01$). When the pretest reasoning score is analyzed jointly with the intervention group, its contribution does not prove to be related to the interaction in a statistically significant way (Model 5). Also, Model 6 does not provide evidence for significant contributions of either low or high pretest scores in the intervention groups to the posttest scores. Hence, we see no clear evidence of magnification or compensation effects.

WM capacity at pretest, age, and gender do not predict posttest scores significantly, which partly corresponds to our expectations. The latter is also non-significant when analyzed in interaction with the intervention group (Model 4). We will return to this issue in the discussion section.

Table 5.4B also shows that the intervention is a significant predictor for post-test scores in the Geleen group ($R^2 = 0.20$). As continuous variables, the reasoning achievement pretest scores make a significant contribution in predicting posttest reasoning achievement improvements (Model 3). The pretest reasoning coefficient indicates that for every unit increase of pretest score the posttest score is predicted to increase by .32 units. Specifically, the coefficient of determination between group and reasoning pretest and posttest scores ($R^2 = 0.38$) indicates that 38% of the variance is explained by these variables. However, their contribution does not prove to be related to the interaction when analyzed jointly with the

intervention group (Model 5). WM capacity, age and gender differences do not predict posttest improvements significantly, even when jointly analyzed with the intervention group (Model 4).

The second reasoning achievements posttest measured the durability of effects after the completion of training. The second posttest was conducted 8 weeks after the training ended which resulted in significant effect sizes of the experimental group compared with the control group even when the control subgroup results improved slightly compared with the prior tests. While we can attribute the latter to natural learning processes, the durable significant results of the experimental group justify the analyses of the experimental *sub*groups' test results (Geleen 1 and Heerlen) in order to retrieve more information. We therefore compared the posttest results of each experimental subgroup with its own pretest scores as well as with the control group's general second post-test scores, and conclude that both experimental subgroups showed significant improvements in their second posttest scores.

The Geleen experimental group posttest results reveal a significant improvement when compared with the general control group scores, as well as when compared with the Geleen control group results $t(37) = -4.44$, $p < .001$, $r = .59$, and when compared with its own pretest results $t(20) = -5.92$, $p < .001$, $r = 0.80$. Notably, in line with the results of the intermediate tests and the posttests, the Heerlen experimental group did not achieve significantly better than the control group results. The group did improve significantly when its second post-test results are compared with its pretest results $t(17) = -2.48$, $p < .05$, $r = .52$. We therefore conclude that test results remain durable due to training. We can also conclude that both the Geleen and the Heerlen experimental group's second posttest scores improved (but not significantly) when compared with their first post-test results. While improvements in this regard could be attributed to maturation and natural learning processes, several studies, including Chapter 4 in this dissertation, suggest that these

Table 5.4A Summary of regression effects of independent variables on posttest reasoning achievements

	Model 1 B (*SE B*)	Model 2 B (*SE B*)	Model 3 B (*SE B*)	Model 4 B *(SE B)*	Model 5 B *(SE B)*	Model 6 B *(SE B)*
Constant	16.19 (.73)	17.42 (1.56)	11.23 (4.36)	11.02 (4.48)	11.11 (4.66)	18.08 (1.63)
Intervention group (0/1)	3.84 (1.07) ***	3.76 (1.07)***	4.03 (1.05)****	4.36 (1.72)*	4.29 (3.27)	2.32 (1.61)
Age 16-18 yrs (0/1)		0.15 (1.12)	-0.19 (1.10)	-0.21 (1.11)	-0.18 (1.11)	0.23 (1.14)
Gender Girl (0/1)		0.78 (1.12)	0.21 (1.10)	0.46 (1.52)	0.21 (1.11)	0.63 (1.13)
Intervention group*Gender Girl (0/1)				-0.52 (2.18)		
Pretest reasoning achievement low score (0/1)		-3.42 (1.36)*				-3.77 (1.89)
Pretest reasoning achievement high score(0/1)		-0.15 (1.39)				-1.48 (1.65)
Pretest reasoning achievement low score*intervention group (0/1)						1.00 (2.62)
Pretest reasoning achievement high score*intervention group (0/1)						3.28 (2.20)
Pretest WM capacity low score (0/1)		-1.86 (1.39)				-2.34 (1.42)
Pretest WM capacity high score (0/1)		-1.17 (1.27)				-1.03 (1.27)
Pretest WM capacity			0.02 (0.10)	0.02 (0.10)	0.02 (0.10)	
Pretest reasoning achievement			0.27 (0.10)**	0.27 (0.10)**	0.28 (0.14)*	
Intervention group*Pretest reasoning achievement					-0.16 (0.20)	
R^2	.15	.25	.23	.23	.23	.28

Notes. *p < .05 **p <.01 ***p =.001 ****p <.001; standard errors between brackets; N=78; Yrs = years.

Table 5.4B Summary of regression effects of independent variables for the Geleen groups

	Model 1 B (*SE B*)	Model 2 B (*SE B*)	Model 3 B (*SE B*)	Model 4 B (*SE B*)	Model 5 B (*SE B*)	Model 6 B *(SE B)*
Constant	17.06 (1.09)	18.16 (2.41)	8.67 (5.44)	6.04 (5.55)	8.54 (5.86)	18.22 (2.58)
Intervention group (0/1)	4.47 (1.48)**	5.10 (1.41)***	4.95 (1.40)***	8.36 (2.49)**	5.20 (4.03)	5.66 (1.89)**
Age 16-18 yrs (0/1)		-1.23 (1.66)	-1.50 (4.55)	-1.75 (1.52)	-1.50 (1.57)	-1.29 (1.69)
Gender Girl (0/1)		0.77 (1.72)	0.23 (1.56)	2.90 (2.23)	0.21 (1.61)	0.49 (1.78)
Intervention group*Gender Girl (0/1)				-4.94 (3.02)		
Pretest reasoning achievement low score (0/1)		-3.12 (1.93)				-1.70 (2.43)
Pretest reasoning achievement high score (0/1)		2.30 (1.87)				2.37 (2.61)
Pretest reasoning achievement low score*intervention group (0/1)						-2.96 (3.00)
Pretest reasoning achievement high score*intervention group (0/1)						-0.25 (3.52)
Pretest WM capacity low score (0/1)		-1.67 (1.80)				-1.81 (1.90)
Pretest WM capacity high score (0/1)		-1.60 (1.75)				-2.02 (1.94)
Pretest WM capacity			0.44 (0.29)	0.14 (0.13)	0.10 (0.13)	
Pretest reasoning achievement			0.32 (0.12)**	0.29 (0.12)*	0.33 (0.16)*	
Intervention group*Pretest reasoning achievement					-0.02 (0.24)	
R^2	.20	.41	.38	.42	.38	.43

Notes. *p < .05 **p <.01 ***p =.001 ****p <.001; standard errors between brackets; N=39; Yrs = years.

improvements can also be caused by the training and regarded as a long-term effect (Chapter 4; Dahlin, 2010; Holmes et al., 2009). In fact, Holmes and Gathercole (2014) state that there's 'preliminary evidence of accelerated learning following training, with significant improvements in math […] and improvements in reading comprehension reported post-training […].' (p. 441), which suggests that WM continues to take advantage from the training, even a long time after the training has ended.

5.4 Discussion and conclusion

Reasoning is an integral part of secondary school courses. This experiment indicates that the combined domain-relevant training of WM capacity and reasoning skills improves school curriculum-based test scores owing to enhanced WM skills and internalized reasoning strategies.

While the experimental group's training leads to (significantly) improved test results, the control group shows no significant improvement in both intermediate and posttest scores in WM-capacity and reasoning achievements. The experimental group's improvements are in agreement with the conclusions of the study in Chapter 4 of this dissertation, in which similar groups underwent the combined training in the history curriculum, and also in earlier research in which either WM capacity or reasoning strategies are trained (e.g. Kramarski and Mevarach, 2003; Olesen et al., 2009).

With regard to the concept of 'metacognition', we note that the term is used by Mevarach and Kramarski (2003) and Kramarski and Mevarach (2003), and is based on Polya's (1945) four-stage problem-solving model. The term has also been used in the present Chapter because of the use of the Mevarech and Kramarski's IMPROVE-training battery. 'Metacognition' refers to one's knowledge of one's own cognitive processes. In the case of the IMPROVE-training battery, the concept of metacognition could also be defined as 'the deduction of general strategies of thinking and action'. According to Fuster's (2003, p. 106-110)

perception-action cycle, these general strategies of thinking and action are reducible to abstractive executive cognitive units (cognits), which are transferable in solving other concrete problems. WM-training helps the student to construct them. This would mean that strategies are trained at a level of higher cognition, not metacognition.

The results of the regression model indicate that, for every unit increase of pretest achievement scores, the score of the reasoning achievements posttest can be predicted to increase by .27 units for all groups combined. Students with a low pretest reasoning achievement score do not typically improve more on their posttest compared with students who have a high pretest score. This sheds light on the question of who benefits the most from cognitive interventions. This question is particularly important regarding students who underachieve in reasoning questions. But an analysis of the intervention groups' pretest reasoning score (Table 5.4A, Model 5) show that pretest score differences do not predict posttest score gains for anyone. Therefore, test-score gains for pretest performers in the present intervention groups could also have been caused by behavioral characteristics, such as improved motivation or concentration. With regard to the latter, the training method requires that students focus their attention on an internal representation. Attention has two components: (1) an inclusionary component consisting of the content in focus; and (2) an exclusionary component consisting of the inhibitory control of interference. Since WM is demonstrably enhanced by concentration and weakened by distractibility, underachievement could easily be caused by distractions. Although motivation and concentration have not been analyzed, minimizing distractions may be a requirement for an effective intervention. Furthermore, enhancing and training attention to improve deductive reasoning could also be regarded as an important outcome measure of the training, while it is regarded as a general cognitive function within which WM is subsumed. The fact that the experimental groups are more likely to be motivated and attentive than the control groups could

therefore explain the better test scores within the experimental groups (J. Fuster, personal communication, August 11, 2015).

The durability of training effects is indicated by experimental and control group test-score differences which remain significant in the second posttest, 8 weeks after the end of the intervention period. These results suggest improvements in WM capacity and the internalization of reasoning structures which, consequently, have durable effects on reasoning test achievements. However, the durability of, specifically, WM capacity improvements could not be explained by the WM tests in the current study for it did not contain a second cognitive posttest to measure for durable effects. Note that the training and testing of WM capacity was not performed in an adaptive form, which is considered to be crucial to continually improve WM capacity and, as far as specifically n-back training is concerned, has proven to lead to (small) positive effects on fluid intelligence (Olesen et al., 2009; Au et al., 2014). Also, WM-capacity training and testing did not contain standardized training batteries (e.g. Automated Working Memory Battery) (Alloway and Gathercole, 2008), but were adapted to contain content-relevant subjects. Consequently, we argue that adaptive, non-standardized, types of the training and test batteries used in the current study are less valid to measure WM-capacity and transfer effects.

Retained knowledge of the training could explain higher WM-capacity test scores compared with control group results, as similar tests have been used during training. This could provide some insight on the extent of possible far-transfer effects on content-based testing (whether this concerns transfer of knowledge or skills or a combination of both) of WM capacity, as well as on the type of tests that could be used in classroom-based interventions. Can far-transfer effects occur to other courses when content-based WM capacity training on a course appears to lead to significant effects in the reasoning tests? Does either standardized or content-based WM-capacity training cause significant improvements in reasoning tests? Should WM capacity be trained in a standardized fashion to further understand WM-training, or should WM-

training in the classroom be adapted to classroom dynamics in order to effectively improve the students' test results? Is WM-capacity training an efficient and effective way to improve reasoning test results? In this regard, training of reasoning skills could in itself provide significant effects on reasoning achievements, since in earlier research (e.g. in Chapter 4 of this dissertation) gains in reasoning achievements were caused by stand-alone reasoning strategy training, which also proved to be durable and therefore internalized (e.g. Goldberg, 2010; Chapter 4). Also, questions can be raised concerning whether WM is enhanced by WM-training alone, or is reasoning training also needed? In line with our model of WM, we suggest that reasoning training must be included in WM-training in order to store reasoning strategy concepts in the novices' long-term memory. This suggestion is also supported by the lack of evidence of far-transfer effects of WM-capacity training alone. Another important question raised is whether reasoning is enhanced by reasoning training alone, or by combined capacity and strategy training. The design of the study can not provide insight into this matter. In Chapter 4 of this dissertation we pointed out that reasoning is enhanced when only reasoning strategy is trained, or when combined training took place compared with only WM-capacity training. However, the question whether reasoning will be enhanced even more when combined training takes place, compared with only reasoning strategy training, remains a topic for further review.

Note that, while the Heerlen experimental group´s WM-capacity test scores demonstrate statistically significant effects in the intermediate test, the posttest scores do not significantly differ compared with the control group. This may be attributed to the fact that the Heerlen group's cognitive posttest did not take place immediately, but, due to planning difficulties, 4 weeks after cognitive training ended. The students´ cognitive WM levels are likely to have decreased during the 4-week time span in which they did not train WM capacity and spent much time on extracurricular activities and less on direct learning activities. Evidently, the sample sizes in all the participating schools are small, and therefore

caution is advocated regarding external validity. Future larger-scale research is required to corroborate the findings presented here.

Teachers often find that students have a lack of motivation to acquire reasoning skills. Despite the fact that teachers acknowledge the importance of teaching reasoning skills to students, they often find it difficult to find approaches to teach these skills effectively to them. Several characteristics of social studies make them particularly dependent on reasoning skills. During social studies courses, students often need to learn facts, or answer reasoning questions by checking the course texts independently. Furthermore, students often get no individual feedback on how they formulated their answer when being trained by using conventional methods. The performance of students in reasoning tests depends strongly on combining content-based knowledge, reasoning skills, and the formulation of answers. The training batteries address these features in a combined manner, and therefore provide a more realistic test preparation. While content-based WM-capacity training helps the students to recognize historical nouns and pictures more efficiently, reasoning-strategy training requires the students to actively retain strategy patterns in their historical knowledge. This implies that the students should have more time to actively work with the retained knowledge and reasoning concepts in the classroom and discuss reasoning problems with their peers. The teacher's role is then to give feedback to the individual, rather than provide a summary of what students need to know, and how students need to reason. Due to the combination of WM capacity, strategy training, and social studies content, the training provides an effective strategy to improve reasoning skills.

6 Improving reasoning abilities in secondary history education: The effect of motivational components in combined reasoning-strategy and working-memory training[14]

Summary

Combined working-memory- (WM) and reasoning-strategy training have been found to improve the historical reasoning achievements of students in secondary education. More generally, motivation and attention are considered important determinants in improving deductive reasoning because they are regarded as general (cognitive) functions within which WM is subsumed. In this study, we hypothesize that more engaged students are more likely to improve their reasoning skills caused by WM-capacity- and reasoning-strategy training to which motivational components are added. We test this hypothesis in an experiment in which four secondary classes of two schools in the Netherlands participated (107 16-year-old students). The experimental groups received the training and were motivated by the teachers. The control groups received the same training but with the exclusion of the motivational incentives. The empirical results indicate that motivation incentives help to maintain motivation during training significantly compared with the control group results. However, the experimental groups do not outperform the control groups in posttest results on reasoning abilities. This could be caused by the use of active control groups, class compilation, and the role of the teacher during the experiment.

[14] Chapter 6 is based on: Ariës, R.J., Ghysels, J., Groot, W. & Maassen van den Brink, H. (2016). Improving reasoning abilities in secondary history education: The effect of motivational components in combined reasoning strategy and working memory training. TIER Working Paper Series 17/12. ISBN: 978-94-003-0128-3.

6.1 Introduction

In Chapters 4 and 5, we found that combined working memory (WM) and reasoning-strategy training improves test scores in secondary educational humanities settings, such as social studies and history. In these chapters we combined WM- and reasoning-strategy training. Improvements of the experimental groups were compared with passive control group conditions (i.e. groups who were not trained differently to pre-training conditions). Therefore, posttest score differences between experimental and passive control groups could also be caused by behavioral characteristics, such as motivation or concentration levels during training. Since the combined training method requires focusing attention on an internal representation, we hypothesize that more engaged students are more likely to improve the reasoning skills as a result of the intervention to which motivational components are added. Motivation and attention are considered important determinants to improve deductive reasoning because they are regarded as general (cognitive) functions within which WM is subsumed. In this regard, attention contains an inclusionary component consisting of the content in focus, and an exclusionary component consisting of the inhibitory control of interference. This is supported by motivation theory which draws on the evidence that cognitive and motivational factors jointly influence academic achievement, causing motivated students to perform better on a trained task due to improved attention (Linnenbrink and Pintrich, 2002).

In this chapter, we hypothesize that test-score gains can be further improved when the intervention includes motivational factors that influence engagement levels. In this regard, Botvinick and Braver (2014) state that motivation and attention impact the intensity and direction of cognitive control (see also: Martin and Steinbeck, 2017). Since the training method requires students to focus their attention, and WM is weakened by distractibility, improving motivation and engagement may be a requirement for an effective intervention, as discussed earlier in Chapter 5. In the present chapter we investigate whether a combined content-based WM- and reasoning-strategy training has an effect on

achievements in school reasoning tests when a combination of three motivational features (i.e. effort-acknowledgements, goal-setting and feedback) is included (Marzano, Pickering and Pollock, 2001). The training was incorporated in history courses in two Dutch higher general secondary education schools. Cognitive control strategies include reasoning-related domain-specific cognitive and meta-cognitive WM tasks for which we modified the original tasks by Alloway and Gathercole (2009) and Mevarech and Kramarski (2003). The content-based approach of WM- and reasoning-strategy training is expected to result in improvements of reasoning achievements (Van der Sluis et al., 2007; Lee et al., 2009; Chapters 4 and 5). For this purpose, the experimental groups receive the content-based WM- and reasoning-strategy training to which motivational components are added. The motivational components consist of verbal praise, goal-setting and feedback, as determined in reward-based models (Botvinick and Braver, 2015). The control groups receive the same training, but without motivational components.

The training is implemented in existing history courses, and: (1) investigates posttest score effects when motivational components are added compared with the active control group's test scores; and (2) explores personal, social, intelligence, and motivational dynamics by means of regression techniques. We conclude that motivation incentives help to maintain motivation during training significantly compared with control-group results. The experimental groups do not outperform the control groups in posttest results on reasoning questions.

Reasoning in history education is defined as 'the process by which central facts and concepts are arranged to build a historical case' (Leinhardt et al., 1994, p.314). Evidence suggests that WM, the cognitive system that monitors higher cognitive processes in the brain, could be trained to improve reasoning (Klingberg et al., 2002; Olesen et al., 2004; St. Clair-Thompson et al., 2010; Alloway, 2012; Loosli et al., 2012; Shipstead et al., 2012; Melby-Lervåg and Hulme, 2013; Holmes and Gathercole, 2014; Titz and Karbach, 2014).

Differences in achievement outcomes in the experimental groups could be caused by individual motivational levels, since WM is enhanced by concentration and weakened by distractibility.[15] In fact, motivational context is commonly regarded as impacting the intensity or direction of cognitive control. This impact varies in 'systematic ways across individuals with (a) the baseline inclination to engage in controlled information processing and (b) general sensitivities to reward and punishment […] which positively predicts responses to incentives during cognitive tasks in reward and control networks' on neural levels (Botvinick and Braver, 2015, p.86). However, still in its early stages, neuroscientific research on the relationship between motivation and cognitive control provides increasing evidence of the relationship of motivation and control networks by means of reward-based decision making. This suggests that motivation during control activities is caused by the interaction of reward and control brain networks. The interaction between these two systems, also known as the 'dual systems model', could explain the maturational gap between cognitive control and affective (reward-based) processes in adolescence (Shulman, Smith, Silva, Icenogle, Duell et al., 2016). The early-maturing reward system peaks in adolescence, during which the cognitive control system has not yet matured, which can confer 'greater flexibility in adjusting one's intrinsic motivations and goal priorities' in adolescence (Crone and Dahl, 2012, p. 636).

In this regard, Botvinick and Braver's (2015) review of the literature on the interface between motivation and cognitive control has found that current research can be categorized into three basic theoretical perspectives: 1. force-field models; 2. resource models; and 3. reward-based models. (1) *Force-field models* strongly rely on the individual's energy investment in cognitive activities which is determined by effective driving forces, which, in turn, is a result of goal importance, mental resources, and task difficulty mechanisms. Action is hereby conceptualized as a result of forces, 'akin to physical forces, that attract

[15] J. Fuster (2015, August 11 – September 11), Personal communication.

an individual toward a goal or else impede progress toward that goal' (p.90). (2) *Resource model*s represent the mechanisms by which limitations of cognitive processing caused by limited resource quantities interact with depletion activities (i.e. the decline of cognitive performance and engagement) which reflect a reduction in the availability of a critical resource. (3) *Reward-based models* focus on reward maximization in which cognitive control is seen as a continuous series of decisions (costs) which pay off when performed well (rewards) in the target task. Cost-benefit maximization is then achieved by including individual domain-specific processing capacity limitations which can also be history-dependent to the task or subject. Although the presented models can be regarded as mutually inclusive alternatives, Botvinick and Braver (2015) point out that the reward-based perspective appears to generate the most promising results, for depletion effects can be prevented or postponed by including strong performance incentives towards motivation. This proves that both force-field and resource models are based on intrinsic motivation, while reward-based models are based on extrinsic motivation. Thus, features of reward-based approaches, such as goal-setting, verbal praise, and feedback, are incentives that are fit to be included in WM and reasoning strategy training. But most of the reward-based research is conducted by using monetary and game incentives or extracurricular training programs. Both incentives proved difficult to implement in schools, as ethical objections obstruct the implementation of monetary incentives in secondary education, while only a few schools can facilitate game-like interventions. Instead, we included a combination of currently accepted reward-based motivation features (effort-acknowledgements (verbal praise), goal-setting and feedback) to boost the students' motivation. These motivation features, commonly regarded as 'external rewards' (Brewer (ed.), 2014, p.15) are elucidated in the Intervention section.

6.2 Methods

6.2.1 Participants

Training took place in two schools for higher general secondary education (HAVO) (Table 6.1). In total, 107 students, divided over four heterogeneous history classes in higher general secondary education (intermediate track), participated in the study (Table 6.2). A priori power analysis indicated that a minimum of 102 subjects in the total sample is required to have 80% power for detecting a medium-sized effect (d=0.5) when employing the .05 criterion of statistical significance. The classes were formed by the schools exclusively based on schedule considerations, and not based on personal characteristics or IQ-measures. In the Netherlands, students can choose one out of four course trajectories (called: profiles)[16] when heading into the fourth year of higher general secondary education. These profiles consist of a number of mandatory courses and a number of non-mandatory courses. The history course is mandatory in two of the four profiles, and non-mandatory in two other profiles. Therefore, history classes are formed by looking at the best possible schedules for each student and teacher. This means that a history class usually contains students from all four profiles. Most students could not freely choose to follow the history courses, since history is a mandatory course for students who specialize in social-sciences and humanities education trajectories.

Table 6.1 Research design

Group name	Control/Experimental group	Teachers
Maastricht	Control	1 male teacher
	Experimental	
Heerlen	Control	1 female teacher
	Experimental	1 male teacher

[16] The four profiles are: Culture and Society*; Economy and Society*; Science and Health; and Science and Engineering (* depicts the profiles in which History is a mandatory course).

Table 6.2 Descriptive statistics and t-tests for equality in expectation per group

	N	Male	Mean age	SES	Avg. prev.	CITO
Total	108	0.36	15.6	-1.06	65.48	539
Experimental group	50	0.40	15.47	-0.90	64.82	540
Control group	58	0.34	15.76	-1.18	66.11	538
t-value		*0.59*	*2.10*	*-1.04*	*0.72*	*-2.05*
Sign.		*0.258*	*0.049*	*0.835*	*0.228*	*0.043*

Notes. Tot.= total; SES = socioeconomic status based on the Netherlands Institute for Social Research (SCP) neighborhood status scores compared with the national average score (0), https://www.scp.nl/english/Publications/Summaries_by_year/Summaries_2012/Neighbourhood_status_development_in_the_Netherlands_1998_2010; Avg. prev. = average history grade previous year; CITO = central end test in primary education provides information about the entry levels in secondary education for each child: entry levels for Higher General Secondary Education usually require a score between 537-544.

Every class contains about two dyslectic students. This can alter test outcomes because of the limitations in verbal impression or expression caused by the condition. An average of two students had repeated the class. Since, amongst normal trajectory students, the 4th year of the Dutch intermediate track of Higher General Secondary Education level can be accessed by graduated lower-track students as well as students who failed in the higher track, additional analyses are conducted in this chapter to control for these variables. The results of these students' tests are included in the main analysis.

We asked the parents for permission to use additional information from the students, such as their socioeconomic status, IQ-test scores, and family composition, in order to control for individual features. Students (N=15) whose parents did not give permission were not included in the additional analyses in which the background information was included. Participants were identified as lower, average, or higher performers, based on their pretest results concerning reasoning questions. All students lived within an average radius of 4 kilometers from their schools, in or near the urban areas of Maastricht and Heerlen in the Dutch province of Limburg.

All teachers had been appointed to the classes prior to the intervention. All teachers were experienced and qualified according to European teaching qualifications. In the lessons, new subjects were introduced by teacher presentation, following whole-class dialogs and discussions to answer textbook questions which students answered during class or at home (conventional instruction method).

6.2.2 Intervention

All groups were given a battery of well-known cognitive- (n-back, Odd One Out) and reasoning-strategy (IMPROVE) training methods. These batteries were identical to the training methods used in Chapters 4 and 5 of this thesis in order to comply with similar experimental conditions. We repeat the following description of the tasks from Chapters 4 and 5 on improvements in historical and social studies reasoning achievements by WM training.

In the n-back task (Appendix A), the participants saw a series of 20 domain-specific nouns in the n-back task which were successively presented in a PowerPoint presentation at the rate of 3 seconds per stimulus. The participants were instructed to remember the sequence in which the nouns were presented. Most of the nouns that appeared were presented in black, but when a green-colored noun appeared on the screen, the participants had to remember the previous noun and write it on an answer form within 5 seconds. The tasks were presented in 1-back, 2-back, and dual 2-back batteries, which are derived from the standardized Automated Working Memory Assessment test battery (Alloway & Gathercole, 2008). During the 1-back task, students had to remember the previous item, consisting of a historical noun. During the 2-back task, students had to remember the penultimate item, consisting of a historical noun. During the dual n-back task, students had to remember the penultimate item, which simultaneously consisted of a historical noun and image.

The Odd One Out (Appendix B) task presented four domain-specific nouns or images on a video screen in a PowerPoint presentation, each for

3 seconds during a total time of 12 seconds. In the sequence, one noun or image was the 'odd one out', which participants had to write down on an answer form when all nouns or images were shown.

All groups were trained on both tasks, the n-back and the Odd One Out. It should be noted that the content of these training tasks, which were content-based history subjects, differed from the outcome tests which contained content from other historical subjects. The dimensions of the cognitive training did not differ from the tests which contained 1-back, 2-back, dual 2-back, three verbal and three visual Odd One Out tasks during a section (30 minutes) of one lesson (50 minutes) per week. The content of the tasks was equivalent to the subjects concerned.

We used the standardized IMPROVE-method (Appendix D) for reasoning-strategy training (Mevarech & Kramarski, 2003). The training was performed in heterogeneous groups that contained three or four participants: one low-, one or two middle-, one high-achieving student. Four open-ended reasoning questions were successively presented, with 10 minutes to answer each question. The question had to be addressed by one of the groups' participants by using the IMPROVE-questionnaire to analyze the reasoning process. When the participant could not answer one of the questions in the questionnaire, the other members of the group could give directions for answering the question. After 10 minutes, a new problem was addressed by a new participant. The teacher participated in each group's reasoning process for 2 minutes per problem in the same way that the other members of the group participated. The tasks were trained in one 50 minute history lesson per week over a 6-week period.

In order to improve the achievements of lower performers on reasoning tests, we added a combination of motivational components to the experimental groups' intervention: 1. acknowledging the students' efforts (i.e. verbal praise); 2. goal-setting; and 3. feedback.

(1) Acknowledging the students' efforts, or verbal praise, is a proven method to encourage students to believe in their capabilities, and to believe that they can achieve better when they work harder. It is

commonly regarded as a powerful strategy to enhance motivation. For instance, in the 'Handbook of Sports Medicine and Science, Sport Psychology' (2009), Robert Weinberg discusses its significance in improving the well-being of athletes, specifically when being 'in a slump or experiencing a plateau in learning or performing' (p.16). Verbal praise has also been identified as being an important factor in enhancing student learning motivation in classroom settings (Hancock, 2002). Hancock's study showed that well-administered verbal praise by a professor caused significant improvements in motivation, time spent on homework and, consequently, test performance. Based on Hancock's (2002) experimental design, in the present study the teacher was instructed to ask each student in the experimental class to record how much the student was engaged during the WM task on the aforementioned answer form. Hereby, students had to place a dot on the line that represented the amount of engagement (0-100%). After waiting a minute for the students in the experimental groups to place the dot on the line, the teacher personally collected the answer form by walking around the room. Each time, the teacher looked briefly at the placement of the dot. If the dot had been placed at or more than 60% of engagement, the teacher said 'good job', 'great work', or 'very good' to the students, which are considered as 'overt expressions of praise' (Hancock, 2002, p.87). If the dot had been placed at less than 60%, then the teacher simply responded by saying 'thank you'. The verbal praise was given by the teacher at such a volume so that all students could hear the response. The teacher also collected the answer forms in the control groups, but did not respond to the amount of engagement.

(2) 'When people make a commitment to attempt to attain a goal, they are likely to compare their performances with the goals as they work on the task' (Schunk, 2003, p.163). Goals that incorporate specific performances are more likely to enhance learning according to Locke and Latham (1990). In our study, the experimental group students were instructed by the teacher to write down specific task-performance goals. At the beginning of each WM task, every student wrote down the

expected total number of good answers on the answer form. Teachers were instructed to check the number at the beginning of every task by walking through the classroom and randomly asking a maximum of 7 students to show the number. The control group students were not instructed to write down their expected number of good answers.

(3) 'Providing explanations during a task can facilitate problem solving performance' (Cheshire et al., 2005, p.1). On the basis of Cheshire et al. (2005), who concluded that performance improvements were the highest when teachers gave feedback and explanations, teachers were instructed to give feedback and explanations to the experimental groups' students during the reasoning strategy task. The control group students were only given teacher feedback when the teacher could respond to the students' reasoning strategy as being either correct or incorrect. Since individual reasoning strategies could contain numerous mistakes or questions, the teacher was instructed to respond to these to the best of his/her knowledge and capabilities.

6.2.3 Data collection

The motivation of all the individual students was assessed 1 week before and after training by digitally administering a modified version of the Tuan, Chin and Shieh (2005) SMTSL questionnaire. The SMTSL questionnaire measures students' motivation specifically toward science learning as opposed to more frequently-used educational psychological questionnaires (e.g. the Motivated Strategies for Learning Questionnaire, MSLQ) (Pintrich, Smith, Garcia and McKeachie, 1991) or the Multidimensional Motivation Instrument (Uguroglu, Schiller and Walberg, 1981), which were mainly developed to evaluate 'pre-determined motivation domains in understanding students' general learning motivation' (Tuan et al, 2005, p. 640). Tuan et al. (2005) point out that investigating student motivation in specific subject-content areas is important, as students may have developed different motivational levels in different areas. The questionnaire makes use of six factors of motivation (i.e. self-efficacy; active learning strategies; science learning

value; performance goal; achievement goal; learning environment stimulation). The formulation of questions in the original questionnaire was modified to fit the specifics of history learning and the Dutch language (Appendix E).

A pretest and a posttest (at the end of the 6-week training) of WM and reasoning achievements were administered to all groups. While the pretest controlled for homogeneity, the posttest measured short-term training effects on WM and reasoning tests. The experimental groups' results and the control group data were compared to evaluate differential increases in reasoning skills.

WM capacity was measured by administrating n-back and Odd One Out tasks, which are commonly regarded as reliable in testing WM capacity, as both tasks engage processes of managing information in WM and discourage the development of task-specific strategies or processes of automatization (Jaeggi et al., 2008; Holmes et al., 2009). Similar to the intervention, test tasks were subject-related, and therefore alterations of standardized tests, and used historical nouns and images to control for verbal and visual components of WM, which are considered important for historical reasoning. During one n-back training task, the participants saw a series of 20 items, which were successively presented in a PowerPoint presentation (3 seconds per stimulus). Three different n-back tasks were administered: 1-back, 2-back, and dual 2-back, which were described earlier in more detail in the Intervention section. For the Odd One Out task, four historical nouns or images were shown in turn in a PowerPoint presentation on a screen (3 seconds per stimulus). While three stimuli were related, a fourth stimulus was the 'odd one out'. Students ideally remembered all sequenced nouns or images, and wrote down the 'odd' stimulus afterwards on the answer form. The test included three series of four nouns and an equivalent number of images-series. Students could score a maximum of 58 points (1 point per good answer) on the WM-capacity test.

Students' answers on the reasoning questions of school tests, being part of school tests which also include knowledge questions, were

analyzed to test for reasoning skills. Each answer was analyzed on the basis of three criteria for meta-cognitive reasoning skills: contextualization; use of sources; and argumentation (Van Drie and Van Boxtel, 2008). The students' use of these criteria was rated by the teacher, and again by the researcher (Appendix C). As 10 reasoning questions were consistently included in each test, the participants scored 1 point per criterion with a maximum of 30 points per test. Furthermore, students' answers on the reasoning questions were also analyzed in order to evaluate the reasoning achievements, and assessed according to historical-content criteria based on examination models. The same school tests were used, but, in this case, only the correct responses to the reasoning questions were rated by the teacher, and afterwards by the principal researcher in order to control for inter-scorer reliability. The participants scored a maximum of 3 points per criterion per reasoning question (with a maximum of 30 points per test).

To further address inter-scorer reliability, as the intervention was implemented at different schools by different teachers, the schools provided additional test-result information to compare the test results with the students' former history test scores, scores of general cognitive ability, and socioeconomic status. It should be noted that all tests were compiled by the participating teachers and the principal researcher in order to standardize test levels between schools. All schools conducted tests on the same historical subjects. Teacher training in advance of the experiment also contributed to having homogeneous examination models between tests.

6.2.4 Design

The experimental groups' scores were compared with the active control groups' scores in reasoning, WM capacity and levels of engagement, since both groups received the combined WM-capacity- and reasoning-strategy training, and only the experimental groups received the additional engagement interventions. Pretests and posttests were employed to evaluate training effects in four 4th Grade classes of the

higher track ‘Higher General Secondary Education’ (HAVO), in which participants were nested in multiple classes in two different schools. In each school, one class was assigned to the experimental condition, while the other class was designated as the active control group.

Posttests from both conditions were evaluated and analyzed independently by the use of paired t-tests to measure the effects on WM capacity, reasoning strategies, reasoning achievements, and training engagement.

6.2.5 Procedure

The groups in both conditions studied the same subjects from the same textbooks. Prior to and during the intervention, all groups were trained in three 50-minute lessons per week. Also, starting conditions were the same for all groups, as teachers used approximately the same weekly instruction methods whereby they introduced new subjects in a presentation, the participants answered and discussed textbook questions with the teacher, and the teachers made some conclusions before addressing a new subject.

To further improve treatment fidelity, the principle researcher trained all participating teachers for 60 minutes 2 weeks prior to pretesting, during which the mechanisms of the training were explained. During the teacher-training, all information on training tasks and tests was provided to the teacher. During the intervention period, the principal researcher monitored the teachers’ skills, as well as the overall training progress by weekly contact.

The structure of the tests and training is elaborated in Table 6.3. Three tests – motivation, WM capacity and reasoning – were employed to measure reasoning achievements and course engagement. Furthermore, the study used additional data to address and evaluate individual differences (by age, gender, socioeconomic status, family composition, learning deficits) and classroom dynamics (gender composition, group size).

10 weeks after the start of the school year, all participants were pretested on WM capacity and motivation. The test was introduced to students by the teachers prior to the test in which the structure and the contents were explained. Topics of the content-based WM-capacity test included historical nouns and images from the studied subject. In the next lesson, students had to digitally fill in the motivation questionnaire, which took about 20 minutes (Appendix E). The official school test to measure knowledge and reasoning took place in a lesson in the following week. The school test was compiled and evaluated by the teachers based on the textbook, in cooperation with the principal researcher, in order to control for overall equality between schools. After the evaluation of the students' answers to the reasoning questions, the heterogeneous cooperative groups were formed to train on reasoning strategies. These groups consisted of one low-, one or two average-, and one high-achiever.

After the 6-week intervention period, posttests were conducted. The posttests on WM capacity and reasoning were compiled based on the historical subject concerned by the teachers and principal researcher. All groups were trained in weekly training cycles that included three 50-minute lessons per week. In the first lesson, the teacher introduced the subject by PowerPoint presentation and the participants answered the questions in the textbook. In the second lesson, reasoning strategy training took place. The third lesson included the 25-minute WM-capacity training. In the next 20 minutes, the textbook questions from the first lesson were evaluated.

Table 6.3 Structure of tests and training

Wk 1	Wk 3	Wk 4	Wk 5	Wk 6	Wk 7	Wk 8	Wk 9	Wk 10	Wk 11
TT	MQ WM1	Reas1	T	T	T	T	T	T MQ	WM2 Reas2

Notes. TT = teacher training; MQ = motivation questionnaire; WM1 = WM capacity pretest; Reas1 = reasoning pretest; T = training; WM2 = WM capacity posttest; Reas2 = reasoning posttest; Wk = week.

6.3 Results

In the Maastricht school, one teacher performed the intervention in both groups. In the Heerlen school, the groups were allocated to two different teachers. All cognition and school pretests used the same historical content which students had mastered prior to the pretest. A post-hoc power analysis shows that, given the achieved sample size (the intervention group = 50, and the control group = 58) and conventional power and significance levels (80% and 95%, respectively), an effect size of at least 0.50 can be detected.

We first analyze the comparability between the experimental and the control groups by using t-tests on differences in gender, mean age, socioeconomic status, average history grade in the previous year, and cognitive tests scores at the end of primary education (CITO scores). In addition, the cognitive and reasoning pretest, as well as the motivation inquiry which was administered prior to the intervention, is compared between groups. This analysis is followed by independent sample t-tests on all posttests in order to check for group differences, and paired samples t-tests for pretest to posttest differences to check for the group effects of the intervention. Then, we evaluate the effects of independent predictor variables on reasoning achievements posttest, and gains in reasoning achievements posttest by means of a multiple regression analysis.

Based on the t-tests of the group characteristics and of the pretests, we observed no significant differences between groups in all categories, except for mean age and IQ scores. (Tables 6.2 and 6.4). IQ-scores were administered at the age of 11, while the average participant is between 15 and 16 years old. Considering the fact that IQ-scores can change through development, we consider gender, socioeconomic status, and previous average grades for history tests as more important predictors for equality of expectations per group. Nevertheless, we use all characteristics as control variables in subsequent multivariate analyses to take potential bias into account.

As a first test of the effect of the intervention, we analyzed the posttest scores assuming equality between groups. Thus, we investigated the experimental effects by analyzing the t-test outcomes of all posttests between groups (Table 6.5). These analyses demonstrate no significant differences between groups with regard to mean scores in cognition, reasoning, and motivation.

Table 6.4 Independent samples t-test results and descriptive statistics for cognition and reasoning pretests and motivation inquiry between experimental and control groups

	Group	Mean Score	SD	df	t	p
Cognition Pretest	Experimental Group	34.46	5.811	102	-1.033	.978
	Control Group	33.33	5.306			
Reasoning Pretest	Experimental Group	16.10	4.089	105	-1.629	.572
	Control Group	14.84	3.883			
Motivation Pretest	Experimental Group	121.04	13.711	105	0.739	.634
	Control Group	123.02	13.837			

Notes. t = t-ratio for each 'Group'; Exp=experimental group; Con=control group; * = significance of values for each 'Group' <0.05.

Table 6.5 Independent samples t-test results and descriptive statistics for cognition and reasoning posttests and motivation inquiry between experimental and control groups

	Group	Mean Score	SD	df	t	p
Cognition Posttest	Experimental Group	38.57	6.483	97	0.172	.254
	Control Group	38.81	7.565			
Reasoning Posttest	Experimental Group	16.16	4.570	99	0.136	.892
	Control Group	16.27	3.188			
Motivation Posttest	Experimental Group	117.06	15.032	101	-0.301	.265
	Control Group	116.25	12.428			

Notes. t = t-ratio for each 'Group'; Exp=experimental group; Con=control group; * significance of values for each 'Group' <0.05.

We also compared each group's pre- and posttest scores for all tests with paired samples t-tests in order to check for achievement effects (Table 6.6). With regard to the maintenance of motivation during training, we observed a significant decline of motivation in both control groups. Motivation levels in the experimental groups, in which motivation incentives were included during the training program, were not significantly affected. We also observed significant improvements in the cognition tests for all groups, except for the Heerlen experimental group.

Despite these effects, both the Maastricht and Heerlen experimental group and the Heerlen control group's achievements in the reasoning posttest did not significantly differ compared with their pretest scores. But the Maastricht control group did improve significantly on the posttest compared with its pretest results.

Table 6.6 Paired samples t-test results and descriptive statistics for cognition and reasoning pretests and posttests and motivation inquiry

		Pretest		Posttest		Difference and effect	
		Mean	SD	Mean	SD	Mean diff.	t
Cognition	Maastricht Exp	34.67	5.247	42.48	4.823	-7.81	-7.551*
	Maastricht Con	34.69	5.350	42.31	5.584	-7.62	-6.493*
	Heerlen Exp	34.39	6.444	35.39	5.655	-1.00	-0.721
	Heerlen Con	31.33	4.836	35.44	7.777	-4.11	-2.347*
Reasoning	Maastricht Exp	15.00	3.398	16.35	4.811	-1.35	-1.569
	Maastricht Con	14.79	3.011	16.43	2.899	-1.64	-2.989*
	Heerlen Exp	17.08	4.454	16.00	4.436	1.08	1.289
	Heerlen Con	15.08	4.736	16.08	3.550	-1.00	-0.972
Motivation	Maastricht Exp	118.68	13.510	116.36	14.585	2.32	0.802
	Maastricht Con	123.71	15.482	119.54	11.228	4.17	2.095*
	Heerlen Exp	121.64	13.506	117.68	15.689	3.96	1.529
	Heerlen Con	121.82	11.994	112.96	12.677	8.86	3.886*

Notes. t = t-ratio for each 'Group'; * = significance of values for each 'Group' <0.05; Exp=experimental group; Con=control group.

To sharpen the analysis, we subsequently developed a series of multivariate regression analyses. First, we conducted multilevel regression analyses, but the multilevel model did not converge due to small sample sizes. Instead, we conducted linear regression analyses to determine whether student characteristics could partly explain the effects of motivation on posttest reasoning achievement and on pretest to posttest gain scores. Age and gender differences, as well as socioeconomic status and the scores of the pretests of WM capacity, motivation, knowledge, and standardized scores in reasoning achievement were included as commonly used and theorized predictor variables. Reasoning achievement test scores were standardized for pretest and posttest scores per school because the groups within each school took the same tests, while the tests between schools differed. Furthermore, teacher, history grades of the previous school year, and IQ-

test scores were included to explain differences in training improvement based either on magnification or on compensation accounts (e.g. Titz and Karbach, 2014). In this regard, it is considered that students who already perform well either benefit the most due to more efficient cognitive resources (magnification account) or benefit less from cognitive interventions, while performing on an optimal level can leave less room for improvements (compensation account).

Tables 6.7 and 6.8 represent the evaluation of these predictor variables that were evaluated through a multivariate regression analysis. Table 6.7 shows the regression effects on standardized reasoning achievements posttest scores. Table 6.8 depicts regression effects on gains in reasoning achievements of pretest to posttest scores per group. Five models were tested. In Model 2 the pretest scores of WM capacity, motivation, knowledge, and reasoning achievement were included as continuous variables. Model 3 represents an extension of Model 2, and shows interactions between the intervention condition and pretest reasoning achievements. Also, three groups of reasoning pretest scores (low-, middle- and high-performers) are distinguished in interaction with the intervention to check for compensation or magnification accounts, because magnification and compensation theories require non-linear specifications. Model 4 shows test interactions between the intervention condition and low-, middle- and high-achievers in reasoning pretests. In Model 5, teacher characteristics are included. Models 2A and 2B are included as additions to Models 2 and 3 and include gender, socioeconomic status, school type, IQ-scores, and previous history grade.

Table 6.7 shows that the intervention which included motivational incentives is not a significant predictor of posttest scores compared with the active control group's test scores. Models 2 and 2A show that there is a significant correlation between the standardized (continuous) reasoning achievements pretest score (p<0.01) and posttest scores. Furthermore, Model 2A implies that the history grade from the previous school year is a significant predictor (b=0.04, p<0.05). Models 2B, 3 and 4 also show the history grade of the previous school year as a significant predictor of

Table 6.7 Regression effects of independent variables on standardized reasoning achievements posttest

	Model 1 B (*SE B)*	Model 2 B (*SE B)*	Model 2A B *(SE B)*	Model 2B B (*SE B)*	Model 3 B (*SE B)*	Model 4 B (*SE B)*	Model 5 B *(SE B)*
Constant	0.127 (*.17)*	- 1.05*(1.04)*	-25.80 *(12.51)**	-25.58 *(12.47)**	-25.57 *(13.67)*	-27.26 (*14.06)*	-27.75 *(15.09)*
Intervention group (0/1)	-.12 *(.25)*	-.23 *(.23)*	-.26 *(.22)*	-1.27 *(.89)*	-1.23 *(.92)*	-2.17 *(2.42)*	-2.52 *(2.97)*
Pretest WM capacity		-.01 *(.02)*	-.01 *(.02)*	-.003 *(.02)*	-.004 *(.02)*	-.002 *(.02)*	-.003 *(.02)*
Pretest motivation		.01 *(.01)*	.01 *(.01)*	.004 *(.01)*	.01 *(.01)*	.004 *(.01)*	.004 *(.01)*
Pretest reasoning achievement (z-score)		.48 *(.11)****	.39 *(.11)****	.27 *(.15)*			
Pretest reasoning achievement low score (0/1)					-.14 *(.72)*	-.36 *(.87)*	-.32 *(.93)*
Pretest reasoning achievement medium score (0/1)					-.12 *(.39)*	-.38 *(.48)*	-.42 *(.50)*
Pretest reasoning achievement high score (0/1) (reference group)							
Pretest knowledge		.01 *(.02)*	.00 *(.02)*	.002 *(.02)*	.002 *(.02)*	.004 *(.02)*	.001 *(.02)*
Intervention group*Pretest reasoning achievement				.06 *(.05)*	.06 *(.05)*		
Intervention group*Pretest reasoning achievement low score (0/1)						.73 *(1.26)*	-.60 *(1.58)*
Intervention group*Pretest reasoning achievement medium score (0/1)						.70 *(.72)*	-.67 *(.83)*
Intervention group*Pretest reasoning achievement high score (0/1) (reference group)							
Gender (0/1)			-.20 *(.24)*	-.22 *(.24)*	-.22 *(.25)*	-.19 *(.25)*	-.20 *(.27)*
SES (0/1)			-.14 *(.23)*	-.19 *(.24)*	-.19 *(.24)*	-.21 *(.25)*	-.19 *(.28)*
IQ			.04 *(.02)*	.04 *(.02)*	.05 *(.03)*	.05 *(.03)*	.05 *(.03)*
History grade previous year			.04 *(.01)**	.04 *(.01)***	.04 *(.01)**	.04 *(.02)**	.04 *(.02)**
Age							*-.73 (.21)*
Teacher Intervention group Heerlen (0/1)							-.02 *(.46)*
Teacher Control group Heerlen (0/1)							-.09 *(.42)*
R^2	.003	.287	.391	.405	.406	.417	.418
N	68	68	68	68	68	68	68

Notes. *p < 0.05 **p <0.01 ***p =0.001 ****p <0.001; standard errors between brackets.

Table 6.8 Regression effects of independent variables on GAINS in reasoning achievements pretest to posttest

	Model 1 B (*SE B)*	Model 2 B (*SE B)*	Model 2A B *(SE B)*	Model 2B B (*SE B)*	Model 3 B (*SE B)*	Model 4 B (*SE B)*	Model 5 B (*SE B*)
Constant	1.47 *(.72)**	3.55 *(4.44)*	-84.66 *(53.54)*	-83.99 *(53.69)*	-86.91 *(58.89)*	-93.29 *(60.87)*	-99.40 *(64.09)*
Intervention group (0/1)	-2.12 *(1.05)**	-1.20 *(.97)*	-1.19 *(.95)*	-4.23 *(3.82)*	-4.15 *(3.94)*	-4.92 *(10.47)*	-9.77 *(12.60)*
Pretest WM capacity		-.00 *(.08)*	.03 *(.08)*	.03 *(.08)*	.03 *(.08)*	.03 *(.08)*	-.01 *(.09)*
Pretest motivation		.03 *(.03)*	.01 *(.03)*	-.002 *(.04)*	-.03 *(.04)*	-.003 *(.04)*	.01 *(.04)*
Pretest reasoning achievement		-2.15 *(.19)****	-2.61 *(.11)****	-2.97 *(.65)****			
Pretest reasoning achievement low score (0/1)					.12 *(3.08)*	.31 *(3.74)*	-.14 *(3.93)*
Pretest reasoning achievement medium score (0/1)					-.08 *(1.69)*	-.68 *(2.07)*	-1.05 *(2.13)*
Pretest reasoning achievement high score (0/1) (reference group)							
Pretest knowledge		.12 *(.07)*	.08 *(.07)*	.08 *(.07)*	.09 *(.07)*	.85 *(.07)*	.02 *(.09)*
Intervention group*Pretest reasoning achievement				.19 *(.23)*	.18 *(.24)*		
Intervention group*Pretest reasoning achievement low score (0/1)						.01 *(5.45)*	2.15 *(6.71)*
Intervention group*Pretest reasoning achievement medium score (0/1)						1.40 *(3.11)*	2.47 *(3.51)*
Intervention group*Pretest reasoning achievement high score (0/1) (reference group)							
Gender (0/1)			-1.06 *(1.03)*	-1.12 *(1.03)*	-1.11 *(1.06)*	-1.03 *(1.08)*	-.99 *(1.42)*
SES (0/1)			-.89 *(1.00)*	-1.04 (1.03*)*	-1.04 *(1.04)*	-1.02 *(1.07)*	-.32 *(1.20)*
IQ			.14 *(.10)*	.14 *(.10)*	.14 *(.11)*	.15 *(.11)*	.18 *(.12)*
History grade previous year			.17 *(.06)***	.17 *(.06)***	.18 (*.06)***	.18 *(.06)***	.16 *(.07)**
Age							-.22 *(.89)*
Teacher Intervention group Heerlen (0/1)							-2.03 *(1.80)*
Teacher Control group Heerlen (0/1)							-2.05 (*1.80)*
R^2	.059	.298	.404	.411	.412	.417	.441
N	68	68	68	68	68	68	68

Notes. *p < 0.05 **p <0.01 ***p =0.001 ****p <0.001; standard errors between brackets.

posttest scores (b=0.04, p<0.05). From the lack of statistical significance of the other independent variables, we conclude that knowledge and motivation in the pretest do not predict reasoning posttest scores significantly when analyzed jointly. Also, according to Model 5, teacher characteristics do not predict improvements in reasoning tests due to the training. Furthermore, differences in gender and socioeconomic status do not predict reasoning posttest scores significantly.

Posttest score improvements could have been caused by motivation levels and/or knowledge levels during training. In the Introduction to this chapter, we hypothesized that more engaged students are more likely to improve their reasoning skills. In this regard, we consider motivation and attention as important determinants in improving deductive reasoning, because they are regarded as general (cognitive) functions within which WM is subsumed. Also, because school-tests consist of domain-specific reasoning questions, we hypothesize that domain-specific knowledge levels contribute to reasoning and formulating answers to reasoning questions. We therefore conducted a mediation analysis to investigate the correlation between the training and posttest results using posttest motivation and knowledge scores, as the posttest scores of motivation and knowledge represent motivation and knowledge during the training. However, these mediators do not significantly predict reasoning posttest scores.[17]

Table 6.8 shows that standardized pretest reasoning achievement scores (Models 1 and 2) are significant negative predictors for pre- to posttest score gains. The pretest reasoning achievement (Models 2 and 2A) indicates that the gain in the posttest score is predicted to decrease by 2.15 and 2.61 units for every unit of increase of the pretest score (R^2 = 0.30 and .40). The pretest scores of WM capacity, motivation and knowledge, as well as gender, scores of general cognitive ability, and socioeconomic status do not predict pretest to posttest gains significantly. Models 3 and 4 indicate that low- and medium-score groups

[17] The results of the mediation analysis can be retrieved by contacting the corresponding author.

of pretest reasoning achievement do not improve significantly in their posttest scores when compared with the high-scoring reference group. The test interactions in Model 4, as well as gender, socioeconomic status, IQ, and school type, do not predict pretest to posttest gain scores. Also, gains in posttest scores are not predicted by teacher characteristics.

6.4 Discussion and Conclusion

In higher levels of secondary education, reasoning is regarded as an important skill. It predicts academic achievement in the humanities due to test compositions which require both knowledge and domain-relevant deductive reasoning. In turn, motivation contributes to academic achievement (Martin and Steinbeck, 2016). This experiment explored the effect of motivation in a combined WM-capacity- and reasoning-strategy training and its effects on school-test outcomes. The experiment suggests that the added motivation incentives play a non-significant role in the training program with regard to test achievements, demonstrating that motivation incentives (of the kind used here) do not add to the training to improve reasoning test achievements.

Table 6.6 shows that the test results have not improved for any group but one control group. This is not in accordance with our conclusions in Chapters 4 and 5 which showed significant improvements in posttest scores compared with pretest achievements with similar groups on humanities courses. Those studies compared the training groups' test scores to passive control groups who received a training that was no different from normal school-practice. Contrary, the present study compares groups that all received WM-capacity- and reasoning-strategy training, but with or without motivation incentives. The different outcomes of the training compared with either active or passive control groups are in accordance with Redick et al. (2015), who state that significant improvements in reasoning due to WM-training generally disappear when compared with active control group results (see also: Chapter 2). In the educational context, the absence of significant improvements can be ascribed to the nature of standard training practice

during humanities courses, in which conservative instruction methods are commonly used. Compared with conservative instruction methods, which often make use of class-based instruction and feedback, the training that is implemented in the present study provides active learning strategies, in which every student participates and receives feedback on an individual level. This, in turn, could have reduced the effect of integrated motivation incentives within the experimental group's training compared with the control group results.

The absence of significant improvements in the experimental group's test scores compared with the control group's results could also have been caused by the nature and practice of the motivation incentives. First, the incentives we utilized to motivate the students of the experimental group (approval, feedback, goal-setting) are, or should be, standard practice in any well-organized school system to enhance the motivation and achievement of students. The absence of such motivation incentives during the training might have adversely biased the education of the control students. Furthermore, not motivating students may have consciously or unconsciously prejudiced teachers and experimenters during the experiment. Second, motivation levels could have also been affected by the WM-capacity training battery that lacked an adaptive form which is considered to improve WM capacity and maintain motivation. Third, as well as being affected by motivation incentives, cognitive control processes are activated by the 'presence of peers, task instructions, strategies and the affective appraisal of the value or priority of performing the task', which are suggested to influence the engagement of cognitive control systems during adolescence (Crone and Dahl, 2012 p. 639). These external factors might have decreased the effect of motivation incentives to transfer of effects in reasoning posttest scores, as implemented by the teacher, to non-significant levels. With regard to the latter, it should be noted that pretest to posttest levels of motivation in the control groups significantly decreased, while the experimental groups successfully maintained their motivation levels during the training. However, the external validity of the results owing to

the small sample sizes in all participating schools should be treated with caution.

Lastly, non-significant improvements could have also been caused by the way that school-tests are compiled. Contrary to national exams, school-exams are compiled by the teachers themselves and consist of different subjects. Pretest and posttest compositions could, for example, have varied in line with domain-specific knowledge and skills, although all school tests during this experiment were composed by the participating teachers and the principal researcher in order to standardize test levels between schools. We argue that, besides humanities studies school tests, more objective measures are required to analyze within-group reasoning-skill levels. For instance, during the present experiment, the pretests commonly contained questions regarding the classical Greek and Roman period, while the posttests generally consisted of questions about the Middle Ages. Accordingly, the students' domain-specific and contextual knowledge levels could have varied in agreement with the historical subject. This, in turn could have biased the posttest scores. We tested for subject-knowledge levels in the pretests and posttests, and included those as a co-variable in our OLS in Tables 6.7 and 6.8. Yet, subject-knowledge levels did not predict pretest to posttest improvements.

The results of the regression analysis further indicate that the posttest reasoning achievements score increases by .31 units for every unit increase of the pretest achievement scores of the intervention group. However, low reasoning pretest score groups generally do not improve more in their posttest compared with high pretest score students (Table 6.7). Furthermore, the test outcomes do not seem to be predicted by gender, socioeconomic status and IQ-scores. We therefore conclude that, in order to develop a training program to improve the performance of specific groups of students, more research has to be conducted to address their specific needs with respect to the acquisition of reasoning skills, as well as to investigate who can benefit from the training. We support Jaeggi et al.'s (2011) argument for a more individual

approach in determining what training regimes and conditions can result in the best transfer effects.

7 Conclusions and Discussion

Evidence on school-based WM training to improve deductive reasoning is still limited. School-based WM training is not yet integrated in secondary education curricula. Consequently, evidence of improved reasoning in education caused by WM training remains poor. The objective of this dissertation was to test a training program which includes content-based WM-capacity- and reasoning-strategy training to improve students' reasoning achievements in humanities courses. We have hypothesized that WM training is effective in educational settings when it is accompanied by reasoning-strategy training with subject-content in the humanities. Chapter 2 of the dissertation provided a review of the literature of the effects of one specific WM task, n-back, on reasoning achievements. It addressed the question whether a standard WM task improves reasoning achievements in education. Chapter 3 of the dissertation provided a synthesis of the literature on WM-capacity- and reasoning-strategy training in primary and secondary education. In addition, it explored the possible effects of the combined training program in education settings for the development of problem-solving skills. Chapters 4 and 5 analyzed the effects of such a training program in two experimental studies. Their purpose was to inform us whether the combined training is more effective to improve reasoning achievements than conventional training or exclusive WM-capacity training. Chapter 6 explores motivational factors related to the combined training, based on the pre-assumption that the success of the combined training is largely determined by the student's motivation during training.

7.1 Key Conclusions

In the Introduction to the dissertation we formulated six research questions. Table 7.1 summarizes the main contributions of this dissertation and the answers to the research questions and refers to the chapters where the questions were addressed. This concluding chapter now combines the findings of the various chapters by discussing the Key Conclusions, policy implications, and limitations of our research. Research

questions 1 and 2 provide arguments for Key Conclusion 1, which argues that there is no convincing evidence that stand-alone WM-capacity training can improve school-based reasoning achievements. Key Conclusion 2 provides information for the effectiveness of a combined WM-capacity- and reasoning-strategy training to improve reasoning achievements, and addresses research questions 3 and 4. Key Conclusion 3 deals with the effect of motivation on training results and addresses research question 6. Finally, Key Conclusion 4 addresses the question to what extent low or high achievers benefit from the training.

We organized the Key Conclusions of this dissertation around four statements.

1. ***There is no convincing evidence that (content-based) working memory training alone improves adolescents' reasoning skills in education.***

WM-theory suggests that WM training might transfer to gains of fluid intelligence (*Gf)* and, possibly, reasoning skills. Contrary to this, we did not find convincing evidence that WM training could produce far-transfer effects to school-based reasoning improvements.

In Chapter 2 we focused on n-back training because it could lead to positive effects on *Gf* and, potentially, reasoning. Our review of the literature indicates that no improvements have been measured in reasoning school-tests when n-back training is incorporated in secondary school curricula and trained over a period of approximately 4 weeks. This observation supports the review results of Redick et al. (2015) and Melby-Lervåg et al. (2016) who also did not find significant far-transfer results to reasoning in education.

When non school-based interventions that focus on WM training in relation to *Gf* outcomes are reviewed, we find that the majority of interventions found no far transfer to reasoning. In Chapter 3 we reviewed whether batteries of WM tasks could serve as an effective training method to enhance school-based reasoning achievements.

Table 7.1 Main contributions in terms of contents per chapter

Research question, (chapter)	Main findings
1 Does n-back training improve school-based reasoning achievements? (Chapter 2)	N-back training generates small significant to non-significant effects on reasoning improvements in reasoning posttest-scores. Future research designs should address methodological considerations and individual differences, before proceeding to implement n-back training in school curricula, such as motivational issues, personalizing training, baseline WM capacity and reasoning skills, and minimizing distractions.
2 Can WM-capacity- and metacognitive training serve as an effective instrument to enhance school-based reasoning achievements? (Chapter 3)	Both could improve reasoning skills, but it remains unclear whether transfer-effects to domain-specific reasoning can be expected. Few studies are found that focus on school-based interventions.
3 Does WM-capacity- + metacognition training, based on subject-matter knowledge in history education, have an effect on students' achievements in tests for which reasoning abilities are required? (Chapter 4)	WM capacity and reasoning strategies of the experimental group improve significantly after 3 and 6 weeks of training. Students achieve significantly better in school tests in which reasoning abilities are tested. The gain in reasoning abilities is also still significant 16 weeks after the training program is completed.
4 Does WM-capacity- and/or metacognitive training, when independently trained, have an effect on students' achievements in tests for which reasoning abilities are required? (Chapter 4)	WM-capacity training does not improve reasoning achievements compared with the passive control group. Reasoning-strategy training does improve reasoning achievements compared with the passive control group and with the WM capacity group.
5 Does the combined training have an effect on students' reasoning achievements in school-tests in other humanities courses (i.e. social studies)? (Chapter 5)	WM capacity and reasoning achievements of the experimental groups improved significantly after 4 training weeks compared with passive control group results.
6 Does the combined training affect reasoning achievements in school-test results when motivational factors are added to the training? (Chapter 6)	When motivational components are added to the training, the experimental group did not improve significantly in reasoning test achievements compared with the active control group test scores.

When training batteries, including a variety of WM tasks, have been used, then the results seem to be ambiguous regarding improvements in reasoning tests. WM interventions appear to be the most effective when implemented in primary education. However, the impact of interventions on reasoning declines in secondary and tertiary education. It could be

assumed that the period between the ages of 5 and 11, contrary to adolescence, may serve as a period of increased susceptibility to the development of reasoning skills.

Alternative explanations for the ambiguous results in secondary and tertiary education are provided by the nature of the students who participated in the interventions. The majority of participants in the n-back training interventions had attended university or preparatory university levels of education (e.g. Chapter 4, Experiment 2). One might consider that this rather homogeneous group already had above-average WM-levels, which has been theorized to limit WM training benefits. In fact, according to the compensation account theories, students who already function at high or optimal levels benefit less from cognitive interventions, because of high cognitive baseline levels which leave less room for improvement (e.g. Titz and Karbach, 2014, p.861).

This suggests that important individual cognitive differences could have determined the extent of transfer to reasoning. For instance, Jaeggi et al. (2011) found that transfer only occurred when the participants also improved considerably on the training task. This suggests that far-transfer gains are most likely to occur with participants who are functioning at below-average WM-levels.

2. ***Combined content-based working-memory-capacity and reasoning-strategy training appears to be an effective instrument to enhance school-based reasoning achievement.***

The second conclusion addresses the effectiveness of a combined WM-capacity- and reasoning-strategy training in school curricula in order to enhance the reasoning achievements of adolescents. In Chapter 3, we theorized that two aspects of WM appear to affect reasoning skills: namely, WM capacity and meta-cognitive WM. We theorize that it is appropriate to include strategies of pattern recognition consisting of metacognitive skills into school-based reasoning training in order to acquire the relevant cognitive tools for reasoning. This can lead to intuitive reasoning based on tacit knowledge. We found few studies that

specifically focus on school-based interventions and no studies that investigated the combined training battery compared with the (active) control groups on the higher general secondary education level. Furthermore, no neuroscientific data are available that support the claim of the possible added value of a dual WM-capacity- and meta-cognitive training battery.

Considering the educational value of the combined training battery in humanities education, which has been the main focus of this dissertation, we can conclude that a combined content-based WM-capacity- and reasoning-strategy training could serve as an effective instrument to enhance school-based reasoning achievements. Two experiments in this regard were conducted. The experiment in secondary history courses (Chapter 4, Experiment 1) investigated the combined intervention compared with the passive control groups. The results show that task-to-test training has considerably improved WM capacity. Even more important, the 6-week training period improved the students' achievements in reasoning tests considerably. Furthermore, the training proved to generate durable effects with regard to reasoning achievement improvements.

Subsequently, the combined training was duplicated in social studies courses in order to analyze whether test outcomes from the history courses training are replicated (Chapter 5). This can give insight into the extent to which the test outcomes in history are applicable in all of the humanities courses. The main difference compared with the experiment in history courses is that the training content has been adjusted to social studies curricula. WM capacity improved significantly and so did overall reasoning achievements in the experimental groups compared with the control groups who had received no specific WM-capacity- or reasoning-strategy training.

Overall, the training effects have shown to be durable considering that these effects remain significant 8 to 16 weeks after the training program was completed. This implies that reasoning structures were internalized in WM. Other causes for the training effects are found in the increased

knowledge of historical concepts and technical terminology due to intense and personalized content-based WM-capacity- and reasoning-strategy training, in which all students were triggered to participate, in contrast to whole-class teaching methods.

Another asset of the training was found in the integration within the course curricula. Level 4 classes on the Higher General Secondary School track usually receive three 50-minute lessons per week of history and/or social studies courses. In this weekly time span, teachers report that, besides executing the training program, enough time is reserved for explaining the chapter contents and for addressing book questions and student questions. By using a content-based approach, teachers and students have been able to focus on the subject matter with the training considered as an additional tool to improve reasoning.

3. Reasoning achievements in school-test results do not seem to be affected by motivation levels when students use the combined training method.

During training, when students reorganize and manipulate retained information to solve a reasoning question or WM task, maintaining attention becomes a major contributor to focusing on the problem concerned. The ability to exert cognitive control over one's thoughts and actions is of crucial importance to the capacity to reason and avoid distractions during a training session (Crone and Dahl, 2012, p.638). But the period of middle adolescence (13-17 years) is regarded as a period during which adolescents have problems in the regulation of affect and behavior (Steinberg, 2005). Motivational aspects appear to influence the activation of cognitive control, such as the 'presence of peers, task instructions, strategies and the affective appraisal of the value of priority of performing the task' (Crone and Dahl, 2012, p. 639, Martin and Steinbeck, 2016). During middle adolescence young people consistently react more strongly to receiving rewards, but not to the expectation of being rewarded, compared with children and adults. Motivation is believed to contribute to the adolescents' focus on the content of the

training task and the inhibitory control of interference. WM is enhanced by concentration and weakened by distractibility. Subsequently, the level of attention is considered to play a major role in the effectiveness of the training.

In Chapter 6, the effect of added motivation incentives (verbal praise, goal-setting, and feedback) to the training was assessed and compared with an active control group. The active control group received the same training and tests as the experimental group, but without added motivation incentives. A paired samples t-test was used to analyze the results for reasoning achievements within groups. When compared with the pretest results, we did not find significant improvements in the reasoning posttest for all but one group, a control group, after a 6-week training period.

Also, posttest results were compared between groups to investigate the experimental effects. The analyses demonstrated no significant differences between groups with regard to the mean scores in reasoning achievement results.

Group motivation levels are measured by comparing each group's pretest and posttest scores regarding the motivation questionnaire. We observed a significant decline of motivation levels in the control group. But motivation levels in the experimental group were not significantly affected. These results indicate that training on WM and reasoning strategies might require huge cognitive loads that can easily affect motivation levels. While the WM training is constructed in a way that it increasingly requires more cognitive load to give correct response, the motivation levels of students are likely to decrease during the training, especially when students discover that their peers are struggling as well. As a result, they can get distracted and can not maintain a high level of attention. As for the reasoning strategy training, students who are regarded as novices in reasoning still have to acquire generic memories-developed schemas to solve problems. according to Goldberg (2010), this development causes a transition from an absence of effective behavior of WM to the formation of effective behavior in reasoning processes. The

results also indicate that the motivation incentives that are used during the training in the experimental group are shown to be effective in maintaining motivation. The higher levels of motivation do not account for significantly improved reasoning test scores compared with the control group when implemented in high school curricula. Our conclusion is that motivation levels do not predict the effects of WM and reasoning strategy training, which is good news for teachers who implement the training into their curriculum, because they can concentrate on the learning process of the students rather than invest time in keeping the students motivated.

4. ***The effects on reasoning tests of the combined working-memory-capacity and reasoning-strategy training do not differ between low or high achieving students.***

The fourth conclusion addresses the effectivity of the training for high achievers and low achievers. Theoretically, training benefits could be attributed to students who already perform well on reasoning questions because they can tap into more efficient resources (i.e. the magnification account) (Titz and Karbach, 2014). On the other hand, they could benefit less from cognitive interventions because they already act on an optimal level, which can leave less room for improvement (i.e. the compensation account). In addition, Jaeggi et al. (2011) found that the transfer of WM training to reasoning only occurs when children have considerably improved while on the training task, which suggests possible far-transfer gains with groups of participants with below-average WM-levels. By focusing on different groups of achievers, we can provide insight into the possible benefits of the training for high or low achievers.

In this connection, we focused on the impact of reasoning pretest scores from low and high achievers in Chapters 5 and 6. Regression analyses with interaction between the intervention and the achievement level of the students showed that the combined WM-capacity- and reasoning-strategy training does not specifically benefit high achievers with respect to reasoning posttest scores in both studies.

Low-achieving students seem to benefit equally compared with high achievers from the implementation of the training in terms of improved reasoning posttest scores. These results are shown in Chapter 5, as well as in Chapter 6, where motivational incentives were added in order to improve the maintenance of the participants' attention. On the basis of these findings, we conclude that the combined training does not discriminate between low or high achievers with higher secondary education in humanities courses, and that they all benefit equally from the training, irrespective of their intellectual, socioeconomic, or behavioral characteristics. The results do not support any of the theories which propose that cognitive training benefits either low or high achievers.

7.2 Methodological issues

Non-school-based WM training is usually used as an instrument to investigate its effects on measures of *Gf* (e.g. matrix reasoning tasks). But matrix reasoning tasks are not to be compared to school-based reasoning in which other aspects, such as knowledge of contents and course-specific skill acquisition, are regarded as being important contributors to school-based reasoning test results. We argue that non-school-based WM training can not be implemented in school curricula in order to train reasoning skills. In this regard, we propose that a number of methodological considerations need to be addressed before implementing WM training in school curricula to improve reasoning skills.

First, a highly debated issue concerns the replicability of far transfer effects if the original training tasks are implemented in school curricula to improve reasoning achievements. This debate centres on the question whether the originally developed general and non-school-based WM training tasks are suitable for implementation in education curricula to generate far-transfer effects to reasoning achievements. The review studies described in Chapters 2 and 3 indicate that general non-school-based WM training does not provide the necessary transfer effects to improve children's reasoning skills in such a way that they improve

significantly in school-based reasoning tests. Therefore, implementing original training tasks one-on-one would be ineffective. Standard WM-training tasks should be tailored to suit the demands of school-based reasoning. But even when the WM tasks are content-based, then the effects on school reasoning test results may still remain substantively insignificant, as Chapter 4, Experiment 2 shows. In this regard, our findings support the proposition that WM-capacity training should be accompanied with reasoning strategy training.

The question whether the content of the originally developed general and non-content based WM training is suitable should now be addressed. Contents in standard WM training are usually focused on *Gf*. For instance, the stimuli in n-back tasks, a frequently-used training task in this PhD-research study, usually consist of letters or numbers. Students in humanities trajectories reason on historical and real-life subjects. We argue in Chapters 2 and 3 that adolescents are not yet able to gain far-transfer effects to content-based deductive reasoning in school tests based on a general non-content-based WM-capacity training. This observation could imply that WM-capacity training is more effective when the contents of the training match the subjects that are addressed and tested in the training period.

Another issue addressed in this PhD-thesis concerns the development of an effective training method to improve reasoning skills and reasoning achievements for students of varying backgrounds and baseline levels. So, we have included several predictor variables in the regression analyses in Chapters 5 and 6 to investigate whether the training favors specific groups. Our findings show that the training does not have different results with different groups of students. Adolescents with low or high achieving baseline levels in reasoning tests do not systematically differ in their posttest scores. Also, gains in posttest results can not be attributed to gender differences, socioeconomic status, or general cognitive ability. Furthermore, motivation levels, and content-based knowledge levels, as well as class and teacher parameters, which could be directly influenced, do not appear to contribute to the variances

in gains of reasoning test achievements. We conclude, therefore, that a one-size-fits-all training is not likely to include all students in the possible benefits of the training.

Lastly, we find that, when motivation incentives are added, motivation levels are higher during training compared with the levels in the active control groups. However, these higher motivation levels do not appear to affect achievement levels in WM-capacity and reasoning posttests compared with the active level control groups. These results do not support the academic literature on the relationship between motivation levels during training and advanced WM levels and reasoning achievements through improved attention levels.

Motivation and attention are considered important determinants in improving deductive reasoning because they are regarded as general (cognitive) functions within which WM is subsumed. In this regard, attention includes an inclusionary component consisting of the content in focus, and an exclusionary component consisting of the inhibitory control of interference. This is supported by motivation theory which draws on the evidence that cognitive and motivational factors jointly influence academic achievement, causing motivated students to perform better on a trained task owing to improved attention (Linnenbrink and Pintrich, 2002). The absence of improved reasoning test-achievement scores due to the maintenance of motivation levels implies that no extra motivation incentives need to be added to the training. But this absence of significant improvements in the experimental group's test scores compared with the control group results could also have been caused by the nature and practice of the motivation incentives, the use of active control groups, or the way that the school-tests are compiled.

7.3 Study limitations and future research

The review and empirical studies in this dissertation have filled some of the gaps in our knowledge and have led to the identification of relevant methodological and policy implications. Nevertheless, they also emphasize that there remains a gap between the theories on the effects

of WM-capacity- and reasoning-strategy training to reasoning and the implementation of such training in school-curricula to improve domain-specific reasoning. These limitations and their implications for future research are discussed below.

The main limitation of the review studies in Chapters 2 and 3 is that improvements in reasoning achievement by WM-capacity training cannot be guaranteed. It is not yet clear which (combination of) WM-capacity tasks can improve school-based reasoning and whether content-based WM training is preferred over non-content-based WM training. Nevertheless, the studies do shed light on the added value and pitfalls of implementing WM training into school-curricula. We argue against the implementation of only WM training as a whole-group training method, and instead suggest that future research should focus on the possible benefits of the training per school-level, age, and cognitive differences. Lilienthal et al. (2013, p.135) conclude that many studies lack a critical analysis of the mechanisms that cause reasoning achievements (e.g. skill internalization and executive attention). Future research should focus on a critical analysis of the mechanisms that cause reasoning achievements through WM training and also focus on motivational aspects that affect the students' concentration levels during training. Studies which investigate the effects of WM-capacity training in our target-group (i.e. adolescents) are still limited, especially with regard to school-based training. Therefore, no firm conclusions can be drawn concerning the effectiveness of the training, and future research is needed to gain more insight on the training implications for this particular group. This is also the main limitation of the review (in Chapter 3) that discusses the theoretical backgrounds of the combined WM-capacity- and reasoning-strategy training. While WM capacity is theorized to use reasoning schemas stored in long-term memory (LTM), more insight is needed into the mechanisms that underlie this cognitive structure. The review in Chapter 3 does not give insight into whether combined WM-capacity- and reasoning-strategy training actually affects the WM function more compared with either only WM-capacity- or reasoning-strategy training.

Future research needs to address this mechanism in order to develop a more effective training that takes into account the duration of the training and possible ceiling effects.

With regard to the empirical studies in Chapters 4, 5 and 6, the main limitations include the small sample sizes and test reliability. The training led to significant improvements in reasoning achievements in two studies, included in chapters 4 and 5. But small sample sizes could have led to underpowered studies which, in turn, could result in type II errors. Also, the groups were randomized at class-level, but not at the individual level, which suggests that there can always be other factors at play that could explain any correlation. Future research should replicate the training with fully randomized groups and include a larger sample. Also, the tests were compiled by the teachers, and the content of the reasoning posttests could not be compared with the content of the reasoning pretests. Testing did not contain standardized test batteries. We argue that these non-standardized test-batteries are less valid to measure the transfer effects of reasoning. Teachers played a role in test-grading, which did not add to the validity of test outcomes, and, furthermore, teachers could be consciously or unconsciously prejudiced during the experiments. Also WM-capacity tests are content-based and thus not standardized. We propose that future research makes use of standardized test batteries to measure content-based reasoning and WM capacity. Other methodological limitations include the use of non-adapted WM training and the use of WM-capacity test batteries that include the same tasks as the training batteries. With regard to the former, it is suggested (and disputed) that adapted training of WM affects WM capacity more than non-adaptive training would. With regard to the latter, the test results of WM capacity could have been influenced by the experimental group students' increased familiarity regarding the task concerned. Future studies should compare test results between adaptive and non-adaptive task training, and use test batteries that are different from the training batteries in order to improve test reliability and test for near-transfer effects. Furthermore, future research needs to focus on the

long-term training effects and the far-transfer effects of the training to other courses in order to improve the efficiency of reasoning training. Another possible research goal would be to explore to what extent student characteristics could predict training success. Through a better understanding of these mechanisms, we could eventually develop personalized training so that low-achieving students can benefit more from the training and make up ground with regard to their middle- and high-achieving peers. Future studies should look further into the behavioral and cognitive characteristics that could predict the impact of the training, such as motivational incentives, knowledge acquisition, learning strategies, the influence of peers, and teacher characteristics in order to produce a teacher- and student-friendly training program that could be based on personalized learning, in order to improve the reasoning skills of those students who need it the most.

On the basis of these considerations, and taking into account the limitations of our dissertation, we propose that the combined training should be integrated into the humanities school-curricula when the current curricula accentuate frontal-learning instruction. When implemented in the curriculum, the training is likely to affect the students' test-scores which would enable more students' to graduate. In this case, educators could pay attention to the procedure of the training program before implementing it in the curriculum. First of all, a minimum training period of 4 weeks could obtain significant improvements in reasoning which would remain durable up to 8 weeks after the end of training. Second, training should be implemented in one and a half lessons (75 minutes) out of a total of three lessons per week, in which the WM capacity is trained for 25 minutes, and reasoning strategies are trained for 50 minutes. Hence, the training becomes part of a larger training cycle in which the teacher can opt to introduce the subject by means of conventional instruction by discussing domain-specific skills and workbook questions with the students. With regard to the WM-capacity training, teachers should include in the training tasks contents that are

related to the subject concerned and include multiple (levels of) WM tasks in order to take into account individual WM-capacity levels. As for the reasoning strategy training, discussion groups should consist of no more than four students. This way, a training cycle could be completed in a single 50-minute lesson, and keep students engaged in the training for an entire lesson. Furthermore, groups should consist of a lower-achiever; one or two middle-achievers and a higher-achiever, so that students can give feedback and help each other in order to answer a reasoning question by means of the IMPROVE-questionnaire. That way, when low-achievers and under-motivated students actively participate in reasoning training and are aided by their peers, the training can contribute to establishing a culture of excellence in secondary schools.

7.4 Policy implications

Despite recent developments in didactic approaches, direct instruction is still a commonly-used method to teach knowledge and reasoning skills. Direct instruction refers to teaching activities that take place in front of the classroom, during which the teacher instructs all students at the same time. This teaching method implies that teachers do not consider the students' personal baseline levels of knowledge and cognition. With regard to knowledge acquisition, we consider the teaching of curriculum content to be an important instrument for students to acquire the specific knowledge essential to address course-specific reasoning questions. Our study indicates that the levels of humanities content (domain-specific factual knowledge) do not predict reasoning skills levels. Therefore, efforts should be made to understand and balance the extent of the transfer of both knowledge and reasoning skills during lessons.

Compared with direct instruction designs, we conclude that our training method is a much more effective instrument for students to acquire the necessary reasoning skills needed to adequately respond to a reasoning question (*d = 1.04-2.20)*. A plausible explanation for the more improved reasoning-achievement levels can be found in the way that

students approach reasoning training. Contrary to direct instruction, the current training assumes that all students are actively engaged in solving reasoning questions and WM-capacity training. Therefore, we conclude that reasoning skills are trained more effectively by using active didactic approaches in order to actively involve all students in the training program.

Both the absence of clear relationships between motivation levels and cognitive levels, at least as observed in higher general secondary education, and the socioeconomic background variables give some insight into the usability of the training in personal learning settings in instructional designs. Personalized learning designs can tailor learning activities to the needs of every individual learner's cognitive skills in order to obtain higher learning outcomes. This type of education has become increasingly integrated in school curricula in order to accommodate the acquisition of, for example, 21st century skills (Office of Educational Technology, 2017, pp.9-10; Onderwijsraad, 2017; Schoolinfo[18], n.d.). While the training generally improves reasoning achievements, we have found no indication for the effectivity of the training in education settings designed to train specific groups (e.g. students with learning disabilities or gifted students). Nevertheless, we argue that the training in general can be implemented in high school curricula to effectively improve the students' reasoning skills in the humanities. However, the training does not provide a clear-cut instrument to target specific groups who need to be trained using personal learning designs. In fact, we conclude that, in order to develop a training program to improve the results of specific groups of students, more research needs to be conducted to address specific needs concerning reasoning-skill acquisition, as well as investigating who can benefit from the training.

[18] Schoolinfo is the author of the website 'Leerling2020' (Student2020), commissioned by the Dutch Secondary Education Council (VO-Raad). The goal of this project is to accelerate the implementation of school-based innovations of personalized learning into Dutch education.

The last policy implication is based on Chapters 4, 5 and 6, in which the effect of the training on reasoning skills at different education levels and in different courses in the humanities is assessed. The findings imply that the training, when focused on course-based content, can be implemented in all courses in the humanities, such as history and social studies, to improve reasoning skills. In that case, we want to emphasize the importance of integrating a standard reasoning training in secondary education to show students that the same reasoning structures can be applied in different courses. By creating awareness in this matter, schools can offer more efficient and effective learning strategies to students and invest in interdisciplinary education concerning the acquisition of reasoning skills. This implies that the training could also result in improved reasoning skills in the lower and higher tracks of education. Lower and higher track students may require different strategies that fit their skill-levels and the requirements of the different tracks by means of skill-acquisition (e.g. Chapter 4, Experiment 2). In the lower tracks, for instance, reasoning requirements could be adjusted to lower cognitive levels. On the other hand, higher tracks education requires more abstract reasoning. We therefore support Jaeggi et al.'s (2011) argument for a more individual approach in determining what training regimes and conditions result in the best transfer effects.

Bibliography

Agin, S. (2001). The effectiveness of using brain based strategies in classroom instruction to enhance student learning. *Learning Disabilities*, Rowan University. MA: 29.

Alloway, T. (2012). Can interactive working memory training improve learning? *Journal of interactive learning research, 23*(3)*,* 197-207.

Alloway, T., Gathercole, S. E., Kirkwood, H. & Elliot, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology, 28*(7), 725-734.

Alloway, T., Gathercole, S. E., Kirkwood, H. & Elliot, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80*(2), 606-621.

Ananiadou, K. & Claro, M. (2009). 21st century skills and the competences for new millennium learners in OECD countries. OECD Education Working Papers, 41, OECD Publishing.

Ariës, R.J. & Cabus, S.J. (2015). Parental homework involvement improves test scores? A review of the literature. *Review of Education, 3*(2), 179-199.

Ariës, R.J., Groot, W. & Maassen van den Brink, H. (2015). Improving reasoning skills in secondary history education by working memory training. *British Educational Research Journal*, *41*(2), 210-228.

Ariës, R.J., Ghysels, J., Groot, W. & Maassen van den Brink, H. (2016). Combined working memory capacity and reasoning strategy training improves reasoning skills in secondary social studies education: evidence from an experimental study. *Thinking Skills and Creativity, 22*, 233-246.

Ariës, R.J.F.J., Ghysels, J., Groot, W. & Maassen van den Brink, H. (2017). Does n-back training in a school-based context improve reasoning? A literature review. TIER Working Paper Series 17/08.

Au, J., Sheehan, E., Tsai, N., Duncan, G.J., Buschkuehl, M., & Jaeggi, S.M.

(2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review, 22*(2). 366-377.

Baddeley, A.D. & Hitch, G.J. (1974). Working Memory. *The Psychology of Learning and Motivation, 8,* 47-89.

Baildon, M.C. & Sim, J.B.Y. (2009). Notions of criticality: Singaporean teachers' perspectives of critical thinking in social studies. *Cambridge Journal of Education, 39*(4), 407-422.

Beatty, E.L. & Vartanian, O. (2015). The prospects of working memory training for improving deductive reasoning. *Frontiers in Human Neuroscience, 9*(56), 1-2.

Bolger, D.J., Mackey, A.P., Wang, M. & Grigorenko, E.L. (2014). The role and sources of individual differences in critical-analytic thinking: a capsule overview. *Educational Psychology Review, 26*, 495-518.

Botvinick, M. & Braver, T. (2015). Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology, 66*, 83-113.

Brewer, B.W. (ed.) (2014). *Handbook of Sports Medicine and Science Sport Psychology*. Oxford: Blackwell Publishing.

Bull, R., Johnston, R. S. & Roy, J. A. (1999). Exploring the roles of the visual–spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, *15*(3), 421-442.

Bull, R. & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental neuropsychology 19*, 273-293.

Cabus, S.J. & Ariës, R. J. (2014). What do parents teach their children? – The effects of parental involvement on student performance in Dutch compulsory education. *Educational Review, 69*(3), 285-302.

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education*. *11*(1), 109-121.

Chein, J.M. & Schneider, W. (2012). The brain's learning and control

architecture. *Current Directions in Psychological Science, 21(2)*, 78-84.

Cheshire, A., Ball, L.J. & Lewis, C. (2005). *Self-explanation, feedback and the development of analogical reasoning skills: Microgenetic evidence for a metacognitive processing account*. Alpha: Sheridan Printing.

College voor Examens (2012). Geschiedenis HAVO - Syllabus centraal examen 2014 [History HAVO - Syllabus central exam 2014]. 33.

College voor Examens (2013). Maatschappijwetenschappen HAVO – Syllabus centraal examen 2014 [Social studies HAVO - Syllabus central exam 2014]. 69.

Colom, R., Román, F.J., Abad, F.J., Chun Shih, P., Privado, J., Froufe, M., Escorial, S., Martínez, K., Burgalata, M., Quiroga, M.A., Karama, S., Haier, R.J., Thompson, P.M. & Jaeggi, S.M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41,* 712-727.

Conway, A.R.A., Kane, M.J. & Engle, R.W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552.

Cowan, N. (1998). *Attention and memory: An integrated framework.* New York: Oxford University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in Brain Research, 169,* 323-338.

Crone, E. (2009). *Het puberende brein* [The adolescent brain]. Amsterdam: Bert Bakker.

Crone, E. & Dahl, R.E. (2012). Understanding adolescence as a period of social-affective engagement and goal flexibility. *Nature Reviews Neuroscience, 13*, 636-650.

Crone, E., Verhoeven, L., Van Aken, M., Bekkering, H., Van den Broek, P., Durston, S., Elzinga, B., Van Gog, T., De Groot, R., Jolles, J., De Jong, T., Kirschner, P., Krabbendam, L., Van der Maas, H., Van

Merriënboer, J., Van der Molen, M. & Veenstra, R. (2016). *Research agenda Brain, Cognition and Education* (Barneveld, S. (ed.)). NIHC NRO, 19.

Davidson, M. C., Amso, D., Anderson, L.C. & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*, 2037-2078.

Dahlin, K. (2010). Effects of working memory training on reading in children with special needs. *Reading and Writing, 24*, 479-491.

De Jong, T., Van Gog, T., Jenks, K., Manlove, S., Van Hell, J., Jolles, J., Van Merriënboer, J., Van Leeuwen, T. & Boschloo, A. (2009). *Explorations in learning and the brain: On the potential of cognitive neuroscience for educational science*. New York: Springer.

Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.

Ericsson, K.A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Fangmeier, T., Knauff, M., Ruff, C.C. & Sloutsky, V. (2006). fMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience, 18*(3), 320-334.

Finn, A.S., Kraft, M.A., West, M.R., Leonard, J.A., Bish, C.E., Martin, R.E., Sheridan, M.A., Gabrieli, C.F.O. & Gabrieli, J.D.E. (2014). Cognitive skills, student achievement tests, and schools. *Psychological Science, 25*(3), 736-744.

Fung, D. & Howe, C. (2012). Liberal studies in Hong Kong: A new perspective on critical thinking through group work. *Thinking Skills and Creativity*, *7*(2), 101-111.

Fung, D.C.L., & Howe, C.J. (2014). Group work and the learning of critical thinking in the Hong Kong secondary liberal studies curriculum. *Cambridge Journal of Education, 44*(2), 245-270.

Fuster, J.M. (2003). *Cortex and Mind: Unifying cognition*. New York: Oxford University Press.

Fuster, J.M. (2006). The cognit: A network model of cortical

representation. *International Journal of Psychophysiology, 60,* 125-132.

Fuster, J.M. (2009). Cortex and memory: Emergence of a new paradigm. *Journal of Cognitive Neuroscience, 21(11),* 2047-2072.

Gathercole, S.E., Pickering, S.J., Ambridge, B. & Wearing, H. 2004. The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*, 177-190.

Gazzaniga, M., Ivry, R., Mangun, G. & Steven, M. (2009). *Cognitive neuroscience: The biology of the mind*. New York: W. W. Norton & Company.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences, 11*(10), 435-441.

Goethe, J.W. (2013). *Gedichte. Ausgabe letzter hand 1827: Berliner Ausgabe.* North Charleston: CreateSpace Independent Publishing Platform. (Original work published 1827).

Gold, J. M., Berman, K. F., Randolph, C., Goldberg, T. E. & Weinberger, D. R. (1996) PET validation of a novel prefrontal task: Delayed response alteration, Neuropsychology, 10(1), 3–10.

Goldberg, E. (2010). *Het sturende brein: Onze hersenen in een complexe wereld [The new executive brain: Frontal lobes in a complex world]*. Amsterdam: Wereldbibliotheek.

Gooding, H.C., Mann, K. & Armstrong, E. (2016). Twelve tips for applying the science of learning to health professions education. *Medical Teacher, 39*(1). 26-31

Hancock, D.R. (2002). Influencing graduate students' classroom achievement, homework habits and motivation to learn with verbal praise. *Educational Research, 44*(1), 83-95.

Holmes, J. & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology, 34*(4), 440-450.

Holmes, J., Gathercole, S.E. & Dunning, D.L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*(4)*, F9-F15.*

Isaacson, W. (2015). *Einstein, de biografie*. Utrecht: Spectrum.

Jaeggi, S. M., Buschkuehl, M., Jonides, J. and Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829-6833.

Jaeggi, S.M., Buschkuehl, M., Jonides, J. & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Psychological and Cognitive Sciences, 108*(25), 10081-10086.

Jaeggi, S.M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J. & Perrig, W.J. (2010). The relationship between n-back performance and matrix reasoning – implications for training and transfer. *Intelligence, 38*, 625-635.

Jaynes, J. (2000). *The origin of consciousness in the breakdown of the bicameral mind*. New York: Houghton Mifflin Co.

Jolles, J. (2010). *Ellis en het verbreinen: over hersenen, gedrag & educatie*. Amsterdam: Neuropsych Publishers.

Jolles, J., De Groot, R., Van Benthem, J., Dekkers, H., De Glopper, C., Uijlings, H. & Wolff-Albers, A. (2006). *Brain lessons: A contribution to the international debate on brain, learning & education, based on the results of an invitational conference organized by the Netherlands Organization for Scientific Research (NWO).* Maastricht: Neuropsych Publishers.

Jurado, M.B. & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology Review, 17,* 213-233.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kane, T., Rockoff, J. & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review, 27*, 615-631.

Kant, I. (2004). *Kritiek van de zuivere rede [Critique of pure reason].* J.

Veenbaas & W. Visser (Trans.). Amsterdam: Boon (original work published 1781).

Karbach, J., Strobach, T. & Schubert, T. (2014). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology, 21*(3), 285-301.

Kimura, D. (1996). Sex, sexual orientation and sex hormones influence human cognitive function. *Current Opinion in Neurobiology, 6*(2), 259-263.

Klingberg, T. (2009). *The overflowing brain: information overload and the limits of working memory.* New York: Oxford University Press.

Klingberg, T., Forssberg, H. & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*(6), 781-791.

Kramarski, B. & Mevarach, Z.R. (2003). Cognitive-metacognitive training within a problem-solving based Logo environment. *British Journal of Educational Psychology, 67*, 425-445.

Kramarski, B. & Mevarach, Z.R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281-310.

Kroesbergen, E.H., Van ‘t Noordeinde, J.E. & Kolkman, M.E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology, 20*(1), 23-37.

Kuhn, D. (1999). A developmental model of thinking. *Educational Researcher, 28*(2), 16-26.

Kuhn, D. & Dean Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*(4), 268-273.

Küper, K. & Karbach, J. (2016). Increased training complexity reduces the effectiveness of brief working memory training: Evidence from short-term single and dual n-back training interventions. *Journal of Cognitive Psychology, 28*(2), 199-208.

Lee, K., Ng, E. & Ng, S. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *Journal of Educational Psychology, 101*(2), 373-387.

Lehman, D.R. & Nisbett, R.E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology, 26*(6), 952-960.

Leinhardt, G., Stainton, C., Virji, S. & Odoroff, E. (eds.) (1994). *Learning to reason in history: Mindlessness to mindfulness.* Hillsdale, Lawrence Erlbaum Associates, Inc.

Lilienthal, L., Tamez, E., Shelton, J.T., Myerson, J. & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20,* 135-141.

Linnenbrink, E.A. & Pintrich, P.R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313-327.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice-Hall.

Loosli, S.V., Buschkuehl, M., Perrig, W.J. & Jaeggi, S.M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18*(1), 62-78.

Luna, B., Garver, K.E., Urban, T.A., Lazar, N.A. & Sweeney, J. (2004). Maturation of cognitive processes from late childhood to adulthood. *Child Development, 75*(5), 1357-1372.

Mackey, A.P., Finn, A.S., Leonard, J.A., Jacoby-Senghor, D.S., West, M.R., Gabrieli, C.F.O. & Gabrieli, J.D.E. (2015). Neuroanatomical correlates of the income-achievement gap. *Psychological Science, 26*(6), 925-933.

Martin, A.J. & Steinbeck, K. (2016). The role of puberty in students' academic motivation and achievement. *Learning and Individual Differences, 53*, 37-46.

Marzano, R.J., Pickering, D.J. & Pollock, J.E. (2001). *Classroom instruction*

that works: Research-based strategies for increasing student achievement. Alexandria (USA): Association for Supervision & Curriculum Development (ASCD).

Maqsud, M. (1998). Effects of metacognitive instruction on mathematics achievement and attitude towards mathematics of low mathematics achievers. *Educational Research, 40*, 237-243.

Melby-Lervåg, M. & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270-291.

Melby-Lervåg, M., Redick, T.S. & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of 'far transfer': evidence from a meta-analytic review. *Perspectives on Psychological Science, 114,* 512-534.

Merriënboer, J. & Sweller, J. (2010). Cognitive load theory in medical education: design principles and strategies. *Medical Education, 44*, 85-93.

Meschberger, F.L. (1990). An interpretation of Michelangelo's Creation of Adam based on neuroanatomy. *JAMA, 264*(14), 1837-1841.

Mevarach, Z. R. & Kramarski, B. (1997). IMPROVE: A Multidimensional Method for Teaching Mathematics in Heterogeneous Classrooms. *American Educational Research Journal, 34*(2), 365-394.

Mevarach, Z. R. & Kramarski, B. (2003). The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. *British Journal of Educational Psychology, 73*, 449-471.

Minear, M., Brasher, F., Brandt Guerrero, C., Brasher, M., Moore, A. & Sukeena, J. (2016). A simultaneous examination of two forms of working memory training: Evidence for near transfer only. *Memory & Cognition, 44*, 1014-1037.

Murnane, R. J. & Willet, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research.* Oxford University Press.

Nagy Jacklin, C., & Martin, L.J. (1999). Effects of gender on behavior and

development. In: Levine, M.D., Carey, W.B., & Crocker, A.C. (eds.), *Developmental-behavioral pediatrics*, 3rd ed. Philadelphia, PA: W.B. Saunders, pp. 100-106.

Nevo, E. & Breznitz, Z. (2014). Effects of working memory and reading acceleration training on improving working memory abilities and reading skills among third graders. *Child Neuropsychology, 20*(6), 752-765.

Newmann, F.M. (1987). Higher order thinking in the teaching of Social Studies: Connections between theory and practice. In: J.F. Voss, D.N. Perkins & J.W. Segal (eds.), *Informal reasoning and education.* Hillsdale, NJ. Routledge. Pp 381-400.

OECD (2008). 21st century learning: research, innovation and policy: directions from recent OECD analyses, OECD/CERI International Conference 'Learning in the 21st century: research, innovation and policy', CERI.

Office of Educational Technology (2017). *Reimagining the role of technology in education: 2017 National education technology plan update*, U.S. Department of Education.

Olesen, P., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75-79.

Onderwijsraad (2017). De leerling centraal? [Student-centric learning?], Onderwijsraad. Paas, F., Van Merriënboer, J.J., & Van Gog, T. (2011). Designing instruction for the contemporary learning landscape. In: K. Harris, S. Graham, & T. Urdan (eds), *APA Educational Psychology Handbook: Vol. 3. Application to learning and teaching*. Washington: American Psychological Association.

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences, 9*(2), 60-68.

Perfetti, C. A., Britt, M. A., Rouet, J.-F., Georgi, M. C. & Mason, R. A. (eds). (1994). *How students use texts to learn and reason about historical uncertainty*. Hillsdale: Lawrence Erlbaum Associates, Inc.

Pintrich, P.R., Smith, D.A.F., Garcia, T. & McKeachie, W.J. (1991). *A*

Manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ), Report Number NCRIPTAL-91-B-004. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning. (ERIC Document Reproduction Service No. ED 338 122).

Polya, G. (1957). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.

Redick, T.S., Shipstead, Z., Harrison, T.L., Hicks, K.L., Fried, D.E., Hambrick, D.Z., Kane, M.J. &Engle, R.W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359-379.

Redick, T.S., Shipstead, Z., Wiemers, E.A., Melby-Lervåg & Hulme, C. (2015). What's working in working memory training? An educational perspective. *Educational Psychological Review, 27*(4)*,* 617-633.

Rotherham, A.J. & Willingham, D. (2009). 21st century skills: the challenges ahead. *Educational Leadership,* 16-21.

Salminen, T., Mártensson, J., Schubert, T. & Kühn, S. (2016). Increased integrity of white matter pathways after dual *n*-back training. *NeuroImage, 133,* 244-250.

Sander, M.C., Lindenberger, U. & Werkle-Bergner, M. (2012). Lifespan age differences in working memory: A two-component framework. *Neuroscience and Biobehavioral Reviews, 36*, 2007-2033.

Schoolinfo (n.d.). Waarom gepersonaliseerd leren [Why personalized learning]. Retrieved March 17, 2017, from http://leerling2020.nl.

Schopenhauer, A. (1988). *Manuscript Remains: Early Manuscripts (1804-1818).* E.F.J. Payne (Trans.). New York: Bloomsbury Academic.

Schug, M.C. & Birkey, J.C. (1983). The development of children's economic reasoning. Joint Council on Economic Education, New York. 22.

Schunk, D.H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal

setting, and self-evaluation. *Reading & Writing Quarterly, 19*(2), 159-172.

Shipstead, Z., Redick, T. S. & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628-654.

Shulman, E. P., Smith, A. R., Silva, K., Icenogle, G., Duell, N., Chein, J. & Steinberg, L. (2015). The dual systems model: Review, reappraisal, and reaffirmation. *Developmental Cognitive Neuroscience, 17,* 103-117.

SLO National Centre of Expertise for Curriculum Development (2007). Examenprogramma geschiedenis HAVO/VWO [Exam program history HAVO/VWO], *1-6*.

SLO National Centre of Expertise for Curriculum Development (2011). Handreiking schoolexamen maatschappijwetenschappen HAVO/VWO [School exam assistance document social studies HAVO/VWO], *1-54*.

Sluijsmans, L. (2013). Evaluatie van het vernieuwde examenprogramma maatschappijwetenschappen voor HAVO [Evaluation of the renewed exam program social studies for HAVO], *SLO, 65*.

Smith, E.E., Geva, A., Jonides, J., Miller, A., Reuter-Lorenz, P. & Koeppe, R.A. (2001). The neural basis of task-switching in working memory: Effects of performance and aging. *Proceedings of the National Academy of Sciences, 98*(4), 2095-2100.

Söderqvist, S. & Bergman Nutley, S. (2015). Working memory training is associated with Long-term attainments in math and reading. *Frontiers in Psychology*, *6*(1711).

Spinoza, B. (2002). *Ethica [Ethics].* H. Krop (Trans.). Amsterdam: Bert Bakker. (original work published 1677).

St Clair-Thompson, H.L. & Gathercole, S.E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59*(4), 745-759.

St Clair-Thompson, H., Stevens, R., Hunt, A. & Bolder, E. (2010). Improving

children's Working memory and classroom performance. *Educational Psychology, 30*(2)*,* 203-219.

Stanovich, K.E. & West, R.F. (2000). Individual differences in reasoning: implications for the rationality debate. *Behavioral Brain Sciences, 23*, 645-726.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences, 9*(2), 69-74.

Stephenson, C. L. & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41*, 341–357.

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

Swanson, H. L., Cooney, J. B. & Brock, S. (1993). The influence of working memory and classification ability on children's word problem solution. *Journal of Experimental Child Psychology, 55*, 374-395.

Swanson, L. & McMurran, M. (2017). The impact of working memory training on near and far transfer measures: Is it all about fluid intelligence? *Child Neuropsychology*. 1-26.

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

Thompson, T.W., Waskom, M.L., Garel, K.-L. A., Cardenas-Iniguez, C., Reynolds, G.O., Winter, R., Chang, P., Pollard, K., Lala, N., Alvarez, G.A. & Gabrieli, J.D.E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE, 8*(5). e63614.

Thornton, S.J. & Vukelich, R. (1988). Effects of children's understanding of time concepts on historical understanding. *Theory and Research in Social Education, 1*, 69-82.

Titz, C. & Karbach, J. (2014). Working memory and executive functions: effects of training on academic achievement. *Psychological Research, 78*, 852-868.

Torney-Purta, J. (ed.). (1994). *Dimensions of adolescents' reasoning about*

political and historical issues: ontological switches, developmental processes, and situated learning. Hillsdale: Lawrence Erlbaum Associates, Inc.

Tuan, H.-L., Chin, C.-C. & Shieh, S.-H. (2005). The development of a questionnaire to measure students' motivation towards science learning. *International Journal of Science Education, 27*(6), 639-654.

Uguroglu, M.E., Schiller, D.P. & Walberg, H.J. (1981). A multidimensional motivational instrument. *Psychology in the Schools, 18,* 279–285.

Van der Sluis, S., De Jong, P. F. & Van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading and arithmetic*, Intelligence, 35*(5), 427–449.

Van Drie, J. & Van Boxtel, C. (2008). Historical reasoning: towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20*(2), 87-110.

Van Merriënboer, J.J. & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-176.

Van Merriënboer, J.J. & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education, 44*, 85-93.

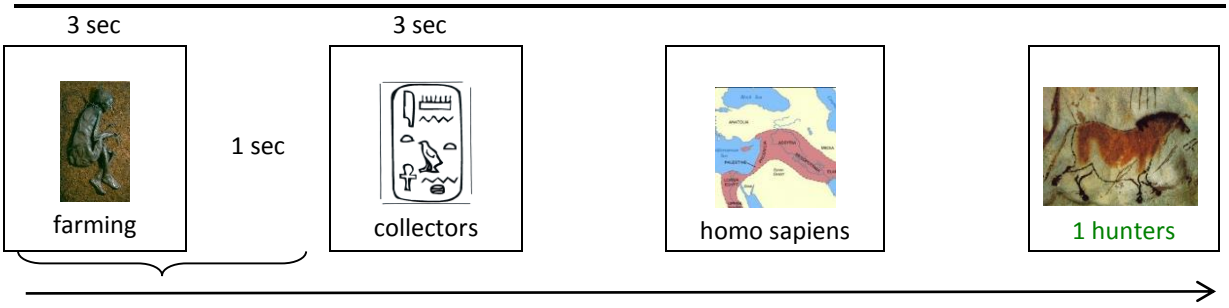
Von Bastian, C.C., & Oberauer, K. (2013). Effects and mechanisms of working memory training: a review. *Psychological Research, 78,* 803-820.

Vos, H. (2001). *Metacognition in higher education*, University of Twente. PhD dissertation: 154 pp.

Voss, J.F., Perkins, D.N. & Segal, J.W. (1991). *Informal reasoning and education*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Appendix A

Example of a dual 2-back task

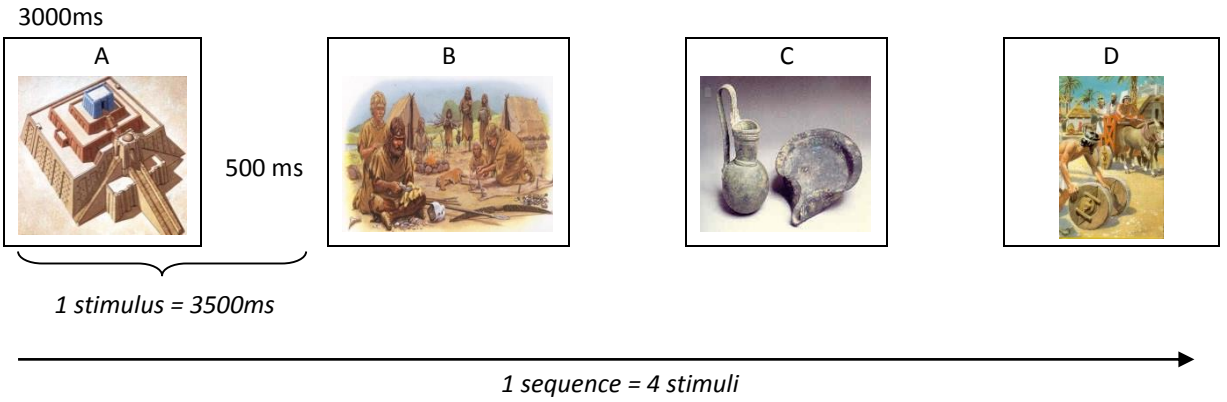


1 sequence = 20 stimuli

Note. The noun ‘hunters’ is colored green which, in the case of a dual 2-back task, means that the noun ‘collectors’ and the image of hieroglyphs should be written down on the answer form.

Appendix B

Example of a visual Odd One Out task



Note. Image 'B' is the Odd One Out. The image depicts a hunter-gatherer civilization. Images 'A', 'C' and 'D' all depict examples of agrarian civilizations.

Appendix C

Example of a school test reasoning question and criteria of analysis

'This painting demonstrates the equality between Pope and the Emperor.'

Question: Is this hypothesis true or false? Present your answer with two arguments based on the painting.

Three criteria for meta-cognitive reasoning:

1. Contextualization: Students should note that the painting is a symbolic representation of the investiture controversy: a series of popes challenged European monarchies over the control of appointments of church officials. (1 point)
2. Use of sources: The controversy, that raised the question of whether the Pope or the Emperor had absolute power, is depicted in this painting as the Donation of Constantine. Roman Emperor Constantine donated lands to the Holy Pope of Rome. (1 point)
3. The painting symbolizes the absolute power of the Pope. Symbols are, for example: the Pope is depicted higher than the Emperor; the Pope is sitting while the Emperor is kneeling. (1 point)

Appendix D

The IMPROVE-questionnaire

IMPROVE-questionnaire (Mevarach & Kramarski, 1997, 2003):

1. What is the problem/task about? (comprehension question)
   'reflect on the problem before solving it. In addressing a comprehension question, students had to read the problem aloud, describe the concepts in their own words, and discuss what the concepts meant or into which category the problem could be classified' (1997, p. 374).

2. Which strategy/principle is appropriate for solving/addressing the problem/task? (Strategic question)
   'which strategies are appropriate for solving the given problem and for what reasons. In addressing the strategic questions, students had to describe the "what"(e.g., "What strategy/tactic/principle can be used in order to solve the problem?"),the "why"(e.g. "Why is this strategy/tactic/principle most appropriate for solving the problem?"), and the "how" (e.g., "How can the suggested plan be carried out?")' (1997, p. 375).

3. How is this problem/task different from what you have already solved? (Connection question)
   'focus on the similarities and differences between the problem they are working on and the previous problem or set of problems that they have already solved. In doing so, students gradually learn to distinguish between the surface and the deep […] structures of a problem' (1997, p.375).

4. What were the difficulties you felt during the solution process? (Reflection question)

Appendix E

The modified SMTSL questionnaire

Directions for students

This questionnaire contains statements about your willingness to participate in this science class. You will be asked to express your level of agreement with each statement. There are no "right " or "wrong" answers. Your opinion is what is wanted. Think about how well each statement describes your willingness to participate in this class.

Draw a circle around

1. if you strongly disagree with the statement
2. if you disagree with the statement
3. if you have no opinion about the statement
4. if you agree with the statement
5. if you strongly agree with the statement

Be sure to give an answer for all the questions. If you change your mind about an answer, just cross it out and circle another.

Some statements in this questionnaire are fairly similar to other statements. Don't worry about this. Simply give your opinion about all the statements.

A. Self-efficacy

1. Whether the science content is difficult or easy, I am sure that I can understand it.
2. I am not confident about understanding difficult science concepts.
3. I am sure that I can do well on science tests.
4. No matter how much effort I put in, I cannot learn science.
5. When science activities are too difficult, I give up or only do the easy parts.
6. During science activities, I prefer to ask other people for the answer rather than think for myself.
7. When I find the science content difficult, I do not try to learn it.

B. Active learning strategies

8. When learning new science concepts, I attempt to understand them.
9. When learning new science concepts, I connect them to my previous experience.
10. When I do not understand a science concept, I find relevant resources that will help me.

11. When I do not understand a science concept, I would discuss it with the teacher or other students to clarify my understanding.
12. During the learning processes, I attempt to make connections between the concepts that I learn.
13. When I make a mistake, I try to find out why.
14. When I encounter science concepts that I do not understand, I still try to learn them.
15. When new science concepts that I have learned conflict with my previous understanding, I try to understand why.

C. Science Learning Value
16. I think that learning science is important because I can use it in my daily life.
17. I think that learning science is important because it stimulates my thinking.
18. In science, I think that it is important to learn to solve problems.
19. In science, I think it is important to participate in inquiry activities.
20. It is important to have the opportunity to satisfy my own curiosity when learning science.

D. Performance Goal
21. I participate in science courses to get a good grade.
22. I participate in science courses to perform better than other students.
23. I participate in science courses so that other students think that I'm smart.
24. I participate in science courses so that the teacher pays attention to me.

E. Achievement Goal
25. During a science course, I feel most fulfilled when I attain a good score in a test.
26. I feel most fulfilled when I feel confident about the content in a science course.
27. During a science course, I feel most fulfilled when I am able to solve a difficult problem.
28. During a science course, I feel most fulfilled when the teacher accepts my ideas.
29. During a science course, I feel most fulfilled when other students accept my ideas.

F. Learning Environment Stimulation
30. I only follow the history course because it is part of the curriculum
31. History is an extracurricular course for me
32. I am willing to participate in this science course because the content is exciting and raries.
33. I am willing to participate in this science course because the teacher uses a variety of teaching methods.
34. I am willing to participate in this science course because the teacher does not put a lot of pressure on me.
35. I am willing to participate in this science course because the teacher pays attention to me.

36. I am willing to participate in this science course because it is challenging.
37. I am willing to participate in this science course because the students are involved in discussions.

Summary

Teachers regularly observe that adolescent students have insufficient reasoning skills to properly answer reasoning test questions. The importance of acquiring reasoning skills in secondary education becomes clear in humanities school-tests which increasingly contain reasoning questions. Policymakers and schools are searching for new didactical approaches to improve adolescents' reasoning skills. One potentially effective strategy is developing learning strategies based on cognitive and brain science, popularized as 'brain-based learning' (Crone et al., 2016). Standardized training batteries have been developed that are theorized to improve reasoning performance while they focus on areas of the brain that deal with reasoning, such as working memory (WM). But these batteries are not yet being implemented in school curricula, and school-based WM training to improve reasoning is still limited and recent. Evidence of improved reasoning in education, caused by WM training, therefore remains poor. This thesis contributes to our knowledge and insight into effective brain-based strategies that improve reasoning skills in secondary education humanities curricula. We evaluate the effectiveness of a combined WM-capacity- and reasoning-strategy training which enables students to acquire reasoning skills to improve their academic achievement. This could provide the necessary tools for both teachers and students to improve school-test results in humanities courses.

Humanities courses in secondary education in the Netherlands include history and social studies, among others. Tests usually consist of deductive reasoning problems. The required problem-solving skills are generally comparable for all humanities courses. They consist of general problem-solving principles, such as analysis, hypothesis, generation, and interpretation, and of domain-specific characteristics that are specific to the particular courses, such as domain-specific knowledge and the meta-concepts of the domain. Each teacher teaches the required knowledge and skills to the students, primarily by using direct instruction, course

book questions and (un)official school tests in order to train problem-solving skills. In this regard, the student should be taught to reorganize and manipulate retained information, which are seen as skills that are highly important to solve reasoning questions. Developing a problem-solving strategy and training in cognitive processing capacity can be primarily accredited to WM functions (Fuster, 2003). WM underlies several cognitive abilities, including logical reasoning and problem-solving. But 'the cortical hubs of the network […] are slow to mature' and sensitive to genetic and environmental factors, which could explain individual differences in reasoning test achievements.

We hypothesize that WM training is only effective in educational settings when accompanied by reasoning strategy training that is subject-based in the humanities. In Chapters 2 and 3 we present review studies that examine the possible effects of a single WM-capacity training and a combined WM-capacity- and reasoning-strategy training. These two chapters set out to address the following questions:

1) Does n-back training improve school-based reasoning achievements?;
2) Can WM capacity + metacognitive training serve as an effective instrument to enhance school-based reasoning achievements?

Chapters 4 and 5 focus on the effect of the combined WM-capacity- and reasoning-strategy training in history and in humanities education. In these chapters, we answer the following questions:

3) Does WM capacity + metacognition training, based on subject-matter knowledge in history education, have an effect on students' achievements in tests for which reasoning abilities are required?;
4) Does WM capacity and/or metacognitive training, when independently trained, have an effect on students' achievements in history tests for which reasoning abilities are required?;

5) Does the combined training have an effect on students' reasoning achievements in school-tests in other humanities courses (i.e. social science)?

Chapter 6 includes the combined intervention in which motivational factors are included to investigate the added value of motivation during the training period. The following question is addressed:

6) Does the combined training affect reasoning achievements in school-test results when motivational factors are added to the training?

Chapter 2 of this dissertation reviews the effects of one specific WM training task, the n-back, on deductive reasoning improvements. The n-back task involves the serial presentation of a stimulus, spaced several seconds apart. The participant must decide whether the current stimulus matches the one displayed *n* trials ago, where *n* is a variable number that can be adjusted up or down to, respectively, increase or decrease cognitive load. In the context of WM training, efforts have focused on flexibly adapting the task difficulty in accordance with the participants' fluctuating performance level by increasing and decreasing the level of *n*. The idea is to keep the participant's WM system perpetually engaged at its limit, thereby stimulating an increase in WM function.

Specifically n-back training, which is a commonly used task in WM task-transfer training studies, could lead to positive effects on general fluid intelligence (*Gf)* (Au et al., 2015; Jaeggi et al., 2008; Beatty and Vartanian, 2015). Ergo, n-back training could theoretically improve deductive reasoning via its effect on *Gf.* N-back requires storage and updating of information in WM and 'interference resolution'. Therefore, more than other WM-training tasks, n-back training could positively affect fluid intelligence (Au et al., 2015). N-back training might stimulate the WM function and, as a result, improve reasoning. In this chapter, we review the literature on the effects of the n-back on deductive reasoning improvements. To enlarge on far-transfer results to reasoning, we

included studies on school-based and non-school-based interventions in our review.

The main characteristic is that interventions must: include some form of n-back and have far-transfer measures to (school-based) reasoning; use a control group; and include a deductive reasoning outcome measure. The literature review shows that all interventions have used a pretest-intervention-posttest design, in which objective academic measurements are used. All but one intervention (Jaeggi et al., 2011) used randomized experimental and control groups. The average sample size was 67 participants. All interventions used identical pretest and posttests to measure for far transfer effects. We have found five significant and eight non-significant results concerning far-transfer effects to reasoning (*Mdn d* = 0.10*, range* -0.35 – 2.20*).*

Four interventions were school-based. This means that the training was implemented in the school curricula. Two studies have found significant results and two studies did not (*Mdn d* = 1.04*, range* 0.66 – 2.20). In conclusion, no far transfer occurs when solely n-back is trained. When n-back training is combined with other training tasks, then the results appear to be ambiguous with respect to far-transfer to reasoning. Eight non-school-based interventions were included in the review. Five out of eight interventions observed no far-transfer to reasoning posttests (*Mdn d*= 0.16*, range* -0.32 – 0.80). Furthermore, significant improvements in posttest results did not occur when compared with the results of an active control group. Many studies lack a critical analysis of the mechanisms that cause reasoning achievements through WM or n-back training, such as skill internalization, other than WM (see also: Lilienthal et al., 2013, p.135). In school-based settings, when WM training is combined with reasoning strategy training, significant improvements do occur in reasoning-test results. One might question the role of n-back in this transfer to reasoning, and whether reasoning strategy training is primarily responsible for transfer results. In this regard, it might be the case that WM (improvements) functions as a subsystem that supports improvements in reasoning through reasoning strategy training. Five of

seven studies in the young adult group used samples of university students. One might consider that this is a rather homogeneous group who already had above average WM levels, a situation which has been theorized to limit WM training benefits. The possible benefits of a combined WM- and reasoning-strategy training to reasoning test-achievements in primary and secondary education is addressed in Chapter 3.

Chapter 3 reviews whether a combined training of WM capacity and reasoning strategy could benefit school-based reasoning skills, while two aspects of WM are theorized to affect the process of problem-solving. (1) WM temporarily stores and manipulates information during complex cognitive activities, such as reasoning. Holmes, Gathercole and Dunning (2009) define WM as the cognitive system that provides temporary storage and manipulation of information in the course of complex cognitive activities. According to Olesen, Westerberg and Klingberg (2004), the amount of information that a person can retain is important for problem-solving. (2) Prescriptive knowledge, the generic memories of effective ways to solve higher-level cognitive problems, is stored in WM (Gazzaniga et al., 2009; Klingberg, 2009; Goldberg, 2010). Experts, unlike novices, possess generic memories-developed schemas or patterns to solve problems (Sweller, 1988). This 'pattern recognition' is based on learning processes which underlie repeated exposure to similar contexts that originate in matched response strategies. According to Goldberg (2010), a transition occurs from the absence of effective behavior of WM to the formation of effective behavior in reasoning processes to address reasoning questions more effectively and efficiently when meta-cognitive components are added to training strategies. Meta-cognitive training causes WM to form blueprints and to plan for rational analysis and analytical methods (Gold et al., 1996; Gazzaniga et al., 2009; Goldberg, 2010).

We have found a total of 16 articles (10 articles concerning WM-capacity training, and 6 articles concerning metacognitive training)

eligible for inclusion in the systematic review. Many different WM-capacity training tasks were used in the studies with training periods that highly differed in duration. Almost all interventions were conducted in (non-)compulsory education, ranging from preschool to middle school. Furthermore, in almost all interventions, the participants were tested for verbal or mathematical reasoning. 50% of the interventions resulted in significant improvements in reasoning achievement, regardless of the type and intensity of training or age. Significant improvements occurred more often in reading compared with math tests. The transfer to reasoning achievement is likely to be more effective when participants are trained by using content-based interventions due to near-transfer effects. Near-transfer effects occur when the content of the training batteries is similar, but not identical, to test-content. Two studies on mathematics used content-based training batteries in which near-transfer effects could occur. Alloway et al. (2012) found positive results in mathematical reasoning, but had a small sample size. In the Kroesbergen et al. (2014) study, content based training resulted in significant improvements in early numeracy. The results that we found in studies that contained general training batteries proved to be more ambiguous. With regard to mathematics, far-transfer effects between general training and tests are less likely to occur. This could be attributed to the specifics of reasoning in mathematics, in which students reason at an abstract level, which can not be compared with reading comprehension or reasoning in the humanities, in which students mainly deal with concrete topics.

As for the metacognitive training, four different training tasks were used in the training studies within training periods that greatly differ in duration. Almost all interventions were conducted in secondary education. Two instructional designs on meta-cognition were reviewed: the method in which small heterogeneous groups work together in a meta-cognitive process; and the method in which meta-cognition is trained individually. Heterogeneous groups that are trained in meta-cognition turn out to achieve better compared with individual training

methods on reasoning. But this significant difference cannot be applied to the transfer of reasoning between different school courses without more research on this specific topic. The review further suggests that five issues need to be addressed in meta-cognitive training methods in order for problem-solving structures to be internalized in long-term memory: (1) Participants need to solve problems by using a training method which includes both self-reflection and reflection proposed by group members. (2) As stated in the Introduction (Chapter 2), evidence suggests that problem-solving structures need to be trained by using content-based knowledge on reasoning questions. (3) Problem-solving skills can best be internalized by training in heterogeneous groups, containing underachievers, overachievers, and average achievers in problem-solving skills. The group is responsible for reflecting on a participants' problem-solving process. (4) Skills are best internalized when a participant follows a step-by-step plan, and verbally presents the reasoning process to the other group members. (5) Training needs to be consistently repeated.

We conclude that both instructional designs could improve reasoning skills. In line with Leinhardt et al. (1994), we suggest that content-based knowledge may be crucial to acquire problem-solving skills because of the alternative views, detailed factual knowledge, broader frames of reference, and other concepts that are frequently used in compulsory education courses. Most of the research on this matter has been conducted on children and adults. This means that no firm conclusions can be drawn about the training of cognitive reasoning abilities for adolescents. Practical restrictions to implement general WM training in school curricula make it difficult to integrate general and content-based cognitive training methods in school practice. The question how to implement the combined training and its effects on improving reasoning-test achievements is addressed in Chapters 4, 5 and 6.

Chapter 4 evaluates two content-based interventions in the history courses of higher general secondary education (HAVO) and pre-university

secondary education (VWO). HAVO students (total N=92) received the combined WM-capacity- and reasoning-strategy training battery in order to analyze its effectiveness in their reasoning achievements in school-tests. WM capacity included n-back and Odd One Out tasks, while reasoning strategy training is performed by using the IMPROVE-method. The IMPROVE-method is a format to develop and structure the reasoning process (Mevarech and Kramarski, 2003). The results are compared with a control group who received a conservative training. WM capacity, reasoning strategies, and reasoning achievements in the school tests of the experimental group improve significantly after 3 weeks of training. The gain in reasoning abilities is also durable and remains significant up to 16 weeks after the training program is completed.

In VWO (total N=63), WM capacity and reasoning strategies are independently trained and compared with control group results to analyze the improvements of both training groups independently. Training of WM capacity did not improve achievements in reasoning tests significantly, but training of reasoning strategies did. The results of both experiments cannot be compared because of cognitive group differences between Experiments 1 and 2 (in terms of educational tracks). These results demonstrate that brain based learning strategies to improve reasoning abilities can easily be integrated in history curricula. Furthermore, Chapter 4 shows that a 6-week training period can significantly improve reasoning abilities of students in secondary education.

Chapter 5 includes the combined intervention in HAVO in social studies courses. The main aim of this intervention is to analyze whether the training duplicates the results of the intervention described in Chapter 4. This is an important question which can give insight into the extent to which the training is applicable in all the arts courses, as well as possible transfer effects which can be investigated in the future, as both history and social studies are courses taught in the humanities curricula.

In three secondary schools in the Netherlands we conducted an experiment that is based on WM training tasks. The aim of the training intervention was to: (1) investigate the effect of the complete training on achievement in reasoning tests; (2) investigate its durability after completion of the training; and (3) investigate the effects of age, gender, pretest reasoning-achievement scores, and WM pretest scores on posttest achievements. The experimental groups' test results (N=39) were compared with the data of the control groups (N=42) that were trained by using 'low-dose' WM tasks, such as evaluating reasoning questions using whole-class teaching methods.

WM-capacity and reasoning achievements improved significantly in the experimental groups after 4 training weeks compared with the control groups. The gain in reasoning abilities was demonstrated in both experimental subgroups and remained significant 8 weeks after the end of training, while the control groups' results did not improve. The experimental groups' improvements are in agreement with conclusions of Chapter 4, in which similar groups underwent the combined training in history curricula. The results of the regression model indicate that, for every unit increase of pretest achievement scores, the score of the reasoning achievements posttest can be predicted to increase by .29 units. Students with a low pretest reasoning-achievement score typically improve less in the posttest compared with high pretest score students. This has important implications for the effectiveness of the training for underachieving students. While it is possible that the training tasks could be going too fast for these students, the absence of an improvement in test scores could also be caused by behavioral characteristics, such as lack of motivation or concentration. With regard to the latter, the training method requires that students focus their attention on an internal representation. Although motivation and concentration were not analyzed, minimizing distractions may be a requirement for an effective intervention. Furthermore, enhancing and training attention to improve deductive reasoning could also be regarded as an important outcome measure of the training, while it is also regarded as a general cognitive

function within which WM is subsumed. To investigate the role of motivation in the training, we added motivational components in a second experiment in history courses in Chapter 6.

Chapter 6 contains an analysis the combined intervention to which motivational factors are added in order to investigate the results of the less-motivated students. In this experiment, we compared the school test results of both groups who had received the combined training, where one group was actively being motivated by the teacher by use of verbal praise, goal-setting, and feedback, which are known to improve attention. Both the HAVO experimental (N=50) and the control (N=58) groups received the training in history courses. The main objective was to analyze whether the addition of motivational components can improve test results even more compared with the control groups.

Group motivation levels are measured by comparing each group's pretest and posttest scores regarding a motivation questionnaire. We observed a significant decline of motivation levels in the control group. But neither were the motivation levels in the experimental group significantly affected. These results indicate that training on WM and reasoning strategies might require huge cognitive loads that can easily affect motivation levels. The results indicate that the motivation incentives that were used during the training in the experimental group were shown to be effective in maintaining motivation. However, the higher levels of motivation do not account for significant improved reasoning test scores. Our conclusion is that motivation levels do not predict effects of WM- and reasoning-strategy training. This allows teachers to concentrate on the learning process of the students rather than investing time in keeping the students motivated.

In Chapter 7, the findings of the review and the empirical studies lead to the formulation of four key conclusions.

(1) There is no convincing evidence that (content-based) WM capacity training alone improves adolescents' reasoning skills in education;
(2) Combined content-based WM-capacity- and reasoning-strategy training appears to be an effective instrument to enhance school-based reasoning achievement;
(3) Reasoning achievements in school-test results do not seem to be affected by motivation levels when students use the combined training method;
(4) The combined WM-capacity- and reasoning-strategy training does not appear to discriminate between low or high achieving students in terms of improvements on reasoning tests.

A number of methodological issues and policy implications arise from our findings. With regard to the methodological issues, we propose that non-school-based WM training can not be implemented one-on-one in school curricula in order to train reasoning skills, because of the absence of other aspects, such as knowledge of contents and course-specific skill acquisition, which are regarded as being important contributors to school-based reasoning-test results. We argue that a general non-school-based WM training does not provide the necessary transfer effects to improve the children's reasoning skills to enable them to improve significantly in school-based reasoning tests. Consequently, standard WM training tasks should be tailored to suit the demands of school-based reasoning. Another concern addressed in this PhD-thesis relates to the development of an effective training method to improve reasoning skills and reasoning achievements for students of varying backgrounds and baseline levels. Our findings do not imply that specific groups of students (based on low- or high-achieving baseline levels, gender, socioeconomic status, general cognitive ability or motivation levels) particularly benefit from the training compared to other groups of students. We conclude that a one-size-fits-all training is not likely to include all students in the possible benefits of the training.

With regard to the policy implications, we conclude that our training method is a much more effective instrument for students to acquire the necessary reasoning skills needed to adequately respond to a reasoning question (*d = 1.04-2.20)*, compared with frontal instruction designs. We propose that the explanation for this difference in outcomes could be that, besides improvements in WM levels, the current training engages all students actively in solving reasoning questions and in WM-capacity training. Reasoning skills are trained more effectively by the use of active didactic approaches. We also argue that the training in general could be implemented in high school curricula to effectively improve the students' reasoning skills in the humanities. However, the training does not provide a clear-cut instrument to target specific groups who need to be trained by having personal learning designs. We also want to emphasize the importance of integrating a standard form of reasoning training in secondary education to show students that the same reasoning structures can be applied in different courses. By creating awareness in this matter, schools can offer more efficient and effective learning strategies to students, and invest in interdisciplinary education concerning the acquisition of reasoning skills. In this case, the findings imply that the training, when focused on course-based content, can be implemented in all courses in the humanities, such as history and social studies, to improve reasoning test achievements.

Curriculum Vitae

Roel Ariës was born in Maastricht, The Netherlands, in 1979. He received a B.Ed. in history education at Fontys Applied University in 2003 and started his teaching activities in secondary education in the same year. While he was teaching, he obtained an M.Ed. in history education at Fontys Applied University in Tilburg in 2007 and an M.Sc. in Evidence Based Innovation in Teaching at Maastricht University's Top Institute for evidence-based Education Research (TIER) in 2011.

After graduation, Roel worked as a university teacher and coordinator of the Minor in Education program, which was a program that allowed Maastricht University students to obtain a teaching degree. His responsibilities in this post included, among others, curriculum development, quality assessment, hrm, and lecturing didactics and pedagogics in both frontal instruction and problem-based learning settings. He was a member of the Founding Committee of the Master's program in Education of Maastricht University, and was responsible for the development of the curriculum. He also obtained a teaching position (university teacher) in the Master of Evidence Based Innovation in Teaching program at TIER. He was the Quality Research Officer for the MEBIT module program that enabled teachers of the Foundation of Secondary Education Limburg (Stichting LVO) to study in MEBIT's modular program. At the same time, he retained a teaching position in history, political sciences, and philosophy in secondary education.

In 2014, Roel successfully applied for a PhD-scholarship, Promotiebeurs voor leraren [Doctoral Grant for Teachers], with the Netherlands Organisation for Scientific Research (NWO). Since then, he has been working on his PhD on a part-time basis, while also holding a teaching position in secondary education. During his PhD-research, Roel worked on practice-based working-memory training in secondary education. In order to extend his knowledge of working-memory training, he made two research visits to MIT's Gabrieli Lab for Cognitive and Affective Neuroscience in Boston, Massachusetts. He also researched

parent involvement strategies to improve test scores (not included in this dissertation). Roel has published in national and international peer-reviewed economics, psychology, and educational sciences journals. He has presented several of his projects at education and neuroscience & education conferences in, amongst other places, Europe, the United States and Canada. Currently, Roel is the Senior Educational Advisor of Zuyd Applied University's Faculty of International Business and Communication. He continues to improve his expertise in education through an active research agenda.

List of publications in peer-reviewed journals

On working memory training (included in this dissertation):

Ariës, R.J., Groot, W., Maassen van den Brink, H. (2015). Improving reasoning skills in secondary history education by working memory training. *British Educational Research Journal, 41*(2): 210-228.

Ariës, R.J., Ghysels, J., Groot, W., Maassen van den Brink, H. (2016). Combined working memory capacity and reasoning strategy training improves reasoning skills in secondary social studies education: evidence from an experimental study. *Thinking Skills and Creativity, 22*: 233-246.

On parental involvement in education (not included in this dissertation):

Ariës, R.J. and Cabus, S.J. (2015). Parental homework improves test scores: a review of the literature. *Review of Education, 3*(2): 179-199.

Cabus, S.J. and Ariës, R.J. (2015). Ouderbetrokkenheid beïnvloedt leerprestaties [Parent involvement improves academic achievements]. *Economisch-Statistische Berichten, 100*(4709): 266-268.

Cabus, S.J. and Ariës, R.J. (2016). What do parents teach their children? On birth order and performance differences in Dutch compulsory education. Educational Review, 69(3): 285-302.

Ariës, R.J. and Cabus, S.J. (2017). Verbeteren toetsscores wanneer ouders betrokken zijn bij het huiswerk van hun kind? Een verslag van een literatuuronderzoek [Do test scores improve when parents are involved in the homework process? A review of the literature]. Tijdschrift voor Onderwijsrecht en Onderwijsbeleid, 4.

Acknowledgements

I would like to begin by expressing my sincere gratitude to my supervisors, Professor Wim Groot (10+6=16) and Professor Henriëtte Maassen van den Brink ('however') for their continuous support during my MEBIT study and my PhD-research over the previous nine years. Thank you for believing in me from the beginning at MEBIT, and for giving me the opportunity to grow. In addition to my supervisors, I would like to thank my daily supervisor, Dr. Joris Ghysels. You have had a profound impact on the quality of the articles, as well as on my academic development, because of your patience and expertise.

I would also like to thank two scientists who gave extensive advice and feedback on the neuroscientific parts of the articles. Dr. Allyson Mackey, thank you so much for your advice on working memory theory and for inviting me to visit the Gabrieli Lab at MIT in Boston. Those research visits really were a once-in-a-lifetime experience. Thank you Professor Joaquín Fuster for generously taking the time to provide the working papers with the necessary theoretical improvements and discuss any questions I had concerning WM function.

My sincere appreciation also goes to the teachers who took the time to participate in the experiments, Hub Meessen, Lilly van Mierlo and Pieter Terpstra from Graaf Huyn College, Jan de Heer from Bernardinuscollege, Roel Vossen and Manon Walthie from Porta Mosana College, and Nico Zijlstra and Ria van Beusekom from SG Sintermeerten. Thanks also to all of the students who participated voluntarily or involuntarily, consciously or unconsciously, in the experiments. The future is yours!

Furthermore, I would also like to acknowledge with appreciation the impact of all current and former colleagues at TIER, teachers, researchers, and students - especially Chris Kranendonk, my fellow Educational Minor Program-coordinator: 'Ad majorem fati gloriam'!

To my father, Jef, and mother, Annie, I just want to say that words can't even express how grateful I am to have grown up in such a loving and caring family. It can't be said enough. Thank you for always being there

for me. Thank you for all the opportunities you gave me. And thank you, Ingrid, my sister, for always being there for me in good and bad times. I'm so proud that you're my sister!

And finally, I want to thank the people who deserve it the most. Thank you Kelly, my beautiful wife, for everything you have given me and life has given us. Thanks for your unwavering support. Thank you for lifting me up. Thank you for your love. And to my two kids, Pip and Lou, for lighting up the sky. You are the joy of my life.

Will it, and set to work briskly! – Friedrich Schiller.

Top Institute for Evidence Based Education Research (TIER)

The Top Institute for Evidence Based Education Research (TIER) is an inter-university institute that conducts research to develop evidence-based education. The Institute has three partners: the University of Amsterdam, Maastricht University, and the University of Groningen, and is located in Amsterdam, Maastricht and Groningen.

TIER contributes to the improvement of the quality of education in the Netherlands by promoting an evidence-based approach as a guiding principle in education policy and practice. It accomplishes this by developing (cost-) effective education interventions that are grounded in sound scientific research. TIER research is funded by the Ministry of Education, Culture and Science and the participating universities through the NWO, and complies with the quality standards and evaluation procedures used by NWO.

The following books recently appeared in the TIER Research Series:

I. C. Haelermans (2012), *On the productivity and efficiency of education. The role of innovations in Dutch secondary education*

II. L. van Welie (2013), *They Will Get There! Studies on Educational Performance of Immigrant Youth in the Netherlands*

III. S. Cabus (2013), *An Economic Perspective on School Dropout Prevention using Microeconometric Techniques*

IV. I. Cornelisz (2013), *School Choice, Competition and Achievement: Dutch Compulsory Education*

V. M. Heers (2014), *The Effectiveness of Community Schools: Evidence from the Netherlands*

VI. N. Ruijs (2014), *Empirical Studies in the Economics of Education*

VII. D. Bartelet (2015), *Basic Mathematics Skills Development in Children: An Evaluative Study on Cognitive and Educational Factors*

VIII. I.Rud (2015), *The Relationship between Youth Crime and Education*

IX J. Coenen (2016), *Teacher Characteristics and Student Performance*

X D. Hidalgo Saá (2017), *Essays in Economics of Education and Training*

XI E. Sneyers (2017), *What works to improve student success: The effects of academic dismissal policies, student grants and institutional mergers on student outcomes*

XII R. Ariës (2018), *An evidence-based approach to working-memory-based training in secondary education to improve reasoning test achievements*