

Finnish primary school pupils' performance  
in learning to learn assessments:  
A longitudinal perspective on educational equity

*Dedicated to my grandfather  
Veikko Lamberg (11.4.1921 – 19.5.2012)  
who would have appreciated this doctoral thesis  
more than anybody else in the world*



Mari-Pauliina Vainikainen

Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity

*Academic Dissertation to be publicly discussed by due permission of the Faculty of Behavioural Sciences at the University of Helsinki, Siltavuorenpenger 3A (Athena-building), Lecture hall 107, on Friday, 14<sup>th</sup> November 2014, at 12 o'clock.*

Supervisors: *Professor Emeritus Jarkko Hautamäki*  
*(special education, educational assessment)*  
*University of Helsinki*

*Professor Jan-Erik Lönnqvist*  
*(social psychology)*  
*University of Helsinki*

*Docent Markku Verkasalo*  
*(psychometrics)*  
*University of Helsinki*

Pre-examiners: *Professor Andreas Demetriou*  
*(psychology)*  
*University of Nicosia, Cyprus*

*Professor Hannu Savolainen*  
*(special education)*  
*University of Jyväskylä*

Custos: *Professor Markku Jahnukainen*  
*(special education)*  
*University of Helsinki*

Opponent: *Professor Benő Csapó*  
*(education)*  
*University of Szeged, Hungary*

ISBN 978-951-51-0185-3 (print)

ISBN 978-951-51-0186-0 (pdf)

ISSN 1799-2508

Unigrafia

2014

**Mari-Pauliina Vainikainen**

**Finnish primary school pupils' performance  
in learning to learn assessments**

A longitudinal perspective on educational equity

---

**Abstract**

During the last decade, Finnish pupils' performance in educational assessment studies has steadily declined. At the same time the differences between pupils – and in the capital area also the differences between schools – have increased, and girls usually outperform boys in most assessed domains. The aim of the present study was to examine how these differences develop during primary education, with a special emphasis on the development of the performance of pupils in need of support for their studies. This was done by following three different samples of primary school pupils in two municipalities: In Helsinki a sample of 608 pupils was followed from the beginning of the first grade to the end of the sixth grade, and in Vantaa two full cohorts ( $N \approx 2000$  in each) were assessed in the first/third grade and again in the third/sixth grade. In the beginning of the first grade the pupils took a learning preparedness test, and teachers evaluated their initial reading skills. At the turn of the third and fourth grade the pupils completed the Finnish learning to learn scales, which addressed a wide scope of cognitive competences and learning-related attitudes. Learning to learn assessments were repeated at the end of the sixth grade before the transfer to lower secondary education. Additional information was collected about pupils' social relationships, task interest and effort as measured by time investment, based on the log files of computer-based assessment. Multiple-group structural equation modelling, repeated measures general linear modelling and variance components modelling were applied in four substudies for testing the hypotheses about the influences of prior cognitive competences, attitudes, interest and effort on performance and about the

different trajectories of their development within municipalities, schools, classes and peer groups.

The results showed that whereas girls were evaluated by their teachers as being slightly better readers already when they came to school, there was no gender difference in pupils' performance in the learning preparedness test. Girls, however, gained slightly more in reading comprehension during the first three years of basic education. Boys in Helsinki outperformed girls in mathematical thinking in the beginning of third grade, but girls closed the gap by the end of the sixth grade. Mothers' lower education and pupils' support needs were related to lower initial competences, but the differences did not increase during the first three years of basic education. In contrast, in regard to reasoning skills pupils with support needs even closed the gap to some extent. The gap between pupils with support needs and others, however, increased from the beginning of the fourth grade to the end of the sixth grade in both municipalities. Between-school differences slightly increased during the six years of follow-up in Helsinki, but in Vantaa the variation remained between classes in schools. From the end of the third grade to the end of the sixth grade girls improved their performance slightly more than boys in both municipalities. The log data analyses of the computer-based assessment in Vantaa revealed that girls' advantage could be completely explained by their more positive attitudes and greater effort as measured by their time investment in the tasks. Reduced time investment and higher levels of detrimental attitudes also provided a partial explanation as to why pupils with identified support needs did not reach their expected level of performance in the sixth grade assessment.

As expected, learning-related attitudes declined with age, but this change was unrelated with the changes in performance. Changes in task interest, however, were a meaningful predictor of later performance. Changes in attitudes and interest happened to some extent in classes and peer groups, and boys – who were also identified as having support needs more often than girls – seemed to be more vulnerable to the influences of their boy classmates both regarding their attitudes and task behaviour in the assessment situation.

---

*Keywords:* learning to learn, cognitive development, learning-related attitudes, task interest, time on task, gender differences, support needs, educational equity

**Mari-Pauliina Vainikainen**

**Oppimaan oppimisen arvioinneissa osoitetun osaamisen  
kehittyminen kuuden ensimmäisen kouluvuoden aikana**  
Seurantatutkimuksia koulutuksellisen tasa-arvon toteutumisesta

---

**Tiivistelmä**

Viime vuosikymmenen aikana suomalaisoppilaiden suoriutuminen koulutuksen arviointitutkimuksissa on laskenut tasaisesti. Samaan aikaan oppilaiden väliset erot – pääkaupunkiseudulla myös koulujen väliset erot – ovat kasvaneet, ja tytöt saavat poikia parempia tuloksia useimmilla arvioinnin osa-alueilla. Tämän tutkimuksen tarkoituksena oli selvittää, miten nämä erot kehittyvät alaluokkien aikana kiinnittäen erityistä huomiota oppilaisiin, joiden on todettu olevan tehostetun tai erityisen tuen tarpeessa. Tutkimuksessa käytettiin kolmea seuranta-arviointiaineistoa kahden kaupungin alueelta: Helsingissä 608 oppilaan satunnaisotosta seurattiin ensimmäisen luokan alusta kuudennen luokan loppuun, ja Vantaalla kahta noin 2000 oppilaan kohorttia seurattiin ensimmäiseltä/kolmannelta luokalta kolmannen/kuudennen luokan loppuun. Ensimmäisen luokan alussa oppilaat tekivät Ensiaskleet-oppimisvalmiustestin. Opettajat myös arvioivat heidän lukutaitonsa koulun aloitushetkellä. Kolmannen ja neljännen luokan vaihteessa oppilaat osallistuivat oppimaan oppimisen arviointiin, joka kattoi laajan kirjon osaamistehtäviä ja asennekyselyitä. Oppimaan oppimisen arviointi toistettiin vielä kuudennen luokan lopussa. Aineistoa koottiin myös oppilaiden sosiaalisista suhteista, tehtäväkiinnostuksesta ja arviointitilanteessa osoitetusta yrittämisestä, jota mitattiin rekisteröimällä vastaamiseen käytetty aika tietokonepohjaisessa arvioinnissa. Useamman ryhmän samanaikaisella rakenneyhtälömallinnuksella, toistomittausten varianssianalyysillä ja varianssikomponenttimallinnuksella testattiin neljässä osatutkimuksessa hypoteeseja aiemmin osoitettujen kognitiivisten taitojen, asenteiden, tehtäväkiinnostuksen ja yrittämisen vaikutuksista arviointitehtävissä suoriutumisi-

seen sekä näiden erilaisiin kehityskulkuihin eri kaupungeissa, kouluissa, luokissa ja kaveriryhmissä.

Tulokset osoittivat, että vaikka opettajat arvioivat tyttöjen lukutaidon koulun alussa hieman poikia paremmaksi, oppimisvalmiustehtävissä ei havaittu sukupuolieroa. Tyttöjen luetunymmärtämistaito kehittyi kuitenkin hieman poikia nopeammin kolmen ensimmäisen kouluvuoden aikana. Helsingiläispojat suoriutuivat tyttöjä paremmin matemaattisessa ajattelussa neljännen luokan alussa, mutta tytöt kuroivat eron umpeen kuudennen luokan loppuun mennessä. Äidin matala koulutustaso ja oppilaan tuen tarve olivat yhteydessä heikompaan lähtötasoon, mutta erot eivät kasvaneet kolmen ensimmäisen kouluvuoden aikana. Päinvastoin tukea saavat oppilaat saivat jopa muita kiinni päättelytaidoissaan. Tukea tarvitsevat oppilaat alkoivat kuitenkin jäädä osoitetussa osaamisessaan muista jälkeen kolmannen luokan jälkeen kummankin kaupungin kouluissa. Koulujen väliset erot kasvoivat hieman Helsingissä kuuden vuoden seurannan aikana, mutta Vantaalla erot pysyivät luokkien välisinä koulujen sisällä. Tyttöjen suoritustaso nousi hieman poikia enemmän kolmannen ja neljännen luokan vaihteesta kuudennen luokan loppuun kummassakin kaupungissa. Vantaan tietokonepohjaisen arvioinnin lokitietojen analyysi osoitti, että tyttöjen paremmuus selittyi täysin heidän myönteisemmällä asenteillaan ja tehokkaammalla yrittämislään, jota mitattiin tehtäviin käytetyn ajan kautta. Muita vähäisempi ajankäyttö ja haitalliset asenteet taas selittivät osin sitä, miksi tukea tarvitsevat oppilaat eivät saavuttaneet omaa ennustettaan kuudennen luokan lopussa.

Oppimista koskevat asenteet laskivat odotetusti iän myötä, mutta muutos ei ollut suoraan yhteydessä tehtäväsuoritukseen. Tehtäväkiinnostuksen muuttuminen sen sijaan ennusti myöhempää suoritusta. Asenteiden ja kiinnostuksen muutos oli osin luokka- ja kaveriryhmätason ilmiö. Pojat – joilla myös todettiin useammin tukitarpeita kuin tytöillä – näyttivät olevan tyttöjä alttiimpia luokkansa poikien vaikutuksille sekä asenteiden että arviointitilanteessa toimimisen osalta.

---

*Avainsanat:* oppimaan oppiminen, kognitiivinen kehitys, oppimiseen liittyvät asenteet, tehtäväkiinnostus, ajankäyttö arviointitilanteessa, sukupuolierot, tehostettu ja erityinen tuki, koulutuksellinen tasa-arvo



# Acknowledgements

Conducting this doctoral thesis while working full-time except for the final spurt has not been an easy task. Of course, my normal duties are quite closely related to what I have done here, but I would have never managed to realise this childhood dream without the large supportive networks I have been lucky enough to have in both my professional and personal life. Since I got my Master's degree twelve years ago I have come a long way, and partially in a very different direction to the one I was originally planning. A series of coincidences opened new opportunities and challenges for me, and as I did not quite know where I was going for a long time, it took me nine years to decide that I even wanted to start doctoral studies. And yet, I have been working with these educational assessment studies all along the way. During all these years, I have not wanted to give up on being a psychologist to become an educational researcher, but recently I have come to understand that these multiple professional identities only support one another.

As this doctoral thesis is as multidisciplinary as my work and educational history are, there are many colleagues and collaboration partners from different work communities I want to thank. First of all, I want to thank my supervisor and mentor, Professor Emeritus Jarkko Hautamäki. It has been a privilege to work so closely with you for such a long time. At this point of my career, I am not even able to distinguish my own thinking from what originally came from you, and I really wish that in the future I would be able to somehow continue the great work you have been doing – even if you once called me a poor successor as my way of seizing new opportunities seems to be as arbitrary as your own has been.

This, of course, is partially due to the nature of the work we are doing: short and intensive multiple projects make detailed planning often impossible, and even if we were extremely busy today, nobody knows what tomorrow brings. By “we” I mean the whole Centre for Educational Assessment, which – regardless of the constant insecurity – must be the best work community in the whole of the University of Helsinki. Sirkku Kupiainen, thank you for co-thinking and co-writing with me and also for often arguing with me about my solutions – feedback is always the key to

improvement. Natalija Gustavson, thank you for always believing in me and taking care of me when I am too absorbed in my work. Thank you also for drawing the sociograms used in the last substudy of this thesis. Ninja Hienonen, thank you for relieving my workload in regard to administration, and for all your help with the practicalities, whatever they may be. I would not manage to do my work if I did not have somebody I can always trust. Mikko Asikainen, thank you for always helping me with my technical problems and, of course, for designing the data collections which provided all the data for this thesis. Thank you also for having been my friend for so many years. Jukka Marjanen, thank you for always being there for me when I want to understand the statistical analyses I have learned mostly by trial-and-error and for teaching me so many new things about the methods. Satu Koivuhovi, thank you for studying the methods together with me – I think together we learn much more efficiently, and that will hopefully continue during the years to come. Jukka Määttänen, thank you for processing the hundreds of thousands of pages of paper-based data into the data files used in this study. Raisa Ahtiainen, Meri Lintuvuori and Touko Hilasvuori, thank you for being a part of this great work community even though our collaboration is not directly reflected in this thesis. I also want to thank Tiina Mäkelä, Maria Beirad, Eeva Rontu, Tomas Valkonen, Tommi Karjalainen, Marja-Liisa Kieksi, and Pekka Arinen for creating this unique work atmosphere even though you have already moved on to new challenges, as well as Sanna Eronen and Laura Kauppinen for bringing me here so many years ago and introducing me to these topics. Risto Hotulainen, you are relatively new here, but we have had a great start for our collaboration and I hope it continues for many more years. Thank you for that. And last but definitely not least, Helena Thuneberg, thank you for your devotion to our collaboration, which nowadays stretches to so many different content areas. Even though not much of it is directly visible in this thesis, your support has been invaluable during this process as you have made me believe in myself and given me emotional support whenever I have needed it.

Outside our closest work community, I want to first thank my two other supervisors, Professor Jan-Erik Lönnqvist and Docent Markku Verkasalo, for providing me new perspectives on my work and for not

letting me narrow my scope too much. I also want to thank you for our research collaboration beyond this thesis, which we hopefully find more time for in the near future.

I want to give special thanks to my pre-examiners, Professor Andreas Demetriou from the University of Nicosia and Professor Hannu Savolainen from the University of Jyväskylä. Your feedback made me believe that I may be on the right track after all. Thank you also for your valuable suggestions, which helped me to improve my work considerably at the last moment before bringing it into the public discussion. And my warmest thanks go to my opponent Benő Csapó from the University of Szeged. I respect your work so much, and it is a great honour to have you here discussing my thesis in public.

I want to express my greatest gratitude to the Education Departments of the Cities of Helsinki and Vantaa for launching the two collaboration projects, within which this study has been conducted. Especially, I want to thank Paula Sermilä, Outi Salo, Marjo Kyllönen and Satu Gripenberg for enabling the Helsinki longitudinal study and the use of its data in this thesis and Eero Väättäinen, Ilkka Kalo and Tapio Lahtero for their determination in starting and implementing the Vantaa panel study, which is completely unique in this country due to its scale. Thank you also for the always pleasant collaboration. In addition, I want to thank the Ministry of Education and Culture and the National Board of Education for understanding the importance of the assessment of the more general outcomes of education. And of course, the greatest thanks must be directed to the schools – the principals, teachers and especially pupils – who have participated in the studies. I also want to thank City of Helsinki Urban Facts for allowing me a grant, which made it financially possible for me to take some time off from my other duties in order to finish this doctoral thesis. Furthermore, I want to thank Nick Kirkwood for proofreading this thesis and for helping me put the articles and prepositions in the correct places – something that seems to be insurmountable for most of us who speak Finnish as a first language.

During recent years I have been fortunate to have had the possibility to collaborate with professionals from very different fields here in Finland. This has been a very important source of professional growth for me, as new perspectives on the same phenomena always bring something

new to my own work. First I want to thank Tuomo Tikkanen, Vesa Nevalainen, Teemu Ollikainen and Annarilla Ahtola for not letting me forget that deep down inside I will always be a psychologist. From the field of special education, I want to first of all thank my mother Aila Vainikainen for introducing me to the development of the field already at a very early age. During the later years, I continued to follow the reform together with Pirjo Koivula, Terhi Ojala, Sirpa Oja, Heikki Happonen, Hanna-Mari Sarlin and Aija Rinkinen, just to name those who I have spent the most time with travelling around this small but geographically big country. Thank you all. Here at the University of Helsinki I want to thank Markku Jahnukainen and Elina Kontu for their recent efforts in gathering research results regarding the reform from different projects towards a common understanding of the development of this field. Of course, I also thank Markku for acting as the custos for my doctoral defence. Markku Niemivirta and Patrik Scheinin, thank you for your work in developing the attitude scales used in this thesis and for our occasional inspirational methodological discussions. Furthermore, I want to thank Hannu Salmi for introducing me to science centre pedagogics and for the effective writing collaboration, which will be a lot of fun also in the near future. I want to thank Arja and Matti Rimpelä for bringing the health and well-being perspective to our assessment results during the recent years and Venla Bernelius for making me understand the mechanisms of differentiation of schools from the perspective of urban geography. And special thanks to Jorma Kuusela for our always uplifting discussions: every time we meet you manage to bring something new to my own view on educational assessment. Also, you were the one who first introduced me to structural equation modelling, which has lately become almost my specialty. Thank you so much.

Recently our assessment studies have gained a much more international perspective due to the close collaboration with the Research Group Computer-Based Assessment of the University of Luxembourg. I want to thank the whole group, but the greatest thanks of all go to Samuel Greiff. You have inspired me, opened many new opportunities for me and also pressured me quite a lot, making me a much better researcher than I could have ever imagined. Without you this doctoral thesis would not be

here at all. Thank you also for being always there for me when I need a friend who really understands the joys and challenges of this work.

I also want to thank all my other friends and apologise that I have lately had so little time for you. Especially I want to thank Sari Nyström who has not allowed me to isolate myself completely from the outside world. Our daily discussions have helped me so much when trying to survive through all this stress.

I want to thank my parents Aila and Markku Vainikainen for your never-ending care and support. My sister and brother, Katariina Popelka and Vesa-Matti Vainikainen, thank you and your families for helping me remember that there are also other things in life outside this research world. Finally, thank you my dear husband Björn Ole Eide for tolerating me during these extremely stressful times and being always there for our boys when I have not been able to be present. And Elias and Daniel, thank you for bringing me joy in my life.

Helsinki, 5 October 2014

Mari-Pauliina Vainikainen



# Contents

1	INTRODUCTION .....	1
2	THEORETICAL BACKGROUND .....	7
2.1	The Finnish educational system as the context of this study .....	7
2.1.1	A short history of comprehensive school.....	7
2.1.2	Comprehensive school nowadays.....	8
2.1.3	The Finnish support system .....	9
2.1.4	Assessing educational outcomes .....	13
2.2	Development of learning to learn skills during primary education	16
2.2.1	Introduction of the Finnish LTL model .....	16
2.2.2	Cognitive development during primary education .....	19
2.2.3	Learning-related attitudes.....	27
2.2.4	Peer influences on learning and learning-related attitudes.....	31
2.3	Performance in a low-stakes assessment situation .....	34
2.3.1	Task interest .....	35
2.3.2	Time investment and effort.....	37
2.4	Summary.....	39
3	DATA .....	41
3.1	Participants .....	41
3.1.1	The Helsinki sample .....	41
3.1.2	The two Vantaa cohorts .....	43
3.2	Measures .....	44
3.2.1	First grade assessments of learning preparedness .....	46
3.2.2	Learning to learn assessments from grade three to grade six.	48

4	SUBSTUDY 1: EXPLAINING TEN-YEAR-OLDS' TEST PERFORMANCE IN TWO MUNICIPALITIES.....	55
4.1	Participants.....	57
4.2	Measures .....	57
4.2.1	Cognitive tasks .....	57
4.2.2	Learning-related attitudes .....	57
4.2.3	Other variables in the model.....	58
4.3	Statistical methods.....	59
4.4	Results and discussion.....	59
4.5	Conclusions of Substudy 1 .....	69
5	SUBSTUDY 2: PREDICTING LATER PERFORMANCE BY CHANGES IN LEARNING-RELATED ATTITUDES AND TASK INTEREST .....	71
5.1	Participants.....	72
5.2	Measures.....	73
5.2.1	Cognitive tasks .....	73
5.2.2	Learning-related attitudes .....	74
5.2.3	Task interest .....	74
5.2.4	Grouping variables .....	75
5.3	Statistical methods.....	75
5.4	Results .....	76
5.5	Conclusions of Substudy 2 .....	85
6	SUBSTUDY 3: GENDER DIFFERENCES, SUPPORT NEEDS, LEARNING- RELATED ATTITUDES AND TIME ON TASK .....	87
6.1	Participants.....	88
6.2	Measures .....	89
6.2.1	Cognitive tasks .....	89
6.2.2	Learning-related attitudes .....	89
6.2.3	Time on task .....	91
6.2.4	Background variables.....	91
6.3	Statistical methods.....	93
6.4	Results .....	93
6.5	Conclusions of Substudy 3.....	102



7 SUBSTUDY 4:SCHOOL, CLASS AND PEER EFFECTS ON THE DEVELOPMENT OF LEARNING TO LEARN SKILLS .....	105
7.1 Participants .....	107
7.2 Measures .....	107
7.2.1 Cognitive tasks, attitude scales and measures of task interest and time investment.....	107
7.2.2 Social relationships in school classes .....	108
7.3 Statistical methods .....	108
7.4 Results.....	108
7.4.1 Social networks of school classes.....	108
7.4.2 Variance components modelling.....	111
7.5 Conclusions of Substudy 4 .....	123
8 DISCUSSION .....	125
8.1 Development of performance in cognitive assessment tasks.....	126
8.1.1 From first grade to the turn of the third and fourth grade .....	126
8.1.2 From the turn of the third and fourth grade to the end of the sixth grade.....	128
8.1.3 Gender differences in the development of performance .....	130
8.1.4 Support needs and the development of performance .....	132
8.1.5 Socio-economic background and the development of performance .....	135
8.1.6 School, class and peer group effects on the development of performance.....	136
8.1.7 Summary of the development of performance in the cognitive tasks .....	140
8.2 Learning-related attitudes: Their development and effects on performance.....	142
8.2.1 Gender differences in the development of learning-related attitudes .....	145
8.2.2 Support needs and the development of learning-related attitudes .....	147
8.2.3 Socio-economic background and the development of learning-related attitudes.....	149

8.2.4 School, class and peer group effects on the development of learning-related attitudes .....	149
8.2.5 Summary of the development of learning-related attitudes...	151
8.3 Task interest and performance in the assessment.....	152
8.3.1 Gender differences in task interest .....	154
8.3.2 Support needs and task interest .....	154
8.3.3 School, class and peer group effects on task interest.....	156
8.3.4 Summary on task interest in relation to performance in the assessment.....	157
8.4 Time investment and performance in the assessment .....	158
8.4.1 Gender differences in time investment.....	159
8.4.2 Support needs and time investment.....	160
8.4.3 School, class and peer group effects on time investment .....	162
8.4.4 Summary on the effects of time investment on performance	163
8.5 Limitations of the study .....	165
8.6 General conclusions and practical implications .....	170
REFERENCES .....	173

APPENDIX 1A: Bivariate correlations of the variables used in the models of Substudy 1 for the Helsinki data

APPENDIX 1B: Bivariate correlations of the variables used in the models of Substudy 1 for the Vantaa data

APPENDIX 2: Bivariate correlations of the variables used in the models of Substudy 2

APPENDIX 3: Bivariate correlations of the variables used in the models of Substudy 3

# 1 Introduction

The effectiveness of basic education is often evaluated through low-stakes educational assessment studies both at a national and an international level. Finland provides an example of a system in which the monitoring of the educational outcomes is based entirely on sample-based assessments which normally do not have any consequences for the participating students at an individual level. In these low-stakes assessments, students' performance has clearly declined between 2006 and 2012. This has been observed in national assessments of different school subjects (Hirvonen, 2012; Kärnä, Hakonen & Kuusela, 2012; Lappalainen, 2011; Rautopuro (Ed.), 2013), assessments of cross-curricular learning to learn skills (Hautamäki, Kupiainen, Marjanen, Vainikainen & Hotulainen, 2013; Kupiainen, Marjanen, Vainikainen & Hautamäki, 2011) and international comparisons like the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA, OECD, 2013a; [see Hautamäki et al., 2013, for a review]).

This phenomenon, however, is not unique to Finland. Similar results have been obtained in several other Western countries too (e.g. OECD, 2013a), and in many countries it has launched intensive public speculation about the reasons for this unwanted development. Most of these results suggest that education in Western countries – for some reason or another – is losing its importance in young people's lives, and due to this they no longer put their best efforts into school work – or into assessment tasks for that matter. However, cross-sectional studies which are typically conducted during secondary education have been able to provide only very limited evidence to support any of these claims. Longitudinal assessment studies are therefore urgently needed to find more evidence-informed explanations of how differences between students, classes, schools, districts and even countries develop, and what kind of factors are related to different paths of this development both with regard to what the students really can do and what they are willing to show in assessments.

Besides the debate around the decline of the results, cross-sectional assessment studies have brought into public discussion several important topics that are related to the equity of education. In Finland, girls have since the beginning of the current educational assessment system in the mid-1990s outperformed boys in the most assessed areas (Hautamäki et al., 1999; 2000; 2002b; 2003; 2005; Hirvonen, 2012; Kärnä et al., 2012; Lappalainen, 2011; Rautopuro (Ed.), 2013), and the latest PISA-results show that the difference is only getting bigger (Kupari et al., 2013). Moreover, the PISA 2012 results show that the increase in the gender differences is a global phenomenon, and girls are starting to perform better also on areas which have traditionally been considered as boys' strengths, e.g. mathematical literacy (OECD, 2013a). In a short review of the American Achievement Test results Kenney-Benson, Pomerantz, Ryan & Patrick (2006) concluded that at the time of writing the review boys had just lost their edge over girls in mathematical assessments while boys were still performing better in science. Thus, the beginning of the increase in gender differences in external assessments can be dated back to the same period when the Finnish assessment results began to decrease (Hautamäki et al., 2013). When big gender differences are observed when assessing 15-years-old pupils, it is already too late to design interventions for addressing the problem behind the results. Therefore, the development of gender differences during earlier school years has to be understood before any interventions can be planned.

Another major concern regarding educational equity is the differentiation of schools, partly due to the differentiation of residential areas but also due to higher-educated parents not choosing the local school for their children (Bernelius, 2013; Kosunen, 2014). The Finnish basic education system is based on the idea of strong local schools providing equal opportunities for learning for everyone, regardless of social or educational background of the family. However, the differentiation of schools contradicts this idea even though in Finland this development is still very moderate compared to many other European countries – and also the other Nordic countries which have a relatively similar history in regard to the basic education system (e.g. Yang Hansen, Rosén & Gustafsson, 2011). In PISA 2006, the segregation of schools in Finland was the lowest of all the participating countries,

both when measured by the distribution of socioeconomic status of pupils and by their performance in the assessment (Willms, 2010). Nevertheless, the phenomenon of school differentiation is emerging in the biggest cities also in Finland. In Helsinki it can already be seen in assessment studies as relatively large between-school differences which are strongly related to the educational background of the parents (Bernelius, 2013; for results regarding very small between-school differences at a national level in Finland still in 2012, see Hautamäki et al., 2013; Kupari et al., 2013). However, it has not yet been shown with Finnish data whether these differences tell anything about students gaining more in schools of higher socio-economical status or if they only report about the background-related differences that have been there since the pupils started school at the age of seven. This is despite the evidence from other countries that both socioeconomic status- and performance level-based segregation of schools are harmful and that policies aimed at increasing inclusion require an understanding of the mechanisms of how pupils are allocated to schools (Willms, 2010). Therefore, if pupils gain more – measured either by cognitive learning outcomes or the development of positive learning-related attitudes – in some schools compared to others, it needs still to be examined to what extent the differences develop at school level and to what extent they depend on the more random effects of classes and peer groups.

The third key element in the discussion about educational equity, related to the so-called local school principle, is the support provided for the weakest learners; in Finland this support has been considered as one of the key factors explaining the country's success in international comparisons (Sabel, Saxenian, Miettinen, Kristensen & Hautamäki, 2011). In the PISA studies the weakest Finnish pupils have usually clearly outperformed their comparison groups in other countries (Kirsch et al., 2002; OECD 2004; 2007; 2010; 2013a) while the differences between better performers have been much smaller. Unlike the differences related to gender and educational background of parents, there are no clear indicators of change in the assessment results from this aspect, even though in PISA 2012 there were slightly more pupils who did not reach the lowest acceptable level compared to earlier cycles (Kupari et al., 2013). The support system has, however, been adjusted between 2007

and 2011 to meet the constantly increasing support needs and to better follow the principles of prevention and early intervention (Thuneberg, Vainikainen, Ahtiainen, Lintuvuori, Salo & Hautamäki, 2013). In order to evaluate the effectiveness of the new support system it is important to pay extra attention to how pupils receiving support are performing in educational assessment studies and how the differences between students of different performance levels develop over time.

The purpose of this study is to look for at least partial answers to the concerns stated above using the data of two longitudinal studies of the development of learning to learn skills in primary school. The Finnish learning to learn assessment method (LTL, Hautamäki et al., 2002a; Hautamäki & Kupiainen, in press) is a low-stakes assessment of cross-curricular skills used for monitoring the effectiveness of education at a municipal and occasionally also at a national level. Some of the first indications of the decline in the Finnish pupils' performance came from a municipal LTL study (Kupiainen, Marjanen, Vainikainen & Hautamäki, 2011), and it is very likely that the same factors affecting the development of these skills, and the pupils' willingness to give their best in the assessment of them, are also visible in the results of other national and international low-stakes assessments.

The two data sets provide interesting opportunities for comparisons in order to understand the mechanisms of how the observed differences develop over time. The first data set is from Helsinki, where a sample of 608 pupils has been followed from the beginning of the first grade to the end of the sixth grade, that is, from the age of 7 to the age of 13. The second data set is from a panel study in Vantaa where two whole age cohorts were followed from the beginning of the first grade to the end of the third grade (cohort 1, about 2000 students) and from the end of the third grade to the end of the sixth grade (cohort 2, about 2000 students). In both cities, the pupils completed the first, the third/fourth and the sixth grade versions of the Finnish LTL tests. However, there are some important differences in the two studies: In Helsinki the data were collected entirely on paper while in Vantaa the last data collection cycle was performed with the computer-based version of the LTL test, which gave an opportunity to utilise log data in evaluating the pupils' effort in the tasks. In Helsinki, additional data about pupils' peer groups were

collected. The different educational policies of the two cities enable interesting comparisons too: In Helsinki it is much more common for parents to choose other than the local school (Bernelius, 2013), while in Vantaa school choice is quite restricted (Varjo & Kalalahti, 2011; Varjo, Kalalahti & Silvennoinen, 2014). Moreover, the differences between residential areas and schools with regard to socio-economical status are bigger in Helsinki, while in Vantaa the differences are mostly within schools.

Using the two longitudinal assessment data sets, this study consists of four substudies. The first substudy examines how pupils' performance in the third/fourth grade LTL cognitive tasks is predicted by their learning preparedness at the beginning of the first grade and how learning-related attitudes explain performance when prior cognitive competence is controlled for. Educational equity is then evaluated by adding gender, mother's education, support needs and the effect of individual schools in the structural equation model in order to see whether they have systematic effects on third/fourth grade performance. The same model is fitted to the two data sets separately to discuss whether the municipal policies regarding school choice, and the fact that the schools in Helsinki are more differentiated also because of the differentiation of residential areas, could produce different patterns in how the pupil-level background variables explain performance beyond prior competence and attitudes.

The second substudy focuses on the role of attitude and interest change from the fourth to sixth grade in explaining sixth grade performance when fourth grade performance and pupils' general cognitive competence are controlled for. In this study, of the cognitive measures only items that were identical at both measurement points were used to be able to make conclusions about the development of the skills. For the same reason, only the Helsinki data were used as the use of computers at the second data collection point in Vantaa would have required a mode effect study before claiming that the cognitive items were identical in both measurements (cf. Csapó, Molnár, & Nagy, 2014). The model specified for testing the effects of attitude and interest change on performance is fitted on girls and boys, pupils with different backgrounds of school achievement and those with different support needs to evaluate whether pupils in some subgroups gain more than in

other groups regarding the development of cognitive competences. Also, it is examined whether some of the subgroups are more vulnerable to the effects of attitude and interest change on later performance.

The third substudy focuses on pupils' effort in assessment tasks as measured by their time investment in them. It has already been shown with a ninth grade national sample that the effect of pupils' detrimental attitudes on performance in the cognitive learning to learn tasks is mediated by time investment in the assessment situation when their prior school achievement is controlled for (Kupiainen, Vainikainen, Marjanen & Hautamäki, 2014). In the present study the same phenomenon is studied with sixth graders using the log data of the Vantaa study. The model specified here is somewhat simpler than in Kupiainen et al. (2014) partly due to the shorter version of the LTL test, but also to enable the use of additional background variables in the model. Thus, the present study takes the next step from the study of Kupiainen and colleagues in regard to examining whether gender and support needs affect the relative roles of mastery and detrimental attitudes in explaining time on task, and how these background variables together with time on task explain performance in a low stakes learning to learn assessment.

The last substudy focuses on peer influences on performance and attitude change from the turn of the third and fourth grade to the end of the sixth grade. Based on sociograms drawn for the Helsinki schools a simplified method for determining approximations of realistic peer groups in primary school is developed, and these groups are used as the lowest level in variance component models on performance and attitude change in addition to the traditional school and class levels for both data sets. The aim of this substudy is to find out if the performance and attitude changes observed in the other substudies have happened at an individual level or if there are systematic group effects: school-level effects which would first of all tell about differences in school culture, class-level effects that would at least partially be related to individual teachers, and peer effects, which would most likely not depend on the school or teachers per se but which would be useful information when designing interventions for enhancing the development of pupils' thinking and learning skills.



## 2 Theoretical background

### 2.1 The Finnish educational system as the context of this study

#### 2.1.1 A short history of comprehensive school

In the 1960s, Finland was moving away from being an agrarian country into becoming a Scandinavian welfare state. The school system, however, still maintained many of the inequities of the old class society and needed therefore complete reform. In the political atmosphere of the 1960s Finland decided to choose the same route the other Nordic countries had already taken – to introduce a comprehensive school system in which instruction is offered to whole age cohorts in shared settings, free of charge, with no differentiation based on prior abilities until pupils are about 15 years old. The reform was implemented gradually starting first in northern Finland before moving downwards to southern Finland. Already in 1965 some variation of the new system had spread to 25 municipalities, and the whole of Lapland had adopted the system by 1972. The last areas to implement the reform in 1977 were the biggest cities of the Metropolitan area (Aho, Pitkänen & Sahlberg, 2006).

The introduction of the comprehensive school system led to other major changes as well. Teacher education was reorganised and moved to universities to secure high quality teaching for every pupil. The revision of the curriculum began in the mid-1960s, and in 1972 the Ministry of Education ordered the new comprehensive school curriculum to be introduced in all schools. Also school textbooks had to be approved by the National Board of Education. The quality of education was monitored by an external school inspection system (Aho et al., 2006).

The decentralisation of administration took place in the 1980s and 1990s. In the mid-1980s municipalities and schools were requested to develop their own curricula, following the principles of the national core-curriculum, and this is still the practice in 2014. Schools were also given a specified amount of teaching hours based on the number of pupils and the freedom to decide how to use their resources in these hours (Aho et

al., 2006). In addition, the controlling of the study materials was loosened. The school inspection system was ceased in the early 1990s, which lead to the development of the modern framework of evaluating educational outcomes (National Board of Education, 1998, English translation 1999).

### **2.1.2 Comprehensive school nowadays**

In the Nordic tradition of a public educational system local school – the school which is geographically located nearest to a pupil's home – has primarily been the school everyone should attend regardless of background or special needs. Even though in Sweden there has been a change towards a more segregated system based on selection during the last 20 years (Yang Hansen et al., 2011), Finland changed its educational legislation in 2011 to make the role of local schools even stronger than before (Thuneberg et al., 2014). In 2012, 96 % of the nine-year comprehensive schools were run by municipalities (the Official Statistics of Finland, [www.stat.fi](http://www.stat.fi)) and followed local curricula which are regulated by the National Core Curriculum (National Board of Education, 2004). Except for a small proportion of pupils with very high special education needs, everybody is to attain the same curricular goals. Even though some special education pupils are still taught in separate schools, of all the countries that participated in PISA 2006, the segregation of schools was the lowest in Finland both when measured by the distribution of socioeconomic status of pupils and by their performance level in the assessment (Willms, 2010).

In Finland, the 9-year compulsory education begins relatively late compared to most other countries (OECD, 2013d). Children begin the first grade in August of the year they turn 7 years old, and until then they – theoretically – can stay at home outside of any formal pedagogical system. Regardless of the fact that Finnish children are not really taught academic skills until they are 7 years old, they do not seem to have a disadvantage later in international comparisons (Kirsch et al., 2002; OECD 2004, 2007, 2010, 2013a). In practice, however, 99 % ([www.stat.fi](http://www.stat.fi)) of Finnish children go to pre-school for one year before beginning compulsory education, and before that most children have already been in daycare. Therefore, when assessing first graders'

competences immediately when they begin their school career as it is done in the present study, the results do not tell purely about their prior competences which are independent from any formal teaching, but they also tell about the outcomes of the pre-school education. Traditionally, pre-schools have been located in daycare centres, and they have only weakly been connected to normal school work. During the last few years however, alongside with the partial educational reform regarding pupil support which is described in the next section, pre-schools have to an increasing degree been transferred both administratively and physically to comprehensive schools. Until now pre-school has concentrated on the basic skills needed for learning mostly through play instead of really teaching academic skills, but it is to be seen if the transfer to schools results in the pre-school year becoming more school-like also in regard to contents – something that has already happened for example in Norway some years earlier. In the present study, first graders' initial competences are assessed in two different municipalities with slightly different educational policies, and it will be interesting to see if there are systematic municipal-level differences in children's preparedness for learning. In a recent Finnish study (Ahtola & al., 2011), transition practices from kindergarten to first grade were found to have an effect on performance in reading and mathematics one year later. The strongest predictor of later performance was a close connection to daily school work already during the pre-school year in the form of regular shared lessons, for instance. This, of course, supports the administrative changes that have been made during the past years when transferring pre-school classes to schools. It also makes it interesting to evaluate how the level of learning preparedness as demonstrated at the school start predicts success in different domains even years later (cf. Duncan et al., 2007), not only for understanding children's cognitive development but also to develop pre-school and transition practices further to secure a smooth school start for all learners.

### **2.1.3 The Finnish support system**

Since the implementation of comprehensive school in Finland, a key component for securing educational equity has been the system for supporting the weakest learners (Graham & Jahnukainen, 2011; Sabel et

al., 2011). The support system can be interpreted as having been relatively effective, as in international comparisons the weakest Finnish pupils have usually outperformed their comparison groups in other countries (Kirsch et al., 2002; OECD 2004; 2007; 2010; 2013a) while the differences between better performers have been much smaller. Nevertheless, the system has been adjusted during recent years to meet the constantly increasing support needs and to better follow principles of prevention and early intervention (Thuneberg et al., 2013).

Earlier, the support system consisted of general support which could be provided without a referral to special education and special education which required an official administrative decision based on a statement from either a school psychologist or a medical doctor (Jahnukainen, 2011). In 2006, the ten biggest municipalities in Finland together expressed their concern regarding the organisation and the functionality of this system. This was mainly due to the forever increasing number of special education referrals which at that time was as high as 8 % of the pupil population, half of which was taught in segregated classes or special schools (Lintuvuori, 2010; Statistics of Finland). As a result (Salo, 2010), a new Special Education Strategy was introduced by the Ministry of Education (2007), and an extensive in-service training programme was started to give the Finnish municipalities means by which to be prepared for the upcoming change in educational legislation (Ahtiainen et al., 2012; Thuneberg et al., 2013).

During the reform, the division of general education and special education was replaced by a three-tiered support model which is based on a high-quality basic education. The starting point of the new model is that – with some exceptions – moving to the next tier is possible when the previous tier has proven to be insufficient. The first tier, *general support*, is meant for everyone, and it should be provided immediately when any concern is raised. The first-tier interventions can be conducted at the school- or class-level, or they can be individually designed for specific pupils. The most common means of support of this tier are differentiation and flexible grouping, remedial instruction and part-time special education either as co-teaching or in a smaller group (National Board of Education, 2011; Thuneberg et al., 2013).

If general support is concluded to be insufficient based on multiprofessionally conducted *pedagogical assessment*, *intensified support* is organised according to an individual learning plan. Intensified support consists largely of the same type of interventions as general support, however their intensity increases and multiple types of interventions are typically implemented simultaneously. The effectiveness of intensified support is monitored systematically and the interventions adjusted according to the individual needs. However, if they fail to provide sufficient support for the pupil, a *pedagogical evaluation* is conducted in multiprofessional collaboration. It can replace or complement the traditionally used psychological or medical statements, and based on it, an official decision on starting *special support* can be made according to an individual education plan. The provision of full-time special education always requires an official decision of special support. However, in the special support tier all the other means of support can also be used, only their intensity is further increased. In some cases the official decision of special support can be made without first providing general and intensified support, but this is possible only if an individual child's support needs are considered as extremely high, and it is very unlikely that the lighter means of support would suffice. If this is the situation, the child has also usually needed a lot of support in daycare and during pre-school, and there is often information available from other health care professionals who have been working with the child during the earlier years of his or her development.

To a certain extent, the Finnish three-tiered support system is comparable with the Response-to-Intervention (RTI) service delivery model in the United States (for an introduction see Burns & Ysseldyke, 2005). RTI refers to the implementation of increasingly intensive evidence-based interventions, which are designed to meet the pupils' needs, based on continuous assessment. RTI is grounded in the provision of multiprofessional consultation at each level of service (Knotek, 2005), and since its implementation it has affected the working practices of several professional groups in schools. For example, for school psychologists in the United States RTI represented a major paradigm shift from the traditional psychometric activities to collaborative planning and evaluating interventions (Powers, Hagans & Busse, 2008;

see also Sheridan & Gutkin, 2000). Even though the role of the Finnish school psychologists has never been strictly limited to testing and making special education referrals, in practice the psychologists still spend more than half of their working hours with individual pupils instead of implementing group level interventions or providing consultations. That was the situation in 2010 even though the special education strategy had been launched three years earlier (Ahtola & Vainikainen, in press), and the result most likely reflects the situation of the other pupil welfare professionals as well, for example social workers. However, since the implementation of the new support model the pressure to change existing practices has been quite hard as the focus on pupil welfare work and multiprofessional collaboration – as well as other aspects of the organisation of support – is moved from individual-centred problem-solving to prevention and school-level early interventions (Ahtola, 2012).

Since the reform has been implemented only recently, the effectiveness of the new support model has not yet been systematically evaluated (cf., Hanushek, Kain & Rivkin, 2002). There is evidence that the principles emphasised in the new model, for example prevention and early intervention, have found their way into the municipal curricula (Vainikainen, Thuneberg & Mäkelä, in press), and according to a nationally representative sample of school principals these principles are relatively well realised at the school-level too (Vainikainen, Thuneberg, Greiff & Hautamäki, submitted). Moreover, according to the official statistics which are collected yearly from all schools, the new tier of intensified support has gradually been taken into schools' practices (Lintuvuori, in press). The present study is probably the first one since the implementation of the reform to look at the effectiveness of the provided support at a child-level, which has been done by following how the differences between children who have been identified as having support needs and others develop over time. If support needs have been adequately identified, these children should perform on average lower than others in educational assessments already during early grades, but with effective support the differences should not increase significantly during the follow-up. However, as children with support needs often also have motivational problems (Thuneberg, 2007), the picture is probably not that simple. Therefore, in the present study it is also studied how

support needs are related to learning-related attitudes, task performance and interest in the assessment tasks and whether the possible *Matthew-effects* (e.g., Bast & Reitsma, 1997; Shaywitz et al., 1995) could be partially explained by them.

#### **2.1.4 Assessing educational outcomes**

Educational outcomes are in many countries evaluated and monitored centrally even if education was organised according to local curricula. Besides providing information about the performance level of pupils on a comparable scale, centralised assessment is used for securing equity of learning opportunities – both in different geographical areas or school types, and for pupils with different backgrounds. Most countries have their own strategies for evaluating educational effectiveness and equity, and only in Europe is there a wide range of approaches and a variety of traditions of practice and research in the field of assessment (The Association of Educational Assessment – Europe, 2012).

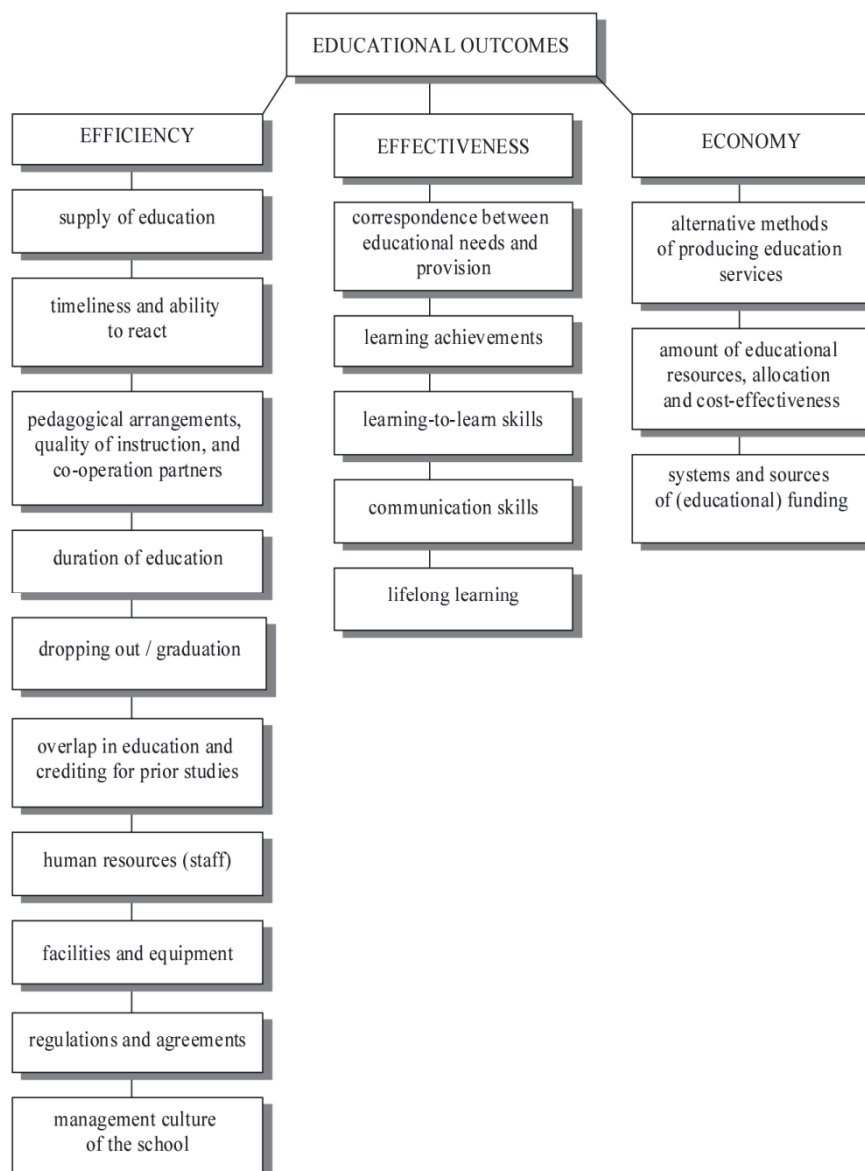
Despite the differences, the national assessment strategies have many common features. Countries often have a nationally coordinated monitoring system of pupils' knowledge of the most important curricular contents even though there are differences in which subjects the monitoring covers and how the target groups or samples are defined. In addition, the importance of more general outcomes of education and prerequisites of life-long learning – so called cross-curricular or transversal skills (see Recommendation 2006/962/EC of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning) – is often also acknowledged in the assessment strategies even though there is a clear lack of well-defined measures of them. The most influential effort to assess competencies that pupils will need in the future, the OECD's PISA, primarily measures application of knowledge acquired at school to real-life issues (OECD, 2013a). Despite an emphasis on knowledge application, most of the PISA-tasks are quite close to curricular contents except for the more general core domain of complex problem-solving implemented in PISA 2012 (OECD, 2013b) and the latest attempt to include collaborative problem-solving in PISA 2015 (OECD, 2013c). Nevertheless, there are considerable limitations to how the results of international comparative assessments can be utilised as

feedback when monitoring development or developing practices in individual schools.

In Finland, the heterogeneity of the pupil population in comprehensive school provoked, already shortly after the implementation of the new educational system, a discussion about educability (see Häyrynen & Hautamäki, 1977). In terms of educational assessment the need for developing more rigorous methods for measuring equity of education increased over the next two decades, and in particular when the school inspection system was ceased in the early 1990s. As a result, A Framework for Evaluating Educational Outcomes in Finland was published in 1995 and in a revised form in 1998 (National Board of Education, 1999, English translation). It divided the outcomes of education into three categories: efficiency, effectiveness and economy. Efficiency referred to the functioning of the educational system, effectiveness in pupil-level outcomes and economy to the successful allocation of resources. The conceptualisations of each category are presented in Figure 2.1. From the perspective of the present study, the conceptualisation of effectiveness is of particular interest as it is directly related to pupil-level measures of competences and attitudes.

The definition of the indicators of effectiveness presented in Figure 2.1. led to two kinds of practical applications. As the first and the most central means of educational assessment, sample-based national assessments were introduced to the key school subjects. However, unlike in many other countries, even in 2014 these assessments are not repeated each year at pre-defined grade levels. Instead, the school subjects and the grade levels to be assessed are defined in a four-year plan for educational assessment (see Ministry of Education, 2012, for the current plan). Typically there are two to three school subjects to be assessed, and a sample of about 5000 pupils participates in each test. The information provided by these assessments of curricular contents is being complemented by international assessments and national thematic assessments, of which learning to learn has been in the evaluation model since the beginning.





**Figure 2.1** The definition of educational outcomes in the Framework for Evaluating Educational Outcomes in Finland (National Board of Education, 1999).  
Figure reproduced with the permission of the National Board of Education.

As shown in Figure 2.1., learning to learn was defined as one of the targets of educational assessment in Finland already in the mid-1990s. As a result, the development of the Finnish learning to learn scales started in 1995, and even though they did not receive the same position in the

national assessment plans as subject-based assessments, several representative assessment studies were conducted in the sixth and ninth grade and in upper secondary education at the turn of the millennium (Hautamäki et al., 1999; 2000; 2002b; 2003; 2005) and again in 2012 after a decade's break.

Recently, the discussion about educability, and assessment and intervention of learning to learn skills has once again become topical. This is partly due to a significant decrease in 15 years-old pupils' performance level in large-scale cross-sectional learning to learn assessments both at the municipal and national level (Kupiainen, Marjanen, Vainikainen & Hautamäki, 2011; Hautamäki et al., 2013), and is especially so as the latest PISA-results show that the phenomenon applies to other areas of assessment too. In addition, due to the recent changes in educational legislation and the pupil support model in 2008 – 2011 (Thuneberg et al., 2013), pupils with very high special education needs are increasingly being taught in local schools; most of them in regular classes with individualised support (see Sabel et al., 2011 for an introduction to the service model). The combination of the increasing heterogeneity of school classes and the weakened position of formal schooling in young people's lives – which has been suggested as an explanation for the decreasing results – makes systematic assessment of cross-curricular skills even more important. Furthermore, system-level assessments will also in the future have to be oriented towards developing practices instead of ranking schools in order to secure that every pupil, regardless of their background, gets equal possibilities for obtaining the basic and transversal skills that are necessary for life-long learning in the changing world.

## **2.2 Development of learning to learn skills during primary education**

### **2.2.1 Introduction of the Finnish LTL model**

As an attempt to evaluate education and its role in creating and maintaining educability, a Finnish model for assessing pupils' learning-to-learn skills was created in 1996 (see Hautamäki et al. 2002a; 2006;

Hautamäki & Kupiainen, 2014). It was developed further during an intensive period of the following seven years when nationally representative large-scale assessment studies (Hautamäki et al., 1999; 2000; 2002b; 2003; 2005) were conducted as a part of the Finnish national strategy for educational assessment. The scales also formed a substantial part of the European learning to learn instrument that was built and piloted in eight countries as a collaboration between the European Commission and the member states (Kupiainen, Hautamäki & Rantanen, 2008; Hoskins & Fredriksson, 2008). Since the method was designed to be used as a means for assessing effectiveness of education – how pupils have at the end of each school level acquired cross-curricular skills they will need in future learning – the assessment tool was not originally built to be diagnostic at an individual level. However, recently the focus has been shifting towards the use of assessment results in developing classroom practices, which are also to meet the needs of the assessed individual pupils to enhance their preparedness for life-long learning. In order to evaluate the predictive validity of the assessment tools and to gain a deeper understanding of the development of learning to learn skills in comprehensive schools, large-scale longitudinal studies have been implemented in collaboration with some of the largest municipalities in Finland (e.g. Kupiainen et al., 2011; Vainikainen, Marjanen, Kupiainen, Gustavson & Hautamäki, 2011).

In the Finnish model, learning to learn is defined as cognitive competences and attitudes and beliefs that support the effective use of them (Hautamäki, Hautamäki & Kupiainen, 2010; Hautamäki & Kupiainen, in press; Hautamäki et al., 2002). Learning to learn is assessed by paper-and-pencil or computer-based group tests that are comprised of cognitive tasks and self-report questionnaires. The attitude scales derive from several different theoretical origins, and the theories that are relevant for the present study are presented later in this chapter.

The cognitive component of learning to learn is measured by tasks that are related to curricular contents, but they require the application of higher-order thinking skills instead of repeating things learned in school subjects. The cognitive competences assessed by the current version of the Finnish LTL scales cover reading comprehension, mathematical thinking skills and more general thinking and reasoning skills. The

theoretical rationale for selecting these competence areas is presented in the next section of this chapter, and the more detailed descriptions of the tasks are found in Chapter 3. Here it is enough to mention that also reading comprehension, which of the areas covered by the tasks is probably closest to the contents of the curriculum, is understood as a higher-order skill: Rather than repeating the contents of texts, the children are expected to understand the main ideas and hierarchically rate facts taken from the texts within the theoretical framework of Kintsch and van Dijk (1978). Another thing worth mentioning here is that, theoretically, the cognitive domain has also comprised problem-solving since the beginning of the development of the scales (Hautamäki et al., 2002), but despite trying out several task types, a permanent solution for their large-scale assessment is still under development. Recently the Finnish longitudinal samples have also been assessed with the MicroDYN tasks for complex problem-solving (Greiff, Wüstenberg & Funke, 2012; Greiff & al., 2013), but the results will not be discussed in this study.

The Finnish conceptualisation of learning to learn is not the only one, and there are different views of how broad the definition should be and to what extent it should cover cognitive competences, beliefs and attitudes, metacognition, learning strategies etc. (e.g. Csapó, 2007; Deakin Crick, 2007; Demetriou, Spanoudis & Mouyi, 2011; Hoskins & Fredriksson, 2008; Moreno & Martín, 2007). The theoretical origins of the Finnish learning to learn model – the understanding of learning as a measurable outcome of more general but modifiable cognitive competences and attitudes that support the use of them - lie in Snow's views of aptitude development and education (see Hautamäki & Kupiainen, 2014, for a more detailed theoretical description of the Finnish model). According to Snow (1996, p.537), “aptitude is an outcome of past educational steps as well as an input to future educational steps”, and he sees “aptitude development” as the most important product of education all along the way. Snow's views of the role of education in enhancing cognitive competences and the affective factors related to it, are presented in a separate section of this chapter. However, before that, it is necessary to take a closer look at the development of cognitive competences in general.

### 2.2.2 Cognitive development during primary education

Demetriou, Spanoudis and Mouyi proposed in 2011 an integrated theory of the developing mind based on findings and concepts from intelligence research, the psychology of cognitive development and cognitive psychology. This theory was selected as the most central theoretical framework of the present study as it – while acknowledging that there are individual differences in children's cognitive competences – has a strong developmental perspective, and it stresses the role of education in enhancing the effective use of the developing competences. It also emphasises the role of consciousness in regulating learning processes which can be equated with the understanding of the role of beliefs and attitudes in the Finnish learning to learn framework. Learning and educational outcomes are clearly not predetermined by biological differences between children, - genetic heritability accounts for only half of the variability in the cognitive abilities that comprise intelligence (Petrill, 1997) – so understanding cognitive development is crucial when trying to develop education that is even more beneficial for all children. On the other hand, the Piagetian developmental view alone cannot explain all the variation between children, and since the emphasis of the present study is partly on children with support needs of different intensity, individual differences need to be taken into account.

Demetriou's model (Demetriou et al., 2011; see also Adey, Csapó, Demetriou, Hautamäki & Shayer, 2007) involves both central and general mechanisms and specialised capacity systems for different domains of knowledge or relations. More specifically, these *specialised capacity* or *structural systems* are coordinated by the *representational capacity system* which interacts with the inference system, and all these systems are monitored and regulated by the *consciousness system*.

The *specialised structural systems* refer to core processes, mental operations and knowledge and beliefs. The spatial, verbal, quantitative, categorical, causal and social reasoning systems have been identified by methods from different theoretical origins, and they are considered as autonomous domains of understanding and problem solving. They may develop at different rates, but they are constrained by the development of the other, higher-level systems (Adey et al., 2007). They can also be trained by means of interventions or more generally through education

(Demetriou et al., 2011). When assessing learning to learn within the Finnish framework, the cognitive component aims at addressing the higher-order skills in the contexts of most of these specialised systems: On the one hand, to get a richer picture of the developmental level and individual differences of the children, and on the other to give the children an opportunity to make up for difficulties in one area with better performances in other areas. This is particularly important when educating and assessing children with different kinds of support needs.

The *inference system* is responsible for connecting and integrating information and operations according to the selected goal. It enables the transfer of meaning from one representation to another based on properties which are typically common for the source and target. Demetriou and colleagues (2011) review studies which show that inductive, analogical and deductive reasoning are based on different inferential mechanisms and they also develop in separate but overlapping waves. Some form of inductive reasoning is present already from birth, and it develops in three main stages from the ages of 6 to 12. In the first stage, children learn to identify patterns or make generalisations based on a single dimension, while in the second stage information can be partly hidden or implied. In the third stage, inductive reasoning is based on theoretical suppositions (Demetriou et al., 2011).

Analogical reasoning means applying the rule learned from one representation to another one. According to Demetriou and colleagues (2011), it can later structure, as a continuum of the development of inductive reasoning, third- and higher-order relationships involving abstract relations which require also cultural knowledge.

Deductive reasoning – making conclusions based on given premises – begins to appear when representations are differentiated and expressed by means of natural language. It is also associated with awareness of cognitive processes and control. Demetriou and colleagues (2011) present evidence that this awareness begins to appear at about the age of five or six, but it takes years before the logic becomes explicit. Later in adolescence young people can handle arguments that are not determinate and specify all implications of an argument.

Many of the Finnish learning to learn cognitive tasks measure the functioning of the inference system, and whereas tasks directed to first

graders measure analogical and also inductive reasoning mainly with spatial and categorical contents, later on there is greater emphasis on deductive reasoning, Piagetian formal thinking and problem-solving.

The *consciousness system* refers to monitoring processes for ensuring the awareness of the goal, evaluative functions for comparing the present state with the goal and control functions for correcting actions. It is also the link between the mind and the personality (Demetriou et al., 2011). The consciousness system covers concepts as metacognition, learning strategies, reflection, self-evaluation and self-awareness, which are also partially addressed by the attitude scales of the Finnish learning to learn assessment method. However, from the point of view of measurement and assessment, it is clearly the most difficult area to measure. The consciousness system develops throughout the whole of childhood and adolescence: Demetriou and Kazi (2006) have shown that at the age of seven most children were aware which mental operation they were applying, and at the age of 14 their self-representations (general self-concept related to the assessed domains) began to be accurate. Self-evaluations began to become more accurate gradually from the age of 11 (cf. Harter, 1999). Demetriou and others (2011) interpret these and other findings as suggesting that self-awareness and self-evaluation of cognitive processes develop in cycles, and in the beginning of the next developmental stage they become more inaccurate again when the stage-specific problem-solving operations and skills become more demanding.

The consciousness system also controls the functioning of the *representational capacity*. It includes modality-specific components for holding information for short periods of time, short-term storage which is available in the modality-specific components for further processing and an executive component. This representational capacity is a more elaborated version of the traditional understanding of working memory, and it comprises two domain-specific systems, the visuospatial sketchpad and the phonological loop. In addition, it has a central executive that regulates the functioning of them (Baddeley & Hitch, 1974), and it is also the core of the information processing models of intelligence (Adey et al., 2007). Demetriou and colleagues (2011) relate the modality-specific components to the specialised structural systems and show that there can be cultural differences regarding how they develop as a result of

education. They also refer to studies pointing out that the short-term memory span for a given type of information is a function of the maximum capacity of the storage and cognitive load which limits the capacity. In general, there is a lot of research about working memory and executive control which support the idea of cognitive competences being modifiable while simultaneously accounting for individual differences (see Adey et. al., 2007). Also the Finnish learning to learn scales comprise tasks for working memory and executive functions for younger pupils, but during later school years they have received less attention in the assessment.

It has been shown that visuo-spatial and phonological short-term memories are two separable cognitive processes which operate somewhat independently from one another (Shah & Miyake, 1996). Halpern (2000) sees this as an especially useful distinction from the perspective of understanding gender differences. Her indirect conclusions can be interpreted as an implication that the evidence she presents regarding girls' superiority in verbal skills and boys' better spatial and quantitative understanding can partly be explained by differences in the functioning of working memory. There are indeed well-documented gender differences in basic brain functions that are related to the evolution of human beings (Gazzaniga, Ivry & Mangun, 1998) which should not be completely forgotten about when making conclusions about the effectiveness of education from the perspective of educational equity. Therefore, the development of gender differences in cognitive abilities needs to be understood before claiming that one or another gender seems to benefit more from formal education.

### ***Gender differences in cognitive competences and achievement***

In the probably most extensive literature review on gender differences in cognitive competences, Halpern (2000) poses a question that is highly relevant for the present study: "How can we ever be certain that what we are labeling sex differences in ability aren't really sex differences in achievement?". Her conclusion is that it will never be possible due to the blurry distinction between ability and achievement and the ways of measuring them. The starting point of the present study is that for some



reason school seems to produce gender differences in achievement, which are even contrary to what possible differences in underlying abilities would suggest, and one of the aims of the present study is to shed light on the mechanisms of the development of these differences in achievement. Cross-sectional studies cannot answer the question of *when* these achievement differences begin to develop and whether there are systematic school, class or peer group effects which could partially explain *why* they happen.

Girls and boys are not different regarding their general cognitive competence (Halpern, 2000), that is, there should not be crucial differences in the functioning of representational capacity and inference system in general. However, there seems to be differences in the specialised structural systems as the evidence from ability studies point systematically to the direction that females exceed males in verbal tasks whereas males perform better in tasks with quantitative and spatial contents (Halpern, 2000). This applies to reasoning as well as to working memory. Halpern reviews biological and psychosocial theories for explaining these findings, and she concludes that while there seems to be both genetic and hormonal differences which may have their origins in evolution (Gazzaniga et al., 1998), also family, peers and broader society – including formal schooling system – have an influence on how gender differences in achievement finally develop. This view is also supported by sociological research on the socialisation of gender roles in school, for instance regarding the connection between academic success and peer acceptance, which seems to be stronger for girls than for boys (Adler, Kless & Adler, 1992). Because of all this, only studying cognitive competences is never sufficient if individual and group-level differences in achievement are to be understood.

It is not a new phenomenon that girls get better school grades also in areas that are traditionally considered as boys' strengths (Kenney-Benson et al., 2006; Kimball, 1989; Wentzel, 1988). Whether or not this tells about real differences in achievement is another question as school grades seem not to reflect only competences or achievement, but they may be affected by effort, attitudes or prosociality (Kenney-Benson et al., 2006; Kupiainen et al., 2014; Wentzell and Caldwell, 1997). Earlier analyses of the Vantaa sixth grade data used in the present study show

that girls got better grades in Finnish language even when their reading comprehension skills were controlled for, while there was no systematic gender bias in mathematic grades (Krkovic, Greiff, Kupiainen, Vainikainen & Hautamäki, 2014). The same study also confirmed that teachers' gender did not have an interaction with gender differences in performance or grades, even though there is relatively much international literature that suggests teachers may treat girls and boys differently (see Jones & Dindia, 2004). Kenney-Benson and colleagues did not look at teacher effects or at different school subjects separately, but they concluded that gender difference in learning strategies accounted for girls' edge over boys in terms of grades. Learning strategies in turn were predicted by holding mastery over performance goals and by refraining from disruptive classroom behaviour, which were both more typical of girls.

Especially lately, girls have also systematically outperformed boys in external assessments of achievement and skills in areas that are not directly related to curriculum contents (e.g. Hautamäki et al., 2013; OECD, 2013a). While in Finland this phenomenon has been visible for a longer time (Hautamäki et al., 1999; 2000; 2002b; 2003; 2005), internationally the change can be dated back to the mid-2000s when girls closed the gap with boys in terms of achievement in external mathematics assessments (Kenney-Benson et al., 2006). At the same time, the overall results began to decline (Hautamäki et al., 2013), so it is more likely that boys' results decreased relatively more than girls' and not that girls have improved their performance dramatically. In fact, this is clearly visible in the latest report of the Finnish PISA-results (Kupari et al., 2013), even though there is also some evidence from learning to learn studies that suggests that the average-performing Finnish girls are not doing as well as they used to do either (Kupiainen et al., 2011).

The focus of the present study is partially on how the development of the cognitive competences of low achievers and children who need support in their studies proceed compared to their agemates without support needs. Here understanding gender differences is particularly important as boys have always been classified as learning disabled, or having socio-emotional problems or other support needs much more often than girls (Henning-Stout & Conoley, 1992; Thuneberg, 2007).

Halpern (2000) reminds that some of these differences are developmental, and boys who mature slower can later catch up with their peers if they are not allowed to fall too far behind. Therefore, boys should benefit even more than girls from the new Finnish support model which emphasises early intervention and general and intensified support (Thuneberg et al., 2013). This, however, cannot yet be answered with the present data, but one indicator of effectiveness of the new support model in the future could be that of stopping the increase of gender differences in achievement. The same of course applies to other background factors as well: Children of parents with lower socio-economic background perform on average lower than those coming from more advantaged homes (Willms, 2010), and one aim of any support system should be not to let these differences increase further.

### ***The role of education in enhancing cognitive competences***

Adey, Csapó, Demetriou, Hautamäki and Shayer (2007) explored the nature of general cognitive ability, showing that despite it being general it is also modifiable by the means of education and intervention. Similarly, Snow stated already 11 years earlier (1996) that education is an aptitude development programme, and intelligence is one of the most important aptitudes to be developed. More recently, Demetriou, Spanoudis and Mouyi (2011) included in their theory of cognitive organisation and development a guideline for educating pupils to gain capacities in effective use and to enhance reasoning, thinking and learning to learn skills, including also self-awareness. Accordingly, one of the most central underlying assumptions in the Finnish learning to learn model is that both the general thinking and reasoning skills, and the measured learning-related attitudes, are partially outcomes of education, and they can be enhanced with both high-quality basic education and specific interventions (e.g., Kuusela, 2000).

Even if Snow's (1996) terminology differs from the concepts used in the present study, he provides a useful distinction between aptitude and intelligence. He defines aptitude as a much wider concept, which includes intelligence as a modifiable subset of aptitudes for learning and problem-solving, particularly in situations involving novel or complex, meaningful

information and incomplete instruction regarding it (cf. Demetriou et al., 2011). Aptitude, on the other hand, is an outcome of past educational steps as well as an input to future educational steps, including all relatively stable cognitive, conative and affective characteristics of persons needed for success in learning performance. According to Snow (1996), this aptitude for new learning should increasingly become the principal goal of education, covering subgoals such as “learning to learn, learning to reason, learning to find and solve problems, learning to be interested and industrious, to persevere, to achieve in the face of novelty, complexity, adversity and change”.

Even though Snow emphasises the importance of enhancing all childrens’ cognitive competences by means of educational interventions for learning and thinking skills and strategies, he discusses two reasons for why individual differences should not be underestimated. Firstly, individual differences can moderate the effects of interventions, and that needs to be taken into account. Secondly, there can be different sources of variance when any competence is measured, so the same recipe does not necessarily work for everyone. In the language of the new Finnish support model, the intensity of the interventions have to be adjusted individually, and if a pupil is not responding to a basic level intervention, it needs to be intensified both in quantity and quality (cf. Thuneberg et al., 2013). The present study does not provide detailed information about any specific means for enhancing learning and thinking skills, but it adds to the knowledge about the sources of individual differences when working with children with identified support needs in school settings. Therefore, the results are relevant when designing practices and policies for enhancing the development of children’s cognitive competences during the first six grades of basic education.

When talking about either assessment or intervention, it is not enough to concentrate only on pupils’ cognitive competences. Affective factors and metacognition play also a central role in the learning process and they should be targets of instruction as well (Demetriou et al., 2011; Snow, 1996). According to Snow (1996), *curriculum effects* (the directly school subject-related processes) are only one thing producing intelligence development, and when designing interventions more emphasis should be on *metacurriculum effects* which can be categorised

into six main groups: 1. perceptual and memory skills and strategies, 2. thinking and reasoning skills and strategies, 3. self-regulation, 4. beliefs and values about learning and thinking, 5. learning to learn from incomplete instruction, and 6. flexible adaptation of knowledge, skills and strategies. Of these, the first two and the last two have been covered at least superficially above, whereas the third and fourth are more closely related to the other domain of the Finnish learning to learn definition, the learning-related attitudes.

### **2.2.3 Learning-related attitudes**

The effects of learning-related attitudes – e.g. motivation, engagement, causality beliefs, and academic self-concepts – on achievement have been studied extensively over the years (see Eccles & Wigfield, 2002 for a review of different theories of motivational beliefs, values and goals in educational contexts). The results indicate that attitudes do play a role in explaining variation of achievement even though their explanatory power has not been very strong in the relatively few studies in which the effects of prior ability or performance have been controlled for (e.g. Aunola, Leskinen & Nurmi, 2006; Gagné & St Père, 2002; Steinmayr & Spinath, 2009). However, there is evidence even from the neurosciences that beliefs can influence learning at the brain level (Mangels et al., 2006), and they also form a central part of the Finnish learning to learn concept. It is to be noted, however, that unlike in some other definitions of learning to learn, the Finnish model stresses the importance of cognitive competences, and the role of attitudes is to secure the effective use of them in learning situations (see Hautamäki et al., 2006).

From the perspective of the present study, the most central theory about learning-related attitudes is the achievement goal theory (see Harackiewicz, Barron, Pintrich, Elliot & Thrash, 2002, for a revision of it). According to the theory pupils can have mastery goals which value learning and understanding things as ends in themselves, and performance goals which are related to performing better than others or trying to avoid looking less able than them. Depending on the definition, goals can also be categorised as intrinsic and extrinsic (cf. motivation theory of Ryan & Deci, 2000, for instance), or approach- and avoidance-oriented (see Dweck, Chiu, & Hong, 1995; Harackiewicz et al., 2002).

Generally, mastery goals have been associated with better educational outcomes (e.g. Kenney-Benson et al., 2006) and engagement (e.g., Tuominen-Soini, Salmela-Aro & Niemivirta, 2012), while performance oriented goals are expected to affect outcomes negatively. However, there is some evidence about that performance-approach goals (trying to look better than others) are not necessarily bad for motivation or achievement (see Eccles & Wigfield, 2002, for a review). Since the purpose of the present study is to understand the reasons for the differences in performance in an assessment situation, and not to contribute to the field of achievement goal theory, performance goals will not be elaborated on further in this study.

Traditionally, mastery goals have covered only goals that value learning as an end in itself (cf. intrinsic motivation), and extrinsic goals (e.g. getting good grades) have been considered as being even harmful for subsequent interest, effort and performance (Patrick, Ryan and Pintrich, 1999). However, earlier learning to learn studies have shown that in Finland also extrinsic goals are clearly related to better performance (Hautamäki et al., 1999; 2000; 2002b; 2003; 2005; see also Tuominen-Soini et al., 2012), and Patrick, Ryan and Pintrich (1999) argue too that extrinsic reasons for doing schoolwork may be better than having no reasons at all. Therefore, in this study the concept of mastery attitudes refers to attitudes relating to both intrinsic and extrinsic mastery goals, pupils' evaluation of the importance of school in general (the last two substudies) and agency beliefs regarding own effort, coming from the theoretical background of action-control beliefs.

Action-control beliefs theory (e.g. Little, Lopez, Oettingen & Baltes, 2001) is another central attitude theory in the Finnish learning to learn framework (Hautamäki et al., 2002). The first component of action-control beliefs, Means-ends beliefs, refers to children's generalised thoughts about the causal power of effort, ability, luck, teachers and other reasons in producing school outcomes. The second component, Agency beliefs, refers to children's beliefs about how much they personally possess or have access to the means of effort, ability, luck and teachers. Control expectancy is similar to the concept of self-efficacy, and it refers to children's expectations of being personally able to produce a desired learning-related outcome without specifying means for it (Little et al.,

2001). In earlier Finnish learning to learn studies (Hautamäki & al., 1999; 2000; 2002b; 2003; 2005), agency and means-ends beliefs regarding effort, and control expectancy, have been positively related to performance while believing in the role of ability or luck in producing educational outcomes has been negatively related to performance.

In the last two substudies of the present study also the role of detrimental attitudes in explaining performance in the assessment is examined. Even though the Finnish LTL test has scales for performance-approach and performance-avoidance goals, they will not be used in the present study due to their weak connection to actual performance in earlier learning to learn studies and to the controversial evidence from international studies (see Eccles & Wigfield, 2002). Therefore, the roots of the detrimental attitudes concept used in this study lie mainly in the action-control theory – in pupils' conceptions of the role of ability and luck in explaining educational outcomes. The only achievement goal - related detrimental attitude concept used in this study is that of self-handicapping strategies (Midgley & Urdan, 2001). Self-handicapping refers to giving up easily, postponing important tasks and not putting the best effort into tasks in order to have an explanation for a poor result. Several studies have found that handicapping is associated with a lower achievement level, and boys have been found to use handicapping strategies more often than girls (Midgley & Urdan, 2001).

### ***Gender differences in learning-related attitudes***

Probably the most comprehensive theoretical model designed to explain gender differences in academic achievement was proposed in 1983 by Eccles (see Eccles, 2011). It is a variation of a more general Expectancy X Value model, and it is based on the idea that the outcome of a cognitive task depends on how much the individual doing the task expects to succeed or fail and how much she or he values the outcome. Eccles (2011) claims that later educational and occupational gender differences depend partly on different choices which are in turn partly explained by gender differences in self-concepts, expectations of the social environment and subjective task values. The central role of self-concepts in the theory makes it interesting as a complementary piece of Demetriou and colleagues' theory of the architecture of the human mind (2011) and

addresses possible causes of gender differences in the consciousness system. The emphasis on subjective task values, on the other hand, is of interest when trying to understand task-specific or situational factors in explaining performance in an assessment situation. This idea will be elaborated on further in the separate section about performance in an assessment situation.

The mechanisms of the development of gender differences in learning-related attitudes are only partially understood. Kenney-Benson and colleagues (2006) suggest that social and biological forces could cause girls and boys to approach schoolwork differently (see also Adler et al., 1992), and as a result they would develop different kinds of goal orientations and learning strategies. The authors review literature about parents treating girls and boys differently in regard to mathematics, by encouraging girls to rely on hard work while believing in boys' abilities. Eccles (2011) complements this interpretation, which is based on literature and her own studies since the 1980s, by discussing more broadly the effects of gender-role ideologies and stereotypes which guide girls to read more and interact with peers while encouraging boys to concentrate on mathematical areas. Kenney-Benson and colleagues believe that parents' gendered attributions for success together with biological differences in activity levels would make girls develop more mastery-oriented goals while boys' goals would be more performance-oriented which would then affect their achievement. The empirical evidence for this gender difference in goal orientations is, however, controversial (Patrick et al., 1999) even though Kenney-Benson and colleagues' own data supported their claims.

Even though empirical evidence does not lead too straightforward conclusions about girls' attitudes being generally more beneficial for learning than boys', it is also obvious that boys demonstrate more socio-emotional problems and disruptive behaviour at school (Halpern, 2000; Thuneberg, 2007). As antisocial behavioural patterns that are already demonstrated during primary education are related to poorer outcomes later in life measured by several different indicators especially for males (Huesmann, Dubow & Boxer, 2009; Olweus, 1979), it is important to study how detrimental attitudes develop and what could be done for preventing them within the possibilities that schools have in influencing



young people's lives. Thuneberg (2007) showed in a study about Finnish pupils' psychological well-being, motivation and school achievement that in the cluster of "unmotivated low-achievers" there were many more boys than girls and more pupils with special education needs than others. When understanding motivation and learning-related attitudes also as socially developing phenomena, research results indicate strongly that system-level or choice-related segregation – based on support needs, gender or socio-economical background – are not recommendable (cf., Willms, 2010), and the Finnish legislative changes emphasising local schools and inclusion are the right way to go.

Since motivation and engagement in school do not develop only at an individual level, in a school context it is particularly interesting to try to understand how pupils influence each others' attitudes. However, there is much more research on teachers' (see Jones & Dindia, 2004, for a review) and parents' roles (see Kenney-Benson et al. 2006, for a review) as socialising agents of motivation and engagement than there are studies on peer influences on them (Ryan, 2000). Even though the first studies providing evidence for school-based peer groups' influence on the development of engagement in schoolwork and school grades were published about 20 years ago (e.g. Berndt & Keefe, 1995), in the field of motivation research many studies reporting about the importance of peer support for school engagement and attitudes still rely on pupils' perceptions of their peers' behaviour instead of using data from peers (e.g. Wang & Eccles, 2012; 2013). Therefore, it is useful to take a look at peer influence literature when searching for more appropriate methods in evaluating classmates' and peer groups' roles in the development of both the cognitive competences and learning-related attitudes which together form the learning to learn skills examined in this study.

#### **2.2.4 Peer influences on learning and learning-related attitudes**

Since Piaget's and Vygotsky's times it has been understood that much of children's learning occurs in social contexts. There is several decades of evidence that peers can have an important role as supporters – or distracters – of learning (eg. Kindermann, 2007; Ryan, 2000; Song & Grabowski, 2006; Wang & Eccles, 2012; 2013; Wentzell and Caldwell,

1997; Wentzel, Filisetti & Looney, 2007), but it has remained under discussion whether the results of mainly cross-sectional studies tell about similar learners seeking each other's company or pupils becoming more like their peers when spending time together. In early studies about peer effects on school success it has already been speculated that school achievement could play a role in what kind of a social group a child belongs to (see Brown and Lohr, 1987). On the other hand, there has been evidence that belonging to a specific group could affect school success if peers are supportive and school-oriented (see Wentzell and Caldwell, 1997). Therefore, when trying to understand the development of learning outcomes especially during early adolescence, it is important to look at the influence pupils can have on each others' learning in school contexts. During the past 10 years a lot of progress have been made in order to understand the mechanisms of peer influence on academic adjustment (Ryan, 2012), but there is still a need for more comprehensive longitudinal research designs and rigorous analysis methods.

There is a long history of research about the importance of children's social goals in their peer relations (e.g. Crick & Dodge, 1994), and lately more attention has been drawn to their effects on academic achievement (Ryan, Jamison, Shin & Thompson, 2012). Social goals are related to group norms which can be very different in different peer groups. These norms can support engagement in learning activities (Hamm, Hoffman & Farmer, 2012), but they can also encourage behaviour that distracts learning or prevents pupils from performing at their own level (see Wentzel, Donlan & Morrison, 2012, for a review). For instance, Wentzel, Filisetti and Looney (2007) found that preadolescents who viewed their friends as having high academic goals behaved in ways that helped promote their own academic achievement, and the same may be true the other way around.

There is also longitudinal evidence that a peer group's level of school engagement predicts changes in children's motivation across time (Kindermann, 2007). The effect in the study of Kindermann was not very strong, but evidence for group influences persisted even when controlling for peer selection and the influence of teacher and parent involvement. Of particular interest regarding the present study is that 11 to 13 years-old

girls' engagement remained relatively stable whereas boys' decreased during the follow-up period, and differences in trajectories coexisted with similarities in processes of how boys and girls selected group members and how they were influenced by their groups (Kindermann, 2007).

Even though pupils' learning – and also participating in educational assessment studies – obviously does not happen in isolation and is affected by school, class and peer group level factors, educational assessment studies often fail to address class- or peer group- level variation that is necessary to understand for utilising the assessment results in developing school practices and designing interventions for enhancing performance. Peer influences on educational outcomes have been studied extensively by economists and social policy analysts with large-scale assessment data (e.g. Hanushek, Kain, Markman & Rivkin, 2003; Harris, 2010; Willms, 2010; Zimmer & Toma, 2000), but they typically equate peer groups with school or neighbour populations and look mainly at the effects of socio-economic background-related variables on school-level outcomes. They often also use aggregate-level data for describing school or at best classroom composition (e.g. Zimmer & Toma, 2000), which can lead to severe overestimations of background variables' effects on performance (Kuusela, 2010). Anyhow, research results from those strands are extremely important in system-level developmental work as they also generally speak against segregation by providing evidence that especially lower-achieving pupils benefit from higher-achieving schoolmates (e.g. Hanushek et al., 2003; Zimmer & Toma, 2000). Accordingly, the concentration of socio-economically disadvantaged pupils in some schools has led to decreased levels in their performance compared to similar pupils in schools with higher average socio-economical status (Willms, 2010).

The economists' perspective and evidence are however not sufficient when designing interventions for enhancing performance or attitudes of individual pupils or small groups of them as pupils are individuals who form naturally differing peer groups. Educational psychologists often study how self-nominated peers or observed peer groups influence each others' learning (e.g. Kindermann, 2007; Ladd, Kochenderfer-Ladd, Visconti, K.J. & Ettekal, 2012; Ryan, 2012; 2000, Wentzel & Caldwell, 1997; Wentzel, Donlan & Morrison, 2012), but their methods are typically

not applicable in large-scale assessments. In addition, the results of the studies are usually applied only in individual-centred contexts, without paying much attention to what could be done at a school-, municipal or even national level to enhance equity of education and raise the performance level of all pupils. Therefore, combining the advantages of two different approaches is fruitful when trying to understand what kind of group-level factors can influence pupils' performance in external assessments. It needs to be understood whether the well-documented decline in the assessment results depends partly on group-level phenomena which could be targeted with interventions. One of the basic assumptions behind the present study is that the change cannot happen only at an individual level, but schools, classes and peer groups play a role in how pupils are willing to perform in low stakes assessments.

## **2.3 Performance in a low-stakes assessment situation**

Performance in an assessment situation does not depend only on cognitive competences, knowledge, more stable belief and attitude structures, or systematic long-term effects produced by schools, classes and peer groups. There are also situational factors, which influence the outcomes of any assessment, and they do not all necessarily depend on the things that have been so far covered in this chapter, even though they are related to them. As Eccles (2011) puts it,

“Participating in a particular task requires the demonstration of the characteristics associated with the task, and whether this requirement is seen as an opportunity or a burden depends on the individual's needs, explicit and implicit motives and personal values, and on the individual's desire to demonstrate these characteristics both to herself and to others.”

Or, as expressed in the words of the Finnish learning to learn framework (Hautamäki & Kupiainen, 2014),

“[pupils] are invited to accept the tasks as their own with all the motivational, goal and aptitude-related conditions attached, and the processes of learning to learn are set in motion in this acceptance. Regardless of the knowledge or skill level of the pupil, the acceptance of the assessment task (or the refusing of it) activates processes that either enhance or hinder flexible intellectual work.”

In the present study it is hypothesised that this task acceptance is not only an outcome of personal, more stable characteristics or group phenomena, but the task or situation itself can be more or less appealing to some pupils or pupil subgroups. In addition, pupils' actual effort in the assessment situation can be constrained by factors, which have little to do with the pupils themselves but are related to the environment in which the assessment is conducted. Therefore, two more research strands related to interest and task performance need to be introduced here.

### **2.3.1 Task interest**

Task interest has been shown to be related to learning and achievement outcomes, at least indirectly, even when the effect of prior ability has been controlled for (Ainley, Hidi & Berndorff, 2002a; Van Yperen, 2003). In other studies, which have often been conducted with older participants than the 7 to 12 years-olds of the present study, task interest has typically accounted for approximately 10 % of the variance of performance. When controlling for initial individual differences, Ainley and colleagues (2002a) showed that the mechanism of the influence was more complex, with interest being related to affective response, the affective response to persistence, and, finally, the persistence to learning outcomes. Nevertheless, all the evidence points in the direction that interest plays a role in explaining performance, and being interested in a task depends only partly on more stable personal characteristics such as achievement goals (for the relationships between achievement goals and task interest, see Hulleman, Durik, Schweigert & Harackiewicz, 2008; Tapola, Veermans & Niemivirta, 2013; Van Yperen, 2003).

Hidi and Renninger (2006) suggest that interest develops in four stages. The first stage is triggered situational interest which can evolve into maintained situational interest. Emerging individual interest can then develop out of the second stage, and finally it can lead to a well-developed individual interest (Hidi & Renninger, 2006). School-age pupils' more long-term interest for school subjects has been shown to decrease when they get older, but it can also develop positively if assignments and the learning environment support it (Renninger & Hidi, 2011). Therefore, it is interesting to see how the individual or subgroup-level differences in these changes in task interest – and in more general

learning-related attitudes – are related to actual performance. As pupils with a more developed individual interest have been shown to have better possibilities to experience related situational interest (Renninger & Hidi, 2011), in a longitudinal perspective this could mean that also task interest is cumulative. It is, therefore, likely that in different pupil subgroups both situational task interest and the underlying general learning-related attitudes develop differently, both because of differences in cognitive competences and peer influences on attitude development.

Situational sources of interest are particularly important when dealing with pupils who do not have prior individual interest in school activities (Ainley et al., 2002a). The definition of task acceptance in the Finnish learning to learn model can be seen to be related to this; novel tasks are expected to trigger situational interest also in pupils who may be less motivated in their normal school work – even if task acceptance is primarily understood to derive from “all the motivational, goal and aptitude-related conditions” (Hautamäki & Kupiainen, 2014), which are not directly depending on the situation. Based on task interest literature (see Renninger & Hidi, 2011, for a review) it is likely that interest and more general learning related attitudes as achievement goals (Hulleman et al., 2008; Van Yperen, 2003), which form a central part of the attitudes covered by the present study, develop hand in hand, and the changes in them are also interdependent. Hulleman and colleagues (2008) proposed a model in which mastery goals enhance subjective task values (cf. Eccles, 2011), which in turn can lead to subsequent interest. Therefore, they suggest an indirect effect of mastery goals on performance through interest and effort.

### ***Gender differences in task interest***

Since the emphasis of the present study is to find out why girls perform better than boys in assessments even if there are no differences in their competences, it is evaluated whether situational task interest could provide a partial explanation to this concern-raising phenomenon. In their study on gender differences in response to literary texts Ainley, Hillman and Hidi (2002b) review literature which suggests that boys can be more vulnerable to the effects of task characteristics, and interesting tasks can enhance their performance significantly. Girls, in contrast, were

much more persistent even though the task content was evaluated as uninteresting. They reflect the result against other studies, concluding that boys and girls tend to be similar on high interest material, but with low interest material boys perform poorer, which is likely due to reduced effort – or what they call persistence. Interestingly, the measure of persistence in both of Ainley and colleagues' studies reviewed here (2002a; 2002b) was based on log data of time on task, which were recoded categorically based on whether the pupils had had enough time to read the tasks properly or not. Therefore, besides comparing girls' and boys' self-reported task interest it is interesting to look at gender differences in time investment also in the present study in which the larger sample enables much more detailed log data analyses.

### **2.3.2 Time investment and effort**

It is a widely acknowledged problem that the results of educational assessments may be influenced by reduced effort, if the assessments do not have any personal consequences for the pupils (e.g., Wise, 2006). Measures for effort have been developed for gaining a deeper understanding of factors influencing performance in an assessment situation (e.g., OECD, 2013a), but until the implementation of computer-based assessments these measures have necessarily been based on self-reports, which have been shown to be relatively unreliable (Wise & Kong, 2005). Log data analysis of time investment has, however, proven to be a much more accurate way of evaluating how much effort pupils really put in doing the tasks (Wise & Kong, 2005). Therefore, also in the present study log data is utilised in order to find out whether some of the differences between pupil subgroups could be explained by differences in the effort they invest in the learning to learn assessments.

Based on Carroll's model (1963), learning is determined by the ratio of the time needed and the time spent on learning. According to Carroll, the time needed depends on pupils' initial competences, their ability to understand instruction, and the quality of instruction. The relationship of pupils' initial competences and the time needed is expected to be negative, that is, pupils with lower initial competences would need more time than others to reach their learning goals. In the present study this means that pupils with identified support needs would need to increase

their time investment in the assessment tasks in order to perform at their own level. However, as the time spent depends both on the time allocated for the assessment and the time an individual pupil is willing to spend on them (Carroll, 1963), it may be that also time investment is influenced by the factors which are expected to influence performance also directly. Indeed, Kupiainen and colleagues found recently (2014) that the effects of detrimental attitudes on test performance were almost completely mediated by time on task. Also in the study of Ainley and colleagues (2002a), affective factors predicted time investment. Thus, there is increasing evidence that affective factors influence task behaviour, which is directly observable in the log data of computer-based assessment.

As log data analysis is a relatively new field of study, very little is known about how the task behaviour of different pupil subgroups differs, and to what extent these differences may explain the group-level differences in performance. In addition, task type can influence the time needed, too. Goldhammer and colleagues showed recently (2014) that in tasks requiring problem-solving, increased time investment predicted better performance regardless of pupils' prior competences, whereas in more routine reading tasks the relationship of time on task and performance was negative. As the learning to learn tasks used in the present study measure higher-order thinking skills instead of repeating the curricular contents, it is expected that increased time investment is associated with better performance in all pupil subgroups. However, group-level differences are expected to be found. For instance, Ainley and colleagues' (2002a) findings on girls' higher persistence indicate that gender differences could be partially explained by differences in time on task. Accordingly, group-level differences based on both gender and support needs may be partially caused by differences in attitudes, which are then mediated by time on task (cf., Kupiainen et al., 2014). In summary, time on task – alongside with task interest and learning-related attitudes – is expected to provide new explanations of why different pupil subgroups' performance seems to develop in different ways in different schools, classes and peer groups.



## 2.4 Summary

In the Introduction above, the Finnish educational system, the strategy for educational assessments and the role of learning to learn in it and the new support model were first presented as the context of the present study. After that, the literature review provided groundings for the hypotheses set in the four substudies below. First, the literature regarding cognitive development during primary school showed that general cognitive competences can be influenced by means of education – that is, it is possible for schools and classes to produce systematic effects on them – , and that the often observed gender differences should not depend on differences in general cognitive competences. However, it showed that initial differences between pupils, which also influence their later performance, are expected, and these are expected to derive partially from pupils' background.

Next, it was described how learning-related attitudes develop and how they gradually begin to predict performance. Some evidence for their possible role in explaining gender differences in performance, and differences between pupils with support needs and others, was also presented. Moreover, it was shown how both school achievement and learning-related attitudes develop partially in interaction with peers. In the last part of the Introduction two situational factors, which may influence performance, were presented. Task-specific interest has been shown to be related to performance especially for boys, whereas one of the possible explanations for girls' superiority in tasks pupils find uninteresting may be their greater effort in the assessment situation. In summary, the literature presented above suggests that the development of performance in low-stakes assessments is a complex phenomenon, which can only be studied in longitudinal settings, taking simultaneously into account several different perspectives on it.



## **3 Data**

This chapter gives an overview of the data used in all the substudies. The Helsinki sample and the two Vantaa cohorts and the procedures of data collection are first described. This is followed by descriptions of all the measures used in the substudies. The statistical methods for each substudy are reported in the respective results chapter.

### **3.1 Participants**

#### **3.1.1 The Helsinki sample**

The Helsinki data were drawn from a nine-year longitudinal study on the development of learning to learn skills during compulsory education. The study is being conducted by the Centre for Educational Assessment at the University of Helsinki on assignment from the Education Department of the City of Helsinki, and it will continue until spring 2016. This study covers the first six years of the longitudinal study, during which only a paper-based assessment (PBA) mode was used.

In autumn 2007, 17 schools were randomly selected from the schools in Helsinki using an equal-probability method that ensured representativeness with regard to socio-economic status. Originally, all 55 of the classes containing first-graders were instructed to participate by the Education Department of the City. However, 19 small classes for children with very high special education needs or completely lacking knowledge of the Finnish language, with one to eight first graders in each, were later excused from participating because the assessment tasks were considered too demanding for their pupils in general. Out of these 19 classes, three decided to take part anyway after reconsideration by the special education teachers. Out of the 17 schools that were instructed to participate, one school with two ordinary classes refused, making the final number of schools 16 and participating classes 40.

Since the study was conducted in collaboration with the Education Department of the city as a longitudinal assessment study of effectiveness

of education, the parents were informed through the Education Department, securing the agreement of all the sampled pupils. In all, 83 % of the parents also returned the attached background information questionnaire about their child's earlier development (see Lönnqvist, Verkasalo & Vainikainen, 2011, who first reported about the first grade data). Also teachers (79 %) filled out evaluation scales of each pupil's learning, behaviour in class and social skills and provided background information on the pupil. As on any school day, at the time of the assessment 5-10 % of the pupils were absent from each class due to sickness or other reasons. This made the final number of assessed pupils 744. The mean age of the pupils at the time of the first data collection was 7.31 years ( $Sd=.31$ ). Girls accounted for 51 % of the pupils.

In addition to the originally sampled schools, four new schools were included in the study in the beginning of the fourth year of the study as many pupils of the original sample had transferred to them. Two of these new schools were normal comprehensive schools built on new residential areas, whereas the other two were selective schools to which pupils typically transfer to in the beginning of third grade. The whole age cohorts of the 20 schools participated in the learning to learn (LTL) assessment in the beginning of fourth grade in 2010 and at the end of sixth grade in 2013, that is, also those almost 300 pupils who were not in the original first grade sample. In the fourth grade assessment in autumn 2010 there were 950 pupils present (53 % girls; mean age  $M=10.22$  years,  $Sd=.33$ ) and in the sixth grade assessment in spring 2013 there were 893 pupils (52 % girls, 4 pupils did not report their gender; mean age 12.81 years,  $Sd=.33$ ). 883 pupils were present in both the fourth and sixth grade assessments (52 % girls). There were 608 pupils that were present in all three data collections.

Also in fourth and sixth grade the class teachers (94 % / 95 %) and parents (82 % / 86 %) filled out questionnaires about the background, learning, working skills, behaviour and social skills of each pupil. Teachers also evaluated pupils' success in the most important school subjects on the normal scale of the Finnish school grades, ranging from 4 (failed) to 10 (excellent).

### 3.1.2 The two Vantaa cohorts

The Vantaa data were drawn from a panel assessment study in which several whole age cohorts within the municipality were followed from 2010 to 2013 (Marjanen, Vainikainen, Kupiainen, Hotulainen & Hautamäki, 2014). The aim of the study was to examine how schools manage to support the development of pupils' learning to learn skills in primary and lower secondary school. This was done both cross-sectionally and longitudinally. In this study only the data from first and third graders in 2010 (third and sixth graders in 2013) were used. The cohorts were selected by the Education Department of the municipality to participate in an educational assessment study conducted by the Centre for Educational Assessment at the University of Helsinki. Even though the whole age cohorts within the municipality were originally selected to participate, some pupils with very high special education needs were excluded after their teachers had considered the assessment tasks too demanding for them. Like in previous cycles of assessment in the same municipality, these decisions were made on an individual basis and whole special education classes were excluded only if all their pupils met the exclusion criteria. In 2010, all the pupils did paper-based (PBA) versions of the LTL measures. In 2013, 20 % of the pupils were randomly assigned to the PBA group while 80 % completed the computer-based (CBA) versions of the tests.

The first grade cohort of 2010 consisted of 2245 pupils in 135 classes in 36 schools. Twelve small special education classes were excluded from the assessment as all their pupils met the exclusion criteria. As on any school day, at the time of the assessment 5-10 % of the pupils were absent from most classes due to sickness or other personal reasons. This made the final number of assessed pupils 2089, of which 48 % were girls and 52 % boys. The mean age of the pupils was 7.25 years,  $Sd=.38$ .

The third grade cohort of 2010 consisted of 2096 pupils in 120 classes in 36 schools. Of these, 14 small classes met the exclusion criteria as a whole, and some pupils were absent at the time of the assessment. In all, 1984 (960 girls and 1024 boys) participated in the assessment. The mean age of the pupils was 9.75 years ( $Sd=.43$ ).

The parents of the pupils were informed about the assessment through a letter signed by the Education Department, and the information leaflet

contained also a background information questionnaire. The questionnaire was filled out by 78 % parents of both first and third graders.

The third grade cohort of 2013 consisted of 2215 pupils in 126 classes in 35 schools. Ten small special education classes were excluded from the assessment as all their pupils met the exclusion criteria, and two regular classes did not complete the assessment tasks within the given timeframe. At the time of the assessment 5-10 % of the pupils were absent from most classes. This made the final number of assessed third graders 2115, of which 1023 were girls and 1069 boys. Eleven pupils did not report their gender. The age of the pupils was  $M=9.83$ ,  $Sd=0.38$ . Of the final number of pupils, 1797 of them had also been present in the first grade assessment in 2010 (869 girls and 928 boys).

The sixth grade cohort of 2013 consisted of 2113 pupils in 118 classes in 37 schools. Due to the exclusion criteria, 8 small special education classes did not participate, and individual pupils were absent from most classes at the time of the assessment. The final number of assessed sixth graders was 1979, of which 986 were girls and 988 boys. Five pupils did not report their gender. The age of the pupils was  $M=12.67$ ,  $Sd=.43$ .

As in 2010, the parents of the participants were informed about the assessment through a letter signed by the Education Department of the City. In all, 81 % of the third graders' and 77 % of the sixth graders' parents filled out the background questionnaire attached to the information leaflet.

Like in the Helsinki study, class teachers filled out questionnaires about the background, learning, working skills, behaviour and social skills of each pupil (response rate  $\approx 93$  % in all four data sets). Except for the first graders, the teachers also evaluated pupils' success in the most important school subjects on the normal scale of the Finnish school grades, which ranges from 4 (failed) to 10 (excellent).

### **3.2 Measures**

All the tasks and scales presented to the pupils in the four substudies are first summarised here in Table 3.1.

**Table 3.1.**

An overview of the tasks and scales used in the four substudies

Scale	Substudy			
	1	2	3	4
<b><u>First grade learning preparedness test</u></b>				
Analogical reasoning	X			X
Visuo-spatial memory	X			X
Following instructions	X			X
<b><u>Third/fourth grade cognitive learning to learn tasks</u></b>				
Mental arithmetics	X	X		X
Arithmetical operations	X	X		X
Reasoning (The Bottles task)	X	X		X
Analogical reasoning	X		X	X
Reading comprehension	X			X
<b><u>Sixth grade cognitive learning to learn tasks</u></b>				
Mental arithmetics		X		X
Arithmetical operations		X	X	X
Mathematical concepts			X	X
Reading comprehension			X	X
Verbal proportional reasoning			X	X
Control of variables			X	X
Reasoning (The Bottles task)		X		X
<b><u>Mastery Attitudes</u></b>				
Learning orientation (Mastery: Intrinsic)	X	X		
Achievement orientation (Mastery: Extrinsic)	X	X	X	X
Agency: Effort	X	X	X	X
Importance of School			X	X
<b><u>Detrimental Attitudes</u></b>				
Means-ends: Ability			X	X
Means-ends: Chance			X	X
Self-handicapping			X	X
<b><u>Task interest</u></b>				
Arithmetical operations		X		X
Mental arithmetics		X		X
Reasoning (The Bottles task)		X		X

### 3.2.1 First grade assessments of learning preparedness

A learning preparedness group test (Hautamäki et al., 2001; Vainikainen, Kupiainen, Marjanen, Gustavson & Hautamäki, 2011; see also Lönnqvist, Vainikainen & Verkasalo, 2012) was administered to the first grade pupils shortly after the school start. The paper-and-pencil assessment tasks were presented to the pupils by their own class teacher as a part of their normal school work. To avoid the effects of exhaustion, teachers were instructed to present only one task to the pupils each day. The test comprised of six non-verbal cognitive tasks in the Helsinki study in 2007 and seven tasks in the Vantaa study in 2010 as well as two drawing tasks which will not be discussed further. Besides the additional task in Vantaa, there were some item-level differences in the test versions, which made the Helsinki test version slightly easier than the Vantaa version as a whole. The adjustments – which consisted mostly of the adding of more difficult items and in some tasks leaving out too easy ones – were done after the Helsinki data collection in order to increase the validity of the test in assessing also better-performing pupils' learning preparedness. As the only study that utilised the first grade data here was substudy 1, in which performance of the two samples was compared, only tasks and items common to both test versions were used. Only three of the cognitive tasks were selected in the final analyses as one of the common tasks was too difficult and one too easy to produce variation between pupils. Thus, only the three tasks used in analyses are described here.

The Analogical reasoning task was adapted from a Dutch geometric analogies test (Hosenfeld, van den Boom & Resing, 1997). The pupils were presented a pair of geometric figures, e.g. a small square on the left and a big square on the right. The task was to apply the same rule when the pupil had to choose a pair from five options for another figure (e.g. a small circle). The transformations included adding an element, changing sizes and positions, halving and doubling, and the maximum number of simultaneous transformations was three. Each analogy was scored dichotomously as correct or incorrect. The Helsinki test version consisted of seven items and the Vantaa version of eight, of which four were common to both. An average score for each pupil was calculated of the four dichotomously coded common items.



Visuo-spatial memory was assessed by a task originally developed by Wilson, Scott and Power (1987) and modified by Logie and Pearson (1997). Even though Halpern (2000) suggests that boys might have an advantage in visuo-spatial memory tasks (p. 91), gender differences have not been observed when using this task with Finnish first graders (Hautamäki et al., 2001; Vainikainen et al., 2011). In the task, the pupils were presented grids of different sizes for three seconds. Some of the squares of the grids were painted black, and some of them were unpainted. After showing the picture, the pupils were asked to reproduce the figure they just saw in an empty grid of the same size. Each grid was scored dichotomously as correct or incorrect. The Helsinki test version consisted of six items and the Vantaa version of eight, of which six were identical to those used in the Helsinki test. An average score for each pupil was calculated of the common items.

The pupils' capacity to follow the teacher's instructions was assessed by a task originally developed by Elkonin (see Raigorodsky [Ed.], 2008) and modified by Hautamäki and colleagues (2001). The task can be understood as measuring both children's inductive reasoning and executive functions. In this task the pupils had to draw a path on an empty 12x5 grid according to the teacher's dictation. The teacher dictated the path step by step, eg. draw two steps forwards...then two steps towards the sun (a picture on the right side of the grid) etc. Halfway through the grid, the teacher stopped dictating, and the pupils had to continue the path according to the same rule. The grids were scored dichotomously as correct or incorrect. The task was identical in both test versions, so an average score of all four items for all the pupils was calculated. Again, even though gender differences in favour of boys have been reported in many types of visuo-spatial tasks (Halpern, 2000), they have not been observed in this or the analogical reasoning task with Finnish first graders (Hautamäki et al., 2001; Vainikainen et al., 2011).

### **3.2.2 Learning to learn assessments from grade three to grade six**

Also in the third/fourth and sixth grade the assessments were conducted by class teachers according to written instructions. The pupils filled out a learning to learn test booklet, comprising of cognitive tasks and questionnaires measuring learning-related attitudes. The pupils were allocated four separate 45-minute sessions for the assessment in Vantaa right before the summer break at the end of the third grade in 2013, and in Helsinki shortly after the summer break in the beginning of the fourth grade in 2010. Despite the difference in pupils' grade level, the results can be interpreted as comparable as it has been shown in many occasions that the pupils' performance level does not increase during the summer break (Cooper, Nye, Charlton, Lindsay & Greathouse, 1996). The sixth grade assessment was conducted in both cities before the summer break in 2013, and the pupils were allocated one 90-minute session without breaks. The time allocated for the assessment had proven sufficient in previous assessments.

#### ***Cognitive tasks in the third/fourth grade***

The tasks of the Finnish learning to learn (LTL) instruments fall into three subareas: reading comprehension, mathematical thinking skills and reasoning skills. The third/fourth grade test versions consisted of two reading comprehension tasks, three mathematical thinking skills tasks and three / four reasoning skills tasks for Helsinki / Vantaa. Only two out of three of the mathematical thinking skills and reasoning tasks were used in this study and are described below.

Reading comprehension was assessed by two tasks based on expository texts. One was a hierarchy-rating task, developed within the theoretical framework of Kintsch and van Dijk (1978) and calibrated on adult interpretation of the text (cf., Lehto, Scheinin, Kupiainen & Hautamäki, 2001; Lyytinen & Lehto, 1998). The pupils were asked to read a one-page text and then to assess 16 statements based on the text as to whether they are a good description of the text as a whole, present important information regarding the content of the text or just refer to minor details in the text. The other reading comprehension task was a shorter text set in a context closer to everyday life, which was adapted

from a Finnish Vocational Guidance Office test. It assesses pupils' ability to understand, analyse and interpret written information with four multiple-response items (Hautamäki & al., 2002). All the items were coded dichotomously as correct or incorrect, and the average score of all the 20 items together was calculated. Thus, the longer hierarchy-rating task received a substantially larger weight in the average scores.

The first task for mathematical thinking skills, the Mental Arithmetics task, was based on the idea of the Arithmetic subscale of the Wechsler Adult Intelligence Scale - Revised (WAIS-R: Wechsler, 1981). The teacher read aloud a mathematical problem (e.g. If you buy two bus tickets and one ticket costs 3 euros 50 cents, how much money do you get back if you give 10 euros?), and the pupils wrote down the answer in their test booklets. The items were coded dichotomously as correct or incorrect. The second task type, the Hidden Arithmetical Operators task (Arithmetical Operations for short) was developed by Demetriou and his colleagues (Demetriou, Pachaury, Metallidou & Kazi, 1996; Demetriou, Platsidou, Efklides, Metallidou & Shayer, 1991). In each item there were one to four hidden operators (e.g.,  $[(5 a 3) b 4 = 6$ . In this task letter a / b stands for: addition (+) / subtraction (-) / multiplication ( $\cdot$ ) / division ( $\div$ )?). The items were coded dichotomously for a correct answer to all of the 1-4 operators in the item. As different combinations of the items of these tasks were used in different substudies, the methods for calculating average scores for mathematical thinking skills are explained in the Measures sections of each substudy.

Of the reasoning tasks used in this study, the understanding of horizontal and vertical axes, sometimes also called spatial reasoning (cf. Demetriou & Kazi, 2006), was measured by the classical Piagetian water-level task (Piaget & Inhelder, 1956; Hautamäki, 1984). A picture of eight empty bottles was presented to the pupils. One of the bottles was standing, and the rest of them were inclined by  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $320^\circ$ ,  $270^\circ$ ,  $225^\circ$  and  $180^\circ$  grades, respectively. The task was to draw a line indicating the water level and mark the area filled with water when each bottle is half full. The bottles were scored dichotomously as correct or incorrect, and then an average score was calculated of the eight coded items. Even if Halpern (2000) uses this task as an example of a task in which boys usually perform better than girls, no systematic gender

differences have been observed in any Finnish data from third to sixth grade.

The third/ fourth grade test version included also a more difficult version of the analogical reasoning task described above for the first grade test. There were some item-level changes in the task between 2010 and 2013, but six items were common for the Vantaa third grade assessment in 2013 and the Helsinki fourth grade assessment in 2010. For substudy 1, average scores were calculated based on these dichotomously coded common items. In substudy 2, that used only the Helsinki data, an average score of all the eight items of the Helsinki test version was used in the analyses.

### ***Cognitive tasks in the sixth grade***

The sixth grade cognitive tasks were very similar to the third/fourth grade tasks. However, in many tasks some of the easiest items were replaced with more difficult ones. The reading comprehension tasks were identical to those used in the third/fourth grade, so the average scores calculated as described above were directly comparable with the third/fourth grade scores. Mathematical thinking skills were measured by three tasks of which two are described above. For Mental Arithmetics five items were identical with items in the third/fourth grade test version, but the three easiest items were replaced by more difficult ones. However, only the common items were used in this study since the Mental Arithmetics task were not used at all in Substudy 3 which cross-sectionally reports sixth graders' results in the computer-based assessment in Vantaa. For Arithmetical operations there were five common items in the Vantaa third grade and Vantaa sixth grade test versions, and four common items in the Helsinki fourth grade and Helsinki sixth grade test versions. However, only three items were present in all of these versions, which made longitudinal comparisons across samples difficult. Therefore, only the Helsinki data were used in Substudy 2.

The third mathematical thinking skills task, Invented Mathematical Concepts which was used only in analyses of the Vantaa sixth grade CBA results, was a modified group-version of Sternberg's Triarchic Test (H-version) Creative Number scale (Sternberg, Castejon, Prieto, Hautamäki

& Grigorenko, 2001) where an arithmetical operator is conditionally defined depending on the value of the digits they combine (e.g., if  $a > b$ , lag stands for subtraction, and else for multiplication). There were eight items with four multiple choice alternatives, which were coded dichotomously for the whole equation.

Sixth graders' reasoning skills were measured by four tasks, of which the first was the Piagetian water-level task described above. All the items of the Bottles task were identical with those used in the other test versions, and in Substudies 2 and 4 this was the only reasoning task which was used in longitudinal analyses. However, this task was not used in Substudy 3 as due to technical issues it was not included in the CBA version of the LTL test. Two tasks, used only when analysing the sixth grade CBA results, measured verbal proportional reasoning (shortly verbal reasoning): Five items were taken from the Bond's Logical Operations Test (Bond, 1995/1976), which operationalises item-by-item each of the schemas of the formal operational stage identified by Inhelder and Piaget (1958). Each multiple-choice item comprised an item of two to four short sentences followed by a set of four or five alternative responses (Bond & Fox, 2012). Five items were adapted from the Missing Premises task of the Ross Test of Higher Cognitive Processes (Ross & Ross, 1979). The pupils were given one premise and the conclusion, and they had to choose from among five alternatives the second premise which would make the conclusion valid. All the items were scored dichotomously as correct or incorrect.

The last reasoning task for sixth graders was Control of Variables, which is a modified version (Hautamäki, 1984) Shayer's (1979) Science Reasoning Tasks 'Pendulum'. It is based on one of the formal schemata identified by Inhelder and Piaget (1958). The pupils were presented with six items in the form of comparisons set in the world of Formula 1 races with four variables: driver, car, tires and track, with two alternatives for each. The pupils were to judge whether the single effect of the driver, car, tires and track could be concluded from the comparison. There were four comparisons with 3 or 4 Yes/No -choices for variables and two comparison-sets to be complemented. The items were coded dichotomously for a correct answer to all of the variables in the item.

### ***Learning-related attitudes***

Learning-related attitudes were measured by presenting questionnaires to third/fourth and sixth graders between the cognitive assessment tasks. Only a part of the scales in the questionnaires were used in this study. In Substudies 1 and 2 “Learning-related attitudes” refer to positive attitudes which support learning. They were measured by scales from two subfields of motivational theory: achievement goal theory (e.g., Harackiewicz et al., 2002) and agency beliefs theory (e.g., Little et al., 2001). From achievement goal theory, *Mastery Intrinsic Orientation*, shortly *Learning Orientation* (“An important goal for me at school is to learn new things.”) and *Mastery Extrinsic Orientation*, shortly *Achievement Orientation* (e.g. “Getting good grades at school is important to me.”) were included, tapping into the internalised value of learning and attainment (cf., Ryan & Deci, 2000). Of agency beliefs the construct *Agency: Effort* (e.g., “I work hard to do well at school.”) was included. In Substudy 4 *Mastery Intrinsic Orientation* was replaced by a more practice-oriented scale *Importance of school* (e.g. “I think we learn useful and important things at school) to achieve comparability with the study of Kupiainen et al. (2014) on the effects of time on task. All the items in the scales were answered both in third/fourth grade and sixth grade with a 7-point Likert-scale in which 1= Not true at all and 7= Very true.

The sixth grade questionnaires included also scales for attitudes detrimental to learning. These were used in addition to the positive attitudes in Substudy 3 when the effects of attitudes on time on task were studied and in Substudy 4 regarding sixth grade results. Two of the detrimental attitude scales come from the theoretical background of means-ends-beliefs: *Means-ends: Chance* (e.g. “Failure at school is mainly due to bad luck.”) and *Means-ends: Ability* (e.g. “Poor marks are due to lack of ability.”; [e.g., Niemivirta, 2002]). The third scale is for *Self-handicapping* (e.g. “I give up easily if my assignments look too demanding.”), deriving from achievement goal theory (e.g., Urdan & Midgley, 2001).

### ***Background variables***

The background variables used in this study were extracted from the pupil background information questionnaire presented before the cognitive tasks or other questionnaires, parent questionnaires and teachers' evaluations of each pupil. Pupils' gender was taken from the self-reports and mothers' educational level from the parent questionnaire. Originally, educational level was asked about with a seven-category multiple-response question, but the answers were recoded as three levels: basic level (only compulsory education), secondary level (upper secondary school or vocational training) and tertiary level (university or polytechnics education). From teacher-evaluations three types of information were used here: For first graders the teachers evaluated each pupils' reading skills when they entered school, that is, what they had learned already before formal schooling began. For third/fourth and sixth graders the teachers reported whether the pupil had received intensified or special support as defined in their individual support documents, prepared according to the new educational legislation. Teachers also evaluated their school achievement in the most important school subjects (mother tongue, mathematics, English, science). The scale was that of the normal school grades in Finland, ranging from 4 (failed) to 10 (excellent).

### ***Additional measures***

In the substudies additional measures for reading skills at school start (Substudy 1), task interest (Substudies 2 and 4), time investment (Substudy 3) and social relationships within classes (Substudy 4) were used. They are described in detail in the Measures section of the substudy in question.





## **4 Substudy 1: Explaining ten-year-olds' test performance in two municipalities**

The first substudy aims at finding predictors for 10-year-olds' performance in the cognitive tasks of a low-stakes learning to learn assessment in two municipalities with slightly different educational policies. It does this by trying to shed light on where the systematic subgroup-level effects, that are often observed in assessments conducted in secondary school, have their origins. Ten-year-olds' performance in the assessment is in this substudy predicted by their learning preparedness and reading skills in the beginning of first grade, and the role of learning-related attitudes – unlike in typical cross-sectional assessments - is studied after controlling for prior cognitive competence. Educational equity is then evaluated by adding pupil-level background variables to the model, one at a time. More specifically, the following research questions were stated:

Q1.1: To what extent is performance in a third/fourth grade low-stakes learning to learn assessment predicted by cognitive competences and reading skills in the beginning of the first grade? Do learning-related attitudes explain performance in the assessment already in the third/fourth grade when prior cognitive competence and reading skills are taken into account?

Q1.2: Are there gender differences in the assessed skills and attitudes and their relationships with each other already during the first years of basic education?

Q1.3.: Are individual support needs, as officially defined by schools, related to the assessed skills and attitudes and their relationships with each other already during the first years of basic education?

Q1.4.: Does mothers' education explain competence differences between pupils already during the first three years of basic education? Do the background-related differences increase over time?

Based on the literature presented above, the following hypotheses were set:

H1.1.: Performance in the first grade learning preparedness test, reading skills at school start and learning-related attitudes are all predictors of performance in the third/fourth grade learning to learn test (Duncan et al., 2009). It is expected that earlier reading skills are stronger predictors of later reading comprehension (Leppänen, Aunola, Niemi & Nurmi, 2008) while the non-verbal cognitive skills measured by the first grade learning preparedness test predict mathematical thinking and reasoning skills relatively well. The additional value of learning-related attitudes in explaining performance is expected to be small but statistically significant (Aunola et al., 2006; Gagné & St Père, 2002; Klauer, 1988; Steinmayr & Spinath, 2009).

H1.2.; Girls are better readers when they start formal schooling, but they should not perform better than boys in non-verbal cognitive tasks (Halpern, 2000). As very large gender differences have been observed in assessments conducted in Finnish secondary schools (Hautamäki et al., 2013; OECD, 2013), it is expected that the differences start to grow already during the first years of formal schooling even though traditionally boys have been better in mathematical tasks and girls in reading (Halpern, 2000). In this, the evidence from earlier studies is controversial. Girls are expected to have more positive learning-related attitudes than boys (Kenney-Benson et al., 2006).

H1.3.: Support needs are related to both cognitive competences and reading skills at school start, indicating that support needs are adequately recognised at school, and on the other hand, that the first grade learning preparedness test has practical relevance. According to the principles of early intervention and support of the renewed Basic Education Act, and the assumption of educational equity, it is expected that the differences between pupils in need of support and the others do not increase during the follow-up period.

H1.4.: Mothers' education is related to both reading skills and cognitive competences at school start (Caro, McDonald & Willms, 2009; Leppänen et al., 2008). International literature also suggests that the differences between pupils with mothers on different educational levels grows over time but that the growth begins to accelerate first when children are about 11 years old (Caro et al., 2009). As one of the equity-related goals of the Finnish educational system is that the initial differences between children should not increase over time, it is hypothesised that in general the differences do not grow during the follow-up period even though Leppänen and colleagues (2008) found a small effect in fourth graders' reading comprehension. Mothers' education is expected to be related to pupils' attitudes as well (cf. Hautamäki et al., 2013, for results from the lower secondary level).

## **4.1 Participants**

In this substudy, the Helsinki and Vantaa samples were compared using the data from the first and third/fourth grade assessments. Therefore, in Helsinki, only the pupils who had belonged to the original first grade sample and had data from the fourth grade were selected (N=608, 51 % girls). Of the Vantaa cohort only the randomly selected 20 % who took the PBA version of the third grade test and had data from first grade (N=371, 51 % girls) were selected to avoid the effects of CBA mode on the results (cf., Csapó et al., 2014).

## **4.2 Measures**

### **4.2.1 Cognitive tasks**

The first grade tasks and their items are described in detail in Chapter 3. From the third/fourth grade test the reading comprehension tasks were used as described in Chapter 3. Of the mathematical thinking and reasoning skills tasks, only tasks and items which were identical in both test versions were used. This left out of the analyses the Mathematical concepts task due to substantial changes of instructions, and both verbal reasoning tasks due to too few common items. Thus, the mathematical thinking skills scores were calculated for each pupil by first averaging the five common items of the Arithmetical Operators task and all the eight items of the Mental Arithmetics task separately, and then averaging the two average scores to balance the different number of items. The reasoning skills scores were calculated by averaging the six common items of Analogical Reasoning and the eight items of the Bottles task separately, and then averaging the two average scores. The reliability of the tasks for both samples is presented in Table 4.1.

### **4.2.2 Learning-related attitudes**

Pupils' learning-related attitudes were measured by the scales described in detail in Chapter 3: Learning Orientation (e.g. "An important goal for me at school is to learn new things."), Achievement Orientation (e.g. "Getting good grades at school is important to me.") and Agency: Effort

(e.g., “I work hard to do well at school.”). All the items in the scales were answered with a 7-point Likert-scale in which 1= Not true at all and 7= Very true. The reliability of the scales for both samples is presented in Table 4.1. It has to be noted that the reliability of some of the scales was somewhat low especially in the Vantaa sample, mainly due to the small number of items per task or the relative difficulty of the reading comprehension items.

**Table 4.1.**

Reliability of the scales for Helsinki and Vantaa (Cronbach's  $\alpha$ )

Scale	Number of items	$\alpha$ Helsinki	$\alpha$ Vantaa
Analogical reasoning first grade	4	.55	.46
Visuo-spatial memory first grade	6	.56	.52
Following instructions first grade	4	.60	.67
Learning orientation third/fourth grade	3	.85	.77
Achievement orientation third/fourth grade	2	.66	.54
Agency: Effort third/fourth grade	3	.77	.74
Mathematical thinking third/fourth grade	13	.76	.67
Reading comprehension third/fourth grade	20	.57	.45
Reasoning skills third/fourth grade	14	.84	.82

#### 4.2.3 Other variables in the model

The background variables used in this study were self-reported gender (0=Boy, 1= Girl), teacher-reported needs of intensified or special support on third/fourth grade (0=No support needs, 1=Support needs) and information about mothers' education that was taken from the third/fourth grade parent questionnaire (1=Basic education, 2=Secondary education, 3= Tertiary education). From first grade teacher-evaluation the question, “How well did the pupil read when he/she came to school?” was used. The scale of that question was from 1= Not at all to 7= Much better than usual.

### 4.3 Statistical methods

Structural equation modeling (SEM) with multiple-group analysis functions was used in AMOS21. Since the deviation from normality of all variables was within the recommended limits (Kline, 2005), maximum likelihood estimation was used. The models were considered as having a good fit with CFI and TLI  $> .95$  and RMSEA  $< .08$ . Also reported are  $\chi^2$  values, but due to the sample size and the large number of variables in the models significant p-values were to be expected. Therefore, they are not considered as an absolute criterion for model fit. Measurement invariance was tested by adding stepwise constrains to the measurement models, first by constraining factor loadings (weak factorial invariance) and then intercepts (strong factorial invariance) across groups while letting the latent mean of one of the groups vary free (Byrne & Stewart, 2006).

### 4.4 Results and discussion

The descriptive statistics for all the variables are presented in Table 4.2. and by groups in Tables 4.3-4.5. The values of the two samples are separated with a vertical bar, always reporting the results of the Helsinki sample first. In case only one value is reported it applies to both samples.

All the other results were analysed by structural equation modelling (SEM). Before specifying the whole path model, the measurement models for the latent factors were tested separately to check their measurement invariance across the two samples. First, a measurement model was specified for the first grade learning preparedness test. The cognitive items that were common for both samples were first parcelled (see Matsunaga, 2008) into task-based average scores, which were then regressed to a first order factor of First grade test score. In order to achieve identifiability and to proceed with the testing of measurement invariance across the two cities, the measurement model was correlated with teacher-evaluations of reading skills at school start. The changes in the fit indices when constraints were added are presented in Table 4.6.

**Table 4.2.**

Descriptive statistics for the variables used in the models for Helsinki and Vantaa

Variable	N	Min	Max	M	Sd
Reading skills first grade	472   371	1.00	7.00	3.77   3.43	2.39   2.20
Analogical reasoning first grade	608   361	0.00	1.00	0.65   0.63	0.29   0.28
Visuo-spatial memory first grade	603   349	0.00	1.00	0.66   0.61	0.24   0.24
Following instructions first grade	605   360	0.00	1.00	0.71   0.72	0.29   0.31
Learning orientation third/fourth grade	581   354	1.00	7.00	5.84   5.89	1.24   1.20
Achievement orientation third/fourth grade	581   354	1.00	7.00	6.03   5.97	1.16   1.17
Agency: Effort third/fourth grade	580   354	1.00	7.00	5.75   5.62	1.13   1.18
Mathematical thinking third/fourth grade	594   331	0.00	1.00   0.81	0.41   0.44	0.21   0.18
Reading comprehension third/fourth grade	604   351	0.00	0.85   0.95	0.38   0.33	0.16   0.14
Reasoning skills third/fourth grade	556   326	0.00	1.00	0.52   0.48	0.26   0.25
Gender (0=Boy, 1= Girl)	608   371	0.00	1.00	0.51   0.49	0.50   0.50
Support needs (0= No support, 1= Support)	563   360	0.00	1.00	0.11   0.22	0.32   0.42
Mothers' education	510   306	1.00	3.00	2.30   2.28	0.63   0.64

The values for Helsinki | Vantaa are separated with a vertical bar

N= Number of responses, Min= minimum value, Max=maximum value, M=Mean, Sd=Standard deviation

**Table 4.3.**

Descriptive statistics by gender for the variables used in the models

Variable	Girls		Boys	
	M	Sd	M	Sd
Reading skills first grade	3.86   3.65	2.34   2.23	3.67   3.21	2.44   2.16
Analogical reasoning first grade	0.66   0.65	0.27   0.28	0.64   0.61	0.30   0.28
Visuo-spatial memory first grade	0.65   0.59	0.24   0.23	0.68   0.62	0.24   0.24
Following instructions first grade	0.71   0.73	0.28   0.31	0.72   0.70	0.30   0.31
Learning orientation third/fourth grade	5.95   6.02	1.09   0.98	5.72   5.74	1.38   1.37
Achievement orientation third/fourth grade	6.20   6.20	1.02   0.96	5.86   5.75	1.27   1.31
Agency: Effort third/fourth grade	5.92   5.77	0.97   1.06	5.57   5.48	1.24   1.27
Mathematical thinking third/fourth grade	0.39   0.44	0.20   0.18	0.46   0.45	0.22   0.18
Reading comprehension third/fourth grade	0.39   0.34	0.17   0.15	0.36   0.32	0.16   0.13
Reasoning skills third/fourth grade	0.52   0.50	0.27   0.24	0.51   0.48	0.26   0.25

The values for Helsinki | Vantaa are separated with a vertical bar

**Table 4.4.**

Descriptive statistics by support needs for the variables used in the models

Variable	No support needs		Support needs	
	M	Sd	M	Sd
Reading skills first grade	3.98   3.79	2.38   2.19	2.45   2.15	2.03   1.71
Analogical reasoning first grade	0.66   0.67	0.29   0.24	0.52   0.49	0.29   0.27
Visuo-spatial memory first grade	0.68   0.63	0.23   0.22	0.57   0.54	0.26   0.27
Following instructions first grade	0.73   0.76	0.28   0.29	0.54   0.60	0.35   0.33
Learning orientation third/fourth grade	5.86   5.91	1.19   1.14	5.56   5.75	1.58   1.41
Achievement orientation third/fourth grade	6.08   6.04	1.09   1.09	5.56   5.72	1.59   1.40
Agency: Effort third/fourth grade	5.79   5.75	1.08   1.08	5.41   5.16	1.44   1.39
Mathematical thinking third/fourth grade	0.45   0.47	0.20   0.17	0.26   0.34	0.20   0.18
Reading comprehension third/fourth grade	0.39   0.34	0.16   0.14	0.29   0.30	0.11   0.14
Reasoning skills third/fourth grade	0.53   0.50	0.26   0.24	0.42   0.41	0.27   0.24

The values for Helsinki | Vantaa are separated with a vertical bar

**Table 4.5.**

Descriptive statistics by mothers' education for the variables used in the models

Variable	Basic		Secondary		Tertiary	
	M	Sd	M	Sd	M	Sd
Reading skills first grade	2.89   3.16	2.43   2.46	3.84   3.20	2.39   2.22	4.49   4.18	2.46   2.06
Analogical reasoning first grade	0.55   0.56	0.30   0.28	0.66   0.62	0.30   0.29	0.72   0.49	0.28   0.26
Visuo-spatial memory first grade	0.65   0.57	0.26   0.23	0.67   0.60	0.26   0.25	0.68   0.63	0.25   0.22
Following instructions first grade	0.71   0.64	0.29   0.35	0.71   0.73	0.29   0.30	0.71   0.79	0.29   0.25
Learning orientation third/fourth grade	5.69   5.71	1.37   1.51	5.82   5.89	1.12   1.17	5.79   5.85	1.35   1.14
Achievement orientation third/fourth grade	5.85   5.90	1.25   1.36	5.99   5.93	1.15   1.16	6.02   6.11	1.26   1.06
Agency: Effort third/fourth grade	5.59   5.52	1.16   1.42	5.71   5.73	1.09   1.07	5.74   5.53	1.19   1.13
Mathematical thinking third/fourth grade	0.37   0.40	0.20   0.19	0.46   0.44	0.21   0.17	0.51   0.48	0.23   0.18
Reading comprehension third/fourth grade	0.35   0.30	0.13   0.17	0.39   0.32	0.17   0.13	0.44   0.36	0.17   0.15
Reasoning skills third/fourth grade	0.47   0.41	0.25   0.23	0.49   0.48	0.25   0.25	0.68   0.54	0.22   0.23

The values for Helsinki | Vantaa are separated with a vertical bar

Note: The information about mothers' education was missing more often than other background variables, and the loss may have not been completely random

In the baseline model the factor loadings and intercept were allowed to vary freely. In the next model, factor loadings were constrained equal but intercepts were allowed to vary across groups. When latent means are not compared and only the structural relationships of the variables examined as is the case in the present study, this would have been the sufficient level of measurement invariance (Steenkamp & Baumgartner, 1998). However, also strong factorial invariance (Byrne & Stewart, 2006) was tested by constraining the intercepts equal across groups and allowing the latent mean of one group to vary free. As can be seen in Table 4.6., this led to a slight decrease of the fit indices, and the  $\chi^2$  model comparison test resulted in a significant p-value. When the constraint of the intercept of the visuo-spatial memory task was released, the fit indices rose again.

Next, measurement invariance of the latent third grade attitude factor was tested accordingly. Again, the items were first parceled into scale-based averages, which were then regressed into a general attitude factor. As can be seen in Table 4.6., the measurement model had an excellent fit, which was not affected by the constraints. Thus, strong factorial invariance was concluded to hold here.

**Table 4.6.**  
Measurement invariance of the latent variables across Helsinki and Vantaa

Model	$\chi^2$	df	CFI	TLI	RMSEA	p
<u>First grade test</u>						
Baseline model	5.871	4	.993	.965	.022	.209
Factor loadings constrained	8.974	6	.989	.963	.023	.212
Measurement intercepts constrained	21.262	8	.951	.877	.041	.006
Measurement intercepts constrained partially	10.529	7	.987	.963	.023	.161
<u>Third/fourth grade attitudes</u>						
Baseline model	6.075	4	.998	.991	.023	.194
Factor loadings constrained	9.202	6	.997	.991	.023	.209
Measurement intercepts constrained	16.424	10	.994	.986	.029	.065

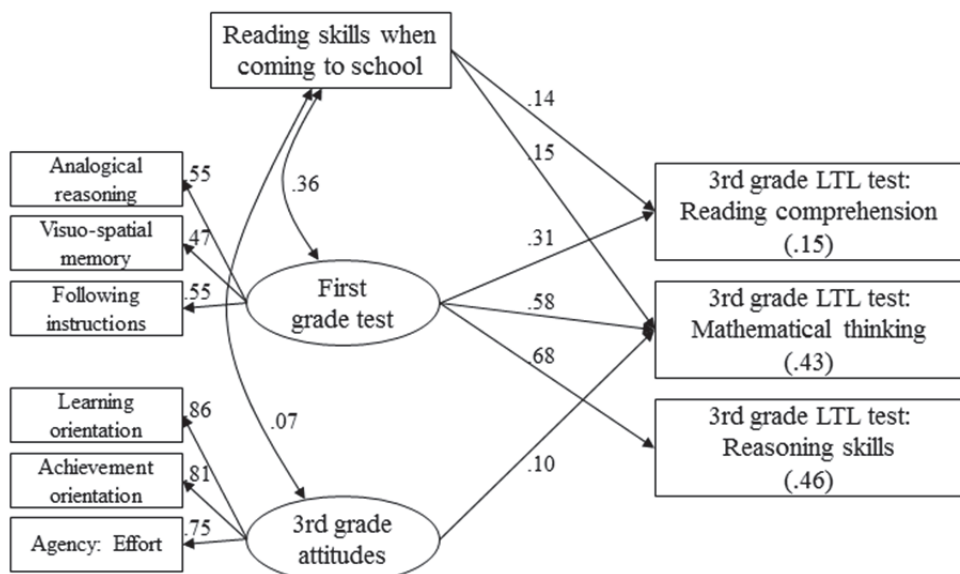
H1.1: In the first hypothesis it was expected that performance in the first grade learning preparedness test, reading skills at school start and learning-related attitudes would all be predictors of performance in the



third/fourth grade learning to learn test. Moreover, earlier reading skills were expected to be a stronger predictor of later reading comprehension while mathematical thinking and reasoning skills would depend more on the non-verbal cognitive competences measured by the first grade test. The role of attitudes was expected to be significant but weak.

A structural equation model was specified for testing this hypothesis. Instead of treating the LTL measure as a test producing a single test score, or a latent factor for general cognitive skills as was done with the first grade test, the subscores of the three task types (reading skills, mathematical thinking skills and reasoning skills) were used in the model separately. The first grade latent test score, teacher-evaluated reading skills when coming to school and third/fourth grade learning-related attitudes were added as predictors of all of them simultaneously. The model fit was good when the model was fitted to the whole data consisting of both samples (CFI=.962, TLI=.955, RMSEA=.030,  $\chi^2=65.335$ ,  $df=56$ ,  $p=.184$  when even residuals were constrained equal).

The next step was to remove insignificant paths from the model. Learning-related attitudes did not correlate with the first grade test score, and they only predicted performance in the mathematical thinking skills task. The final model without the background variables is presented in Figure 4.1.



**Figure 4.1.** Predicting performance in the third/fourth grade learning to learn test. Weights, intercepts, means, covariances and residuals constrained equal across the two samples. Numbers in parentheses indicate the share of accounted for variance.

Figure 4.1 shows that in the whole sample from the two cities, performance in the first grade test was a strong predictor of performance in mathematical thinking ( $\beta=.58$ ,  $p<.001$ ) and reasoning ( $\beta=.68$ ,  $p<.001$ ) tasks. It also predicted reading comprehension relatively well ( $\beta=.31$ ,  $p<.001$ ). Reading comprehension was also slightly predicted by initial reading skills ( $\beta=.14$ ,  $p<.001$ ), which also predicted performance in the mathematical tasks ( $\beta=.15$ ,  $p<.001$ ). The path from initial reading skills to reasoning skills was close to 0 and not statistically significant, and it was therefore removed from the model. Contrary to the expectations, learning-related attitudes only predicted performance in the mathematical tasks ( $\beta=.10$ ,  $p=.001$ ), and the other two paths were removed from the model. As expected, the relationship with mathematical thinking skills was not strong either. There was a moderate correlation between the first grade test factor and earlier reading skills ( $r=.36$ ,  $p<.001$ ) and a weak but statistically significant correlation between learning-related attitudes and reading skills at school start ( $r=.07$ ,  $p<.05$ ).

Next, it was examined whether the structural relationships were similar in the two samples. The factor loadings and all intercepts but one (see above) were constrained equal across groups and the latent mean of the first grade test was allowed to vary freely (there was absolutely no group difference in the latent attitude mean). The model fit the data well (CFI=.990, TLI=.984, RMSEA=.018,  $\chi^2=92.843$ ,  $df=71$ ,  $p=.042$ ). When the structural weights were constrained equal, the fit indices dropped slightly. The very small decrease was caused by the relationship of initial reading skills and later reading comprehension, which seemed to be stronger in Helsinki than in Vantaa ( $\beta=.16$ ,  $p<.001$  vs.  $\beta=.06$ , ns.). Releasing this path increased the fit indices again.

Altogether only 15 % / 14 % (Helsinki / Vantaa) of the variance of third/fourth grade reading comprehension was explained, whereas for mathematical thinking and reasoning skills the shares of explained variance were 43 % / 49 % and 46 %, respectively. It was concluded that hypothesis 1.1. was fully supported regarding mathematical thinking skills, but for reading comprehension and reasoning skills the support was only partial: Reading comprehension was not predicted by learning-related attitudes at all, and the share of explained variance was lower

than expected. For reasoning skills the only – albeit good – predictor was earlier cognitive competence as measured by the first grade test, and earlier reading skills or learning-related attitudes did not explain them at all.

H1.2.: The second hypothesis was that in reading there would be a small gender difference in favour of girls already in the beginning of basic education, but the differences would grow and be visible too already in other areas during the first three school years. However, the evidence regarding gender differences in assessments conducted during later school years is controversial: Traditionally boys have been better in mathematical tasks and girls in reading, but lately girls have outperformed boys in most assessed areas. Girls were also expected to have more positive learning-related attitudes than boys.

To test the hypothesis a dummy-coded gender variable (1=girl) was added to the model. The fit of the model in which the factor loadings and other intercepts but one were constrained equal was good (CFI=.987, TLI=.979, RMSEA=.019,  $\chi^2=112.871$ ,  $df=83$ ,  $p=.016$ ). Constraining the structural paths equal across groups decreased the fit indices slightly, and the best fit was achieved by releasing the constraints from initial reading skills to later reading comprehension as above, and from gender to mathematical thinking (CFI=.983, TLI=.975, RMSEA=.021,  $\chi^2=129.625$ ,  $df=91$ ,  $p=.005$ ).

There was no gender difference in the first grade test for non-verbal cognitive competences, but girls were as expected slightly better readers when they came to school. However, the effect was very small ( $r=.07$ ,  $p<.05$ ). Gender was also related to third grade performance. The effect from gender to reading comprehension was weak ( $\beta =.07$  |  $\beta=.08$  for Helsinki | Vantaa,  $p<.05$ ), but boys' advantage in the third grade mathematical tasks seemed clearer. However, as the best model fit was achieved by not constraining this path equal across the two samples, it was noted that this effect applied only to Helsinki ( $\beta=-.17$ ,  $p<.001$  vs.  $\beta=-.06$ , ns.).

Girls' learning-related attitudes in the third/fourth grade were more positive ( $r=.16$  /  $.17$ ,  $p<.001$ ), and adding gender into the model also increased the effect from learning-related attitudes to mathematical thinking skills from  $.10$  to  $.12/.13$  for Helsinki/Vantaa. This suggests that

girls can cover up the disadvantage in mathematical tasks with positive attitudes and possibly also greater effort (not measured here), which could be an explanation for older girls' superiority in most assessed areas. This made it important to study the development of learning-related attitudes, task interest and effort closer in the following substudies. With the variables used in this substudy, hypothesis 1.2. was supported partially: The results gave support to girls being slightly better readers already when they come to school and – even when controlling for initial differences – they performed better in reading comprehension tasks in the third/fourth grade. The hypothesis was also supported regarding girls' more positive attitudes. It was not supported regarding girls' superiority in all assessed areas: Gender did not predict third/fourth grade reasoning skills, and in mathematical thinking the boys of the Helsinki sample performed better than the girls.

H1.3.: According to the third hypothesis, the cognitive competences measured by the first grade test would together with teacher-evaluated initial reading skills be related to individual support needs as defined officially by the schools. As an indication of educational equity, support needs should not predict third/fourth grade performance as the differences between children in need of support and the others should not increase over time.

The hypothesis was tested by adding a dummy-coded support variable (1=school-defined need of intensified or special support) in the model in which the effects of gender were already taken into account. Support needs turned out to be unrelated to reading comprehension and mathematical thinking, and the paths were removed from the model. The best fit indices were achieved by constraining all the paths from support needs to other variables equal across groups (CFI=.982, TLI=.974, RMSEA=.020,  $\chi^2=150.422$ ,  $df=108$ ,  $p=.004$ ). For Vantaa, 14 % of the variance of reading comprehension, 49 % of mathematical thinking skills and 46 % of reasoning skills were explained with this model whereas for Helsinki the percentages were 17 %, 45 % and 48 %, respectively.

As expected, support needs were strongly related with cognitive competences as measured by the first grade test ( $r=-.42/-0.36$ ,  $p<.001$ ) and also with reading skills at school start ( $r=-.26/-0.24$ ,  $p<.001$ ) for Helsinki/Vantaa. Support needs correlated negatively with third/fourth

grade attitudes ( $r=-.10/-0.09$ ,  $p<.01$ ), indicating that pupils in need of support had somewhat weaker learning-related attitudes in both samples. Support needs were also related to gender but only very weakly ( $r=-.07/-0.06$ ,  $p<.05$ ). Support needs did not directly predict reading comprehension or mathematical thinking, so the differences observed in the beginning of compulsory education had not increased in any of the two samples. However, support needs predicted third/fourth grade reasoning skills positively ( $\beta=.11/.15$ ,  $p<.001$ ), indicating that the pupils in need of support had to some extent actually managed to close the gap with other pupils. As the reasoning items were partially overlapping with the items of the first grade test, this can be interpreted as an indicator of the effectiveness of the support system.

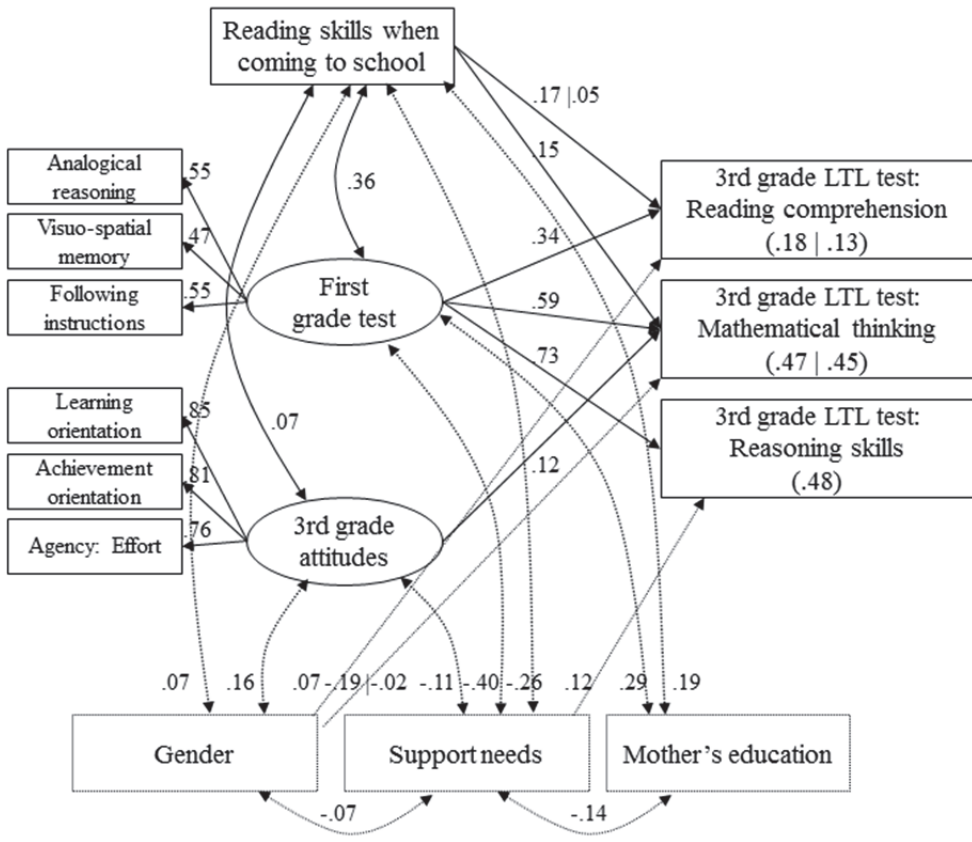
The results supported H1.3. in that the support needs as defined by schools and the first grade skills measured in this study were related and the differences did not increase over time. Surprisingly, support needs even predicted reasoning skills positively, which were interpreted as a sign of a well-functioning support system.

H1.4.: The last hypothesis was that mothers' education would be related to both reading skills and cognitive competences at school start. International literature also suggests that the differences between pupils with mothers of a different educational level would grow over time especially during higher grades, but one of the equity-related goals of the Finnish educational system is that this would not happen. Mothers' education has also been shown to be related to pupils' attitudes at a secondary level and it was expected that the difference would be visible already in third/fourth grade.

Mothers' education was added in the model as a background variable in which 1=basic education, 2= secondary education and 3= tertiary education. Mothers' education did not predict the third grade test scores, and the paths were removed from the model. After that, the best fitting model was that with constraints on all the paths from mothers' education to other variables (CFI=.983, TLI=.976, RMSEA=.018,  $\chi^2=166.149$ ,  $df=126$ ,  $p=.010$ ). For Vantaa, 14 % of the variance of reading comprehension, 48 % of mathematical thinking skills and 46 % of reasoning skills were explained with this model, whereas for Helsinki the percentages were 17 %, 45 % and 48 %, respectively.

The final model is presented in Figure 4.2. The coefficients for Helsinki and Vantaa are separated with a vertical bar, reporting the results of the Helsinki sample first. As expected, mothers' education correlated with both reading skills and cognitive competences in the beginning of first grade, ( $r=.18/.20$  and  $r=.28/.29$ ,  $p<.001$ , respectively) and it was also related to support needs ( $r=-.13/-.10$ ,  $p<.001$ ).

It was concluded that H1.4. was supported regarding the relationships between mothers' education and initial skill differences already in the beginning of basic education. The results supported also the equity principles of the Finnish educational system as the differences between pupils with different educational backgrounds to that of their parents did not increase during the three years of follow-up.



**Figure 4.2.** Predicting performance in the third/fourth grade learning to learn test. The statistically significantly different results of Helsinki | Vantaa are separated with a vertical bar – all the other paths are constrained equal across the samples. Numbers in parentheses indicate the share of accounted for variance.

## 4.5 Conclusions of Substudy 1

This substudy shows that almost half of the variance of ten-year-olds' performance in learning to learn tasks that measure mathematical thinking and reasoning skills can be explained by their initial cognitive competences. Initial reading skills and learning-related attitudes have some additional value, but their effects are very weak compared to the effects of initial cognitive competences. Reading comprehension could be predicted less well with the variables used in this study. In general, the results indicate that the development of childrens' learning to learn skills is not predetermined by their initial competences even though they explain a great share of their variance.

This substudy also shows that when initial competences are controlled for, girls' advantages in reading begin to show already during the first three years of compulsory education. In Helsinki, an opposite gender difference was observed in mathematical thinking. The results also show that girls have more positive attitudes already as early as in the turn of the third and fourth grades.

Regarding the other background factors, this substudy provided evidence for the realisation of equity of education. There were initial differences based on both mothers' education and support needs, but the differences did not increase over time. In contrary, the pupils receiving support even managed to close the gap to some extent in reasoning.





## **5 Substudy 2: Predicting later performance by changes in learning-related attitudes and task interest**

The purpose of substudy 2 is to find out how changes in learning-related attitudes and task interest are related to changes in actual performance in low-stakes assessment tasks in different pupil subgroups. More specifically, the study aims at answering the following research questions:

Q 2.1: How well does 10-years-old pupils' performance in a low-stakes assessment predict their performance in a similar assessment 2.5 years later when their general cognitive competence is taken into account?

Q 2.2: Do changes in learning-related attitudes predict later performance above earlier performance when general cognitive competence is taken into account?

Q 2.3: Do changes in task interest predict later performance above earlier performance when general cognitive competence is taken into account?

Q 2.4: Do changes in learning-related attitudes and task interest have an independent contribution in explaining the variance of later performance above earlier performance and general cognitive competence?

Q 2.5: Are there gender differences in how changes in learning-related attitudes and task interest predict later performance when earlier performance and general cognitive competence are taken into account?

Q 2.6: Are there differences between pupils on different academic achievement levels, including pupils in need of support, in how changes in learning-related attitudes and task interest predict later performance when earlier performance and general cognitive competence are taken into account?

Based on the literature presented in the introduction, and the results of substudy 1, the following hypotheses were set:

H 2.1: Later performance can be relatively well predicted by earlier performance. However, the prediction is not complete due to the effects of education on general cognitive competences (Adey et al., 2007). The measured

competences improve over time, but the development is quite slow (Molnár, Greiff & Csapó, 2013).

H 2.2: A change in learning-related attitudes predicts later performance positively (Renninger & Hidi, 2011). However, the effect is expected to be relatively weak when earlier performance and general cognitive competence are taken into account (Ainley et al., 2002a). The attitude change is expected to be negative from the fourth to sixth grade (Demetriou & Kazi, 2006).

H 2.3: A change in task interest predicts later performance positively (Ainley et al., 2002a; Van Yperen, 2003). The effect is expected to be relatively weak when earlier performance and general cognitive competence are taken into account. As with attitudes, the change is expected to be negative from the fourth to sixth grade (Renninger & Hidi, 2011).

H 2.4: Changes in task interest and learning-related attitudes correlate (Renninger & Hidi, 2011), but they both have an independent contribution in explaining test performance.

H 2.5: As girls are usually more generally learning-oriented (Kenney-Benson et al., 2006) and task-specific interest affects boys' performance more than girls' (Ainley et al., 2002b), the change in learning-related attitudes explain girls' sixth grade LTL test score while boys' performance is explained by the change in task interest.

H 2.6: Academic achievement and support needs are related to the level of performance in the LTL test, and there is a small interaction effect: skillful pupils who do not receive support gain more over time in general thinking skills compared to their lower-performing schoolmates (Shaywitz et al., 1995). More skillful pupils also have more positive attitudes, and their attitudes and task interest decrease less over time.

## 5.1 Participants

Only the Helsinki data were used in this substudy, and of the full dataset only the data from pupils who were present in both the fourth and sixth grade assessments were selected. The Vantaa data were not used due to the switch to the CBA test mode between the two data collections. Of the Helsinki sample, data from both measurement points were available for 883 pupils (52 % girls).

## 5.2 Measures

### 5.2.1 Cognitive tasks

Of the learning to learn assessment battery, only mathematical thinking and reasoning skills tasks were used in Substudy 2, and of those tasks only items which were identical for both age groups were used. It should be noted, however, that the tasks comprised of more items than the ones used in this study as there were also easier items in the fourth grade test version and more difficult items in the sixth grade version. The questions regarding task-specific interest (described below) referred always to the task as a whole, not only to the items used in this study, and this has to be kept in mind when interpreting the results. The tasks and their reliability is presented in Table 5.1. The reliability was acceptable even though it was lower than expected for the fourth grade Arithmetical operations task.

In the mathematical thinking skills test five items were taken from the Mental Arithmetics task and four items from the Hidden Arithmetical Operators task described in Chapter 3. The items were first coded 0-1 for a correct answer, and then the average scores were calculated of the five coded items for Mental Arithmetics and of the four items for Arithmetical Operators. The final scores for mathematical thinking skills were calculated by averaging the two average scores.

Pupils' reasoning skills were measured in both data collection points by the classical Piagetian water-level task described in Chapter 3. An average score was calculated for each pupil of the eight 0-1 -coded items. The final test scores for grades 4 and 6 were calculated by averaging once more the two task type average scores from each grade.

In addition to the assessment tasks common for fourth and sixth graders, the fourth grade test version included the analogical reasoning task as described in Chapter 3. An average of all the eight 0-1 -coded items was used in this study for controlling the effects of general cognitive competence.

**Table 5.1.**Reliability of the cognitive tasks and attitude scales in 2010 and 2013 (Cronbach's  $\alpha$ )

Scale	Number of items	$\alpha$ 2010	$\alpha$ 2013
Arithmetical Operations	4	.52	.60
Mental Arithmetics	5	.65	.64
Reasoning skills (Bottles task)	8	.89	.81
Analogical Reasoning	8	.78	
Learning orientation	3	.85	.80
Achievement orientation	2	.66	.76
Agency: Effort	3	.77	.79

### 5.2.2 Learning-related attitudes

Pupils' learning-related attitudes were measured by the scales described in detail in Chapter 3: Learning Orientation (e.g. "An important goal for me at school is to learn new things."), Achievement Orientation (e.g. "Getting good grades at school is important to me.") and Agency: Effort (e.g., "I work hard to do well at school."). All the items in the scales were answered with a 7-point Likert-scale in which 1= Not true at all and 7= Very true. The reliabilities of the scales for both age groups are presented in Table 5.1. The attitude change scores for each of the subscales were calculated by first averaging the three-item scales for each year, and then subtracting the fourth grade average from the sixth grade average. Thus, a positive change score indicated a positive change.

### 5.2.3 Task interest

After each task the pupils were asked to evaluate the task and themselves when doing it with three statements by using the 7-point Likert-scale described above. Of these, the statement "The task was very interesting" was used in this study. The task interest scale comprised of this statement presented after each of the three cognitive tasks. The task interest change scores for each of the variables were calculated by subtracting the fourth grade response from the sixth grade response. Thus, a positive change score indicated a positive change.

### 5.2.4 Grouping variables

The background variables used in this study were extracted from teacher evaluations except for self-reported gender. Teachers evaluated their pupils' school achievement in four central school subjects (mother tongue, mathematics, English, science). The scale was that of the normal school grades in Finland, ranging from 4 (failed) to 10 (excellent). An average score was calculated of the fourth grade school grades, and the pupils were divided into four equally sized groups based on these (low performers  $< 7.25 \leq$  lower average  $< 8.25 \leq$  higher average  $< 9.00 \leq$  high performers). School grades were available for 94 % of the fourth grade pupils. In the sixth grade, teachers also reported if the pupil had received intensified or special support for studies. In analyses regarding support, the pupils were divided into two groups: No support needs (N=693) and Support needs (N=109). The information regarding support was missing for 9.2 % of pupils.

### 5.3 Statistical methods

Paired samples t-test and repeated measures GLM were used in SPSS18 for analysing the effects of change in performance, attitudes and task interest in different subgroups. For analysing the role of attitude and task interest change in explaining performance structural equation modeling (SEM) was used in AMOS21. Since the deviation from normality of all variables was small (skewness and kurtosis between -1 and 1), maximum likelihood estimation was used (see Kline, 2005). The models were considered to have a good fit with CFI and TLI  $> .95$  and RMSEA  $< .08$ . Also reported were  $\chi^2$  values, but due to the large sample size significant p-values were to be expected. Therefore, they are not considered as an absolute criterion for model fit. Measurement invariance was tested by adding stepwise constrains to the measurement models, first by constraining factor loadings (weak factorial invariance) and then intercepts (strong factorial invariance) across groups while letting the latent mean of the others but the reference group vary free (Byrne & Stewart, 2006).

## 5.4 Results

The descriptive statistics for all the variables for the whole sample are presented in Table 5.2 and by groups in Tables 5.3 and 5.4. Fourth grade descriptives are presented on the left side of the vertical bars, sixth grade descriptives on the right. The change variables applicable are presented under each variable.

**Table 5.2.**

Descriptive statistics for the variables used in the model

Variable	N	Min	Max	M	Sd
Learning to learn test score fourth   sixth grade	882   883	0.00	1.00	0.44   0.64	0.23   0.21
Analogical reasoning fourth grade	844	0.00	1.00	0.50	0.31
Learning orientation	839   874	1.00	7.00	5.86   5.42	1.19   1.13
Learning orientation: Change	831	-4.67	5.00	-0.41	1.38
Achievement orientation	839   868	1.00	7.00	6.06   5.66	1.11   1.20
Achievement orientation: Change	826	-5.50	6.00	-0.39	1.45
Agency: Effort	839   873	1.00	7.00	5.77   5.27	1.10   1.14
Agency: Effort: Change	830	-4.33	6.00	-0.47	1.38
Task interest: Arithmetical Operations	795   796	1.00	7.00	3.99   2.76	2.09   1.89
Task interest: Arithmetical Operations: Change	721	-6.00	6.00	-1.21	2.51
Task interest: Mental Arithmetics	780   822	1.00	7.00	3.26   2.94	2.02   1.70
Task interest: Mental Arithmetics: Change	734	-6.00	6.00	-0.33	2.24
Task interest: Reasoning task	761   840	1.00	7.00	4.28   3.93	2.07   2.71
Task interest: Reasoning task: Change	728	-6.00	6.00	-0.34	2.71

The values for the fourth |sixth grade are separated with a vertical bar

N= Number of responses, Min= minimum value, Max=maximum value, M=Mean, Sd=Standard deviation

**Table 5.3.**

Descriptive statistics by gender for the variables used in the models

Variable	Girls		Boys	
	M	Sd	M	Sd
Learning to learn test score fourth   sixth grade	0.41   0.60	0.17   0.18	0.44   0.59	0.18   0.17
Analogical reasoning fourth grade	0.51	0.31	0.48	0.30
Learning orientation	5.94   5.42	1.10   1.11	5.78   5.44	1.29   1.15
Learning orientation: Change	-0.53	1.29	-0.29	1.46
Achievement orientation	6.19   5.73	0.99   1.14	5.92   5.58	1.21   1.27
Achievement orientation: Change	-0.45	1.38	-0.32	1.52
Agency: Effort	5.91   5.36	0.98   1.12	5.62   5.17	1.22   1.16
Agency: Effort: Change	-0.52	1.29	-0.41	1.48
Task interest: Arithmetical Operations	4.03   2.75	2.00   1.85	3.93   2.77	2.21   1.93
Task interest: Arithmetical Operations: Change	-1.27	2.33	-1.12	2.71
Task interest: Mental Arithmetics	3.28   2.84	1.94   1.56	3.28   3.06	2.11   1.84
Task interest: Mental Arithmetics: Change	-0.40	2.14	-0.26	2.35
Task interest: Reasoning task	4.53   4.12	2.17   2.01	4.02   3.71	2.37   2.12
Task interest: Reasoning task: Change	-0.41	2.60	-0.26	2.84

The values for the fourth | sixth grade are separated with a vertical bar

**Table 5.4.**

Descriptive statistics by support needs for the variables used in the models

Variable	No support needs		Support needs	
	M	Sd	M	Sd
Learning to learn test score fourth   sixth grade	0.44   0.61	0.17   0.17	0.32   0.44	0.17   0.17
Analogical reasoning fourth grade	0.52	0.30	0.34	0.28
Learning orientation	5.86   5.42	1.18   1.12	5.67   5.30	1.44   1.29
Learning orientation: Change	-0.23	1.34	-0.33	1.74
Achievement orientation	6.09   5.66	1.06   1.20	5.73   5.42	1.47   1.29
Achievement orientation: Change	-0.41	1.40	-0.23	1.82
Agency: Effort	5.80   5.30	1.06   1.11	5.52   4.95	1.43   1.36
Agency: Effort: Change	-0.48	1.32	-0.51	1.82
Task interest: Arithmetical Operations	3.90   2.72	2.07   1.85	4.05   2.97	2.39   2.17
Task interest: Arithmetical Operations: Change	-1.18	2.48	-1.04	3.00
Task interest: Mental Arithmetics	3.21   2.92	1.99   1.38	3.33   2.87	2.30   1.92
Task interest: Mental Arithmetics: Change	-0.26	2.23	-0.62	2.52
Task interest: Reasoning task	4.27   3.87	2.25   2.03	4.13   3.85	2.44   2.25
Task interest: Reasoning task: Change	-0.35	2.68	-0.41	3.14

The values for the fourth | sixth grade are separated with a vertical bar

H2.1: The first hypothesis was that later performance in the cognitive tasks could be relatively well predicted by earlier performance but that the development would be relatively slow due to the cross-curricular nature of the assessed competences. Fitting a simple regression model on the whole data showed that the fourth grade LTL test score was as expected a good predictor of the sixth grade LTL test score ( $\beta=.53$ ), and alone it explained 28 % of the variance of the sixth grade score. By adding fourth grade analogical reasoning skills in the model as a control variable, the share of the explained variance increased to 34 %. As expected, there was a relatively strong correlation between analogical reasoning skills and the fourth grade LTL test score ( $r=.42$ ,  $p<.001$ ). In this model, both the fourth grade LTL test score ( $\beta=.42$ ), and analogical reasoning skills ( $\beta=.27$ ) had a relatively strong independent contribution in explaining the sixth grade LTL test score.

Before testing further hypotheses by adding more variables in the model, or dividing the sample to subgroups, it was first tested by paired samples t-test whether the pupils' sixth grade LTL test score differed from their fourth grade score. The sixth grade mean of .64 for the whole sample was statistically significantly ( $t=-27.967$ ,  $p<.001$ ) higher than the fourth grade mean of .44. It was concluded that H2.1 was supported and that it was meaningful to test the further hypotheses regarding what explains the change beyond cognitive development.

H2.2: The second hypothesis was that a change in learning-related attitudes would predict later performance positively but that the effect would be relatively weak when earlier performance and general cognitive competence are taken into account. Moreover, the attitude change was expected to be negative from the fourth to sixth grade. Paired samples t-test showed that for the whole sample the change was negative and statistically significant in all the three motivational scales (Learning orientation:  $t=-8.663$ ,  $p<.001$ ; Achievement orientation:  $t=7.734$ ,  $p<.001$ ; Agency: Effort:  $t=9.774$ ,  $p<.001$ ). Thus, it was meaningful to use the attitude change scores in structural equation modeling when explaining the sixth grade LTL test score.

For structural equation modeling the attitude change scores of the three motivational scales were regressed to a first order latent factor of Attitude change. The latent factor was added to the model specified for

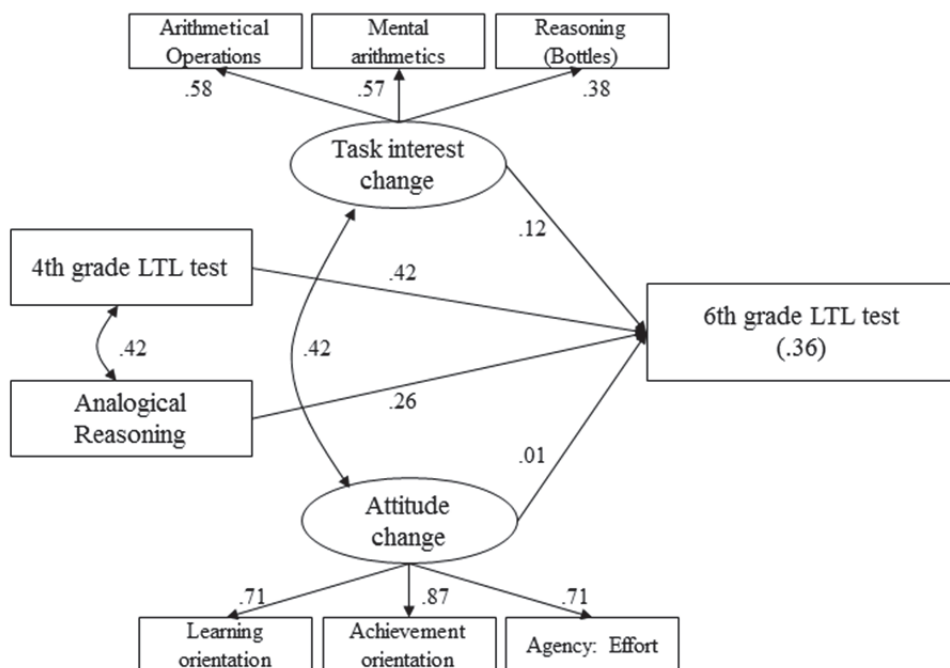


testing H1 as a third predictor of the sixth grade LTL test score. The model fit was good (CFI=.997; TLI=.991; RMSEA=.025;  $\chi^2=12.721$ ,  $df=8$ ,  $p=.122$ ). Attitude change was uncorrelated with the fourth grade LTL test score and the fourth grade analogical reasoning skills. Moreover, it predicted the sixth LTL test score only marginally ( $\beta=.06$ ,  $p=.07$ ), decreasing the path coefficient from analogical reasoning from  $\beta=.27$  to  $\beta=.26$ . It was concluded that H2.2 was supported regarding the decrease of learning-related attitudes from grade 4 to 6 but not regarding the expected effect of attitude change on the sixth grade LTL test score.

H2.3: In the third hypothesis it was expected that a change in task interest would predict later performance positively but that the effect would be relatively weak when earlier performance and general cognitive competence are taken into account. As with attitudes, the change was expected to be negative from the fourth to sixth grade. Paired samples t-test showed that for the whole sample the change in task interest was negative and statistically significant in all the three cognitive tasks (Arithmetical operations:  $t=12.892$ ,  $p<.001$ ; Mental arithmetics:  $t=4.035$ ,  $p<.001$ ; Reasoning skills:  $t=3.377$ ,  $p<.001$ ). Thus, it was meaningful to use the task interest change scores in structural equation modeling when explaining the sixth grade LTL test score.

For structural equation modeling the task interest change scores for the three cognitive tasks were regressed to a first order latent factor of task interest change. Like with learning-related attitudes, the latent factor was added to the model specified for testing H1 as a third predictor of the sixth grade LTL test score. The model fit was good (CFI=.985; TLI=.960; RMSEA=.036;  $\chi^2=17.909$ ,  $df=8$ ,  $p=.022$ ). Change in task interest was also uncorrelated with the fourth grade LTL test score and the fourth grade analogical reasoning skills. The change in task interest predicted the sixth grade LTL test score quite weakly, but statistically significantly ( $\beta=.14$ ,  $p<.001$ ), and it increased the share of the explained variance of the sixth score from 34 % to 36 %. Despite being uncorrelated with the two other predictors in the model, task interest decreased the path coefficient from analogical reasoning to the sixth grade LTL test score from  $\beta=.27$  to  $\beta=.26$ . It was concluded that H2.3 was fully supported.

H2.4: The fourth hypothesis was that changes in task interest and learning-related attitudes would correlate, but they would both have an independent contribution in explaining the sixth grade LTL test score. However, as the change in learning-related attitudes turned out to be a poor predictor of the sixth LTL test score regarding H2.2., it was unlikely that it would explain any unique variance of the sixth grade score in the model with both latent change variables simultaneously. The model was nevertheless specified in order to understand the relationship of the two latent change variables, and to test possible subgroup differences in sections H2.5. to H2.7. The complete model, fitted on the whole sample is presented in Figure 5.1. The model fit was good (CFI=.993; TLI=.998; RMSEA=.021;  $\chi^2=34.662$ ,  $df=24$ ,  $p=.074$ ).



**Figure 5.1.** Predicting performance in the sixth grade learning to learn test by fourth grade performance in the same items and changes in task interest and attitudes, controlling for analogical reasoning skills in fourth grade. Numbers in parenthesis indicate the share of accounted for variance.

Figure 5.1 shows that there was a relatively strong correlation between the changes in task interest and learning-related attitudes ( $r=.42$ ,  $p<.001$ ). However, having the two latent change variables in the model

simultaneously did not increase the share of explained variance of the sixth grade LTL test score compared to the model in which only task interest was used. Moreover, it can be seen that the effect of the two correlating latent factors went completely via the change in task interest, and the direct effect of Attitude change on the sixth grade LTL test score was even smaller than it was without having task interest in the model. However, since the direct effect of Attitude change on the sixth LTL test score was not statistically significant in H2.2, the significance of the indirect effect was not specifically tested either. It was concluded that H2.4 was supported regarding the relationship between changes in task interest and learning-related attitudes but not regarding the unique contribution of both of them in explaining the sixth grade LTL test score.

H2.5: In the fifth hypothesis it was assumed that a change in learning-related attitudes would explain the girls' sixth grade LTL test score while for boys the changes in task-specific interest would have more explanatory power. It was first tested whether there were significant gender differences in how attitudes and task interest had changed from the fourth to sixth grade. Repeated measures GLM indicated that girls' learning-related attitudes were on a higher level in the fourth grade, but of the individual attitude scales in Learning orientation there was a small interaction effect (Wilks' lambda=.993,  $F=6.219$ ,  $p<.05$ ), which was the main reason that the gender difference for learning-related attitudes was no longer statistically significant in the sixth grade. Thus, girls' learning-related attitudes decreased slightly more than boys' attitudes. There were no statistically significant gender differences in the change of task interest, but girls were more interested in the reasoning task at both data collection points ( $M=4.53/4.12$  (girls) and  $M=4.02/3.70$  (boys),  $t=3.215/3.046$ ,  $p<.01$ ).

Next, it was tested if girls and boys had improved their performance in the LTL test in a similar way from the fourth to sixth grade. In the fourth grade there was a small gender difference favouring boys ( $t=-2.53$ ,  $p<.05$ ) while in the sixth grade there was no difference. Repeated measures GLM showed that the interaction was statistically significant (Wilks' lambda=.994,  $F=5.013$ ,  $p<.05$ ), indicating that the girls closed the gap in test performance during the follow-up time.

Finally, the SEM model specified in H2.4 was fitted simultaneously on girls and boys. To enable group comparisons, measurement invariance of the latent factors was first tested separately for the measurement models. To have sufficient degrees of freedom, this was done simultaneously for both measurement models, allowing them to correlate with each other. The fit indices were excellent (CFI=1.000; TLI=1.000; RMSEA=.000;  $\chi^2=6.219-8.265$ ,  $df=16-24$ ,  $p=.986-.999$ ) for the models: the unconstrained baseline model, the one with factor loadings constrained equal across groups and the one with equal intercepts. Even strict measurement invariance (measurement residuals constrained equal, see Byrne & Stewart, 2006) held here (CFI=1.00; TLI=1.00; RMSEA=.003;  $\chi^2=32.278$ ,  $df=32$ ,  $p=.453$ ). Moreover, there were no statistically significant differences in the latent means of the two factors, so the small interaction observed in GLM did not show here.

Next, the whole model of H2.4. was fitted simultaneously on girls and boys. The model with paths constrained equal fitted the data excellently (CFI=1.00; TLI=.999; RMSEA=.004;  $\chi^2=62.740$ ,  $df=62$ ,  $p=.450$ ), and it was concluded that the paths presented in Figure 5.1. were similar for both girls and boys. In other words, for both genders the change in task interest predicted the sixth grade LTL test score beyond earlier performance and analogical reasoning skills and the change in learning-related attitudes – despite the relatively strong correlation with the change in task interest – did not predict sixth grade performance at all.

As there was a gender difference in favour of boys in the fourth grade tasks, constraining structural means equal caused the model fit indices to drop to some extent. Therefore, the best fit for the whole model (CFI=.999; TLI=.999; RMSEA=.004;  $\chi^2=65.020$ ,  $df=64$ ,  $p=.441$ ) was achieved when the mean of the fourth grade test score was allowed to vary across group. It was concluded that H2.5 was not supported, neither regarding girls' higher learning-related attitudes (except in the fourth grade) and boys' higher task-specific interest, nor regarding the gender differences in the relative role of these in explaining the sixth grade LTL test score.

H2.6: In the sixth hypothesis academic achievement and support needs were expected to be related to the level of performance in the LTL test both in the fourth and sixth grade. Small interaction effects were

expected: that skillful pupils who do not receive support would gain more over time compared to their lower-performing schoolmates and to those who receive support. Moreover, it was expected that the level of school achievement and support received would be related to the level of motivational attitudes and task interest.

First, it was tested whether there were differences between the four school achievement groups, and pupils receiving support for studies vs. those who did not, in how attitudes and task interest had changed from the fourth to sixth grade. Repeated measures GLM indicated that there were statistically significant differences between school achievement groups in Achievement orientation and Agency – Effort ( $F=5.055/6.875$ ,  $p<.001$ , respectively) so that higher performing pupils reported also higher achievement orientation and effort, but there was no interaction in how the attitudes developed over time in different groups. In Learning orientation the groups did not differ from each other. When pupils receiving support for studies were compared to those who did not receive support, there were statistically significant differences only in Learning orientation. In the fourth grade pupils who did not receive any support reported significantly higher Learning orientation but their results declined to the level of their classmates who received support in the sixth grade. The interaction was statistically significant (Wilks' lambda=.993,  $F=5.498$ ,  $p<.05$ ). Task interest had developed in a similar way in all school achievement groups, and there were no statistically significant group differences in the levels of task interest. The development of task interest was also similar for pupils receiving support and for those who did, not but pupils who received support reported – a little surprisingly – higher interest in the Arithmetical operations task at both data collection points ( $F=4.813$ ,  $p<.05$ ).

Next it was tested if pupils in the four school achievement groups, and those who received support vs. those who did not, had improved their performance in the LTL test in a similar way from the fourth to sixth grade. As expected, the school achievement groups differed from each other also in the LTL test ( $F=119.051$ ,  $p<.001$ ), and there was a small but statistically significant interaction too (Wilks' lambda= .988,  $F=3.301$ ,  $p<.05$ ): the differences between groups increased slightly over time. The interaction was much stronger when pupils who received support were

compared to those who did not (Wilks' lambda=.977,  $F=18.733$ ,  $p<.001$ ): pupils who received support performed significantly lower in the LTL test ( $F=80.842$ ,  $p<.001$ ) and the difference increased over time.

Next, the SEM model specified in H2.4 was fitted simultaneously on the four school achievement groups. To enable group comparisons measurement invariance was first tested for the two factors separately. Constraining measurement weights (CFI=.992, TLI=.985, RMSEA=.014,  $\chi^2=51.310$ ,  $df=44$ ,  $p=.209$ ) and then measurement intercepts (CFI=.995, TLI=.993, RMSEA=.010,  $\chi^2=60.757$ ,  $df=56$ ,  $p=.309$ ) did not decrease model fit indices much compared to the baseline model (CFI=1.000, TLI=1.000, RMSEA=.000,  $\chi^2=29.583$ ,  $df=32$ ,  $p=.589$ ), so the model was concluded as being measurement invariant. There were also no statistically significant differences in the latent means of the two factors.

When the whole model was fitted on the school achievement groups simultaneously, the fit indices varied between CFI=.964-.977, TLI=.947-.957, RMSEA=.018-.020,  $\chi^2=124.742-182.507$ ,  $df=126-138$ ,  $p=.007-.026$  when more constrains were added gradually, the lowest figures referring to the model in which also structural weights were constrained equal. It was concluded that the path coefficients were not statistically significantly different for different groups, but, as could be expected based on the repeated measures GLM results, constraining structural means resulted in an extremely poor model fit. Thus, there were statistically significant group differences in their performance level but not in how it was related to the changes in attitudes or task interest. In other words, for all school achievement groups the change in task interest predicted the sixth grade LTL test score in addition to earlier performance and analogical reasoning skills, and the change in learning-related attitudes – despite the relatively strong correlation with the change in task interest – did not predict the sixth grade performance at all.

Finally, the model was simultaneously fitted on pupils receiving support and those who did not. The measurement models were concluded as being measurement invariant separately by constraining factor loadings and intercepts equal, and constraining structural weights equal did not decrease model fit indices either from CFI and TLI=1.000; RMSEA=.000;  $\chi^2=34.145-56.699$ ,  $p=.516-.617$ . Constraining structural

means equal, however, decreased the fit indices to CFI=.983; TLI=.976; RMSEA=.021;  $\chi^2=93.957$ ,  $df=64$ ,  $p=.009$ . The results pointed in the same direction as the statistically significant GLM results: pupils receiving support did not perform as well as their classmates who did not receive any support, but there were no group differences in how changes in attitudes or task interest predicted performance.

It was concluded that H2.6 was supported regarding the differences in the LTL test score, both between the four school achievement groups and between pupils receiving support vs. those who did not receive it. The expected small interaction effects were also found: skillful pupils and those who did not receive support gained more over time compared to their lower-performing schoolmates and to those who received support. However, in regard to the group differences in the level of motivational attitudes and task interest, H2.6 was only supported on the level of some of the individual scales. Belonging to different school achievement groups was related to Achievement orientation and self-reported effort while receiving support was related to higher interest in one of the tasks and to lower Learning orientation in the fourth grade.

## 5.5 Conclusions of Substudy 2

As a conclusion of all the hypotheses and results presented above it can be summarised that when earlier performance in similar tasks and analogical reasoning skills are taken into account, changes in learning-related attitudes alone do not later affect the LTL test score of girls and boys, of pupils at different levels of school achievement or of pupils who need support for their studies. Changes in task interest, however, predict later performance over and above earlier performance and analogical reasoning skills for all these subgroups – even if the level of both task interest and learning-related attitudes decline quite evenly in all subgroups from the fourth to sixth grade. However, even though changes in learning-related attitudes do not affect performance directly, they are strongly related to changes in task-specific interest, and they are therefore to be taken seriously when discussing the factors that are possibly affecting the results of low-stakes assessments.





## **6 Substudy 3: Gender differences, support needs, learning-related attitudes and time on task**

The third substudy concentrates on pupils' behaviour in the assessment situation, looking at how time investment is related to attitudes and performance. The aim of the study is to not only replicate with younger participants the study of Kupiainen and colleagues (2014) regarding the mediating effect of time investment on performance but also to take the next steps of looking at how gender and support needs affect the relationships of learning-related attitudes, time investment and performance. Research on this relatively new area of interest is of great importance as it might provide at least partial answers to the unexplained growth of gender differences and help in understanding why the gap between pupils in need of support and the others often increase despite provided support.

The exact research questions of Substudy 3 are:

Q 3.1.: Does time on task mediate the effects of sixth graders' mastery and detrimental attitudes and prior ability on the LTL test score as it did for three years older pupils in the study of Kupiainen et al. (2014)? Having longitudinal data available, is GPA the best indicator of prior ability in the model or should a measure of general cognitive competence be used instead?

Q 3.2.: Are gender differences in low-stakes assessment scores at least partially explained by differences in time on task and mastery and detrimental attitudes?

Q 3.3.: Do support needs affect the relationship between prior ability, learning-related attitudes, time on task and test score?

Q 3.4.: How do support needs and gender together change the relationships between prior ability, attitudes, time on task and the test score?

Based on theories presented in the introduction the following hypotheses were set:

H 3.1.: Time on task mediates the effects of detrimental attitudes, but for mastery attitudes and prior ability the mediation is very weak and only partial (Kupiainen et al., 2014). A measure of general cognitive competence instead of GPA makes the interpretation of the results easier as an external measure is freer of the effects of attitudes used in the same model.

H 3.2.: Girls have more mastery attitudes and less detrimental attitudes (Kenney-Benson et al., 2006; see also Huesmann et al., 2009; Olweus, 1979) and a higher test score (OECD, 2013; Hautamäki et al., 2013) even though there is no gender difference in girls' and boys' general cognitive competence (Halpern, 2000; see also Substudy 1). More positive attitudes make girls invest more time on assessment tasks which is then reflected in their better performance.

H 3.3.: Support needs increase the time needed (cf. Carroll, 1963), and they are related to lower prior ability (cf. Substudies 1 and 2). However, based on the principles of equity in the Finnish educational legislation it is not expected that support needs would predict the sixth grade test score when prior ability is taken into account. Support needs are however often related to more negative attitudes (Thuneberg, 2007), and an indirect effect through attitudes is expected.

H 3.4.: More boys than girls are usually identified as having support needs (Henning-Stout & Close-Conoley, 1992; Thuneberg, 2007; see also Substudy 1), and a negative correlation between support needs and gender is expected. However, both background variables are expected to have independent effects on the other variables.

## 6.1 Participants

The participants of this substudy were the 80 % of pupils from the Vantaa sixth grade cohort who completed the computer-based version of the LTL test in spring 2013 (N=1543, 49.2 % girls). Longitudinal data comprising also the third grade results in analogical reasoning were available for 1303 of them.

## 6.2 Measures

### 6.2.1 Cognitive tasks

Since this study had primarily a cross-sectional approach, all the available data for sixth graders' cognitive LTL competences were used. Thus, both reading comprehension measures, both mathematical tasks used in CBA (the Mental Arithmetics task was only in the PBA version and therefore excluded from this study) and the three reasoning tasks (again, the water-level task was only in PBA and could not be used here) were used. Instead of specifying a latent test score factor like in the earlier study of Kupiainen et al. (2014), test score was used as a manifest variable in path modelling. The reason for using manifest variables in modelling instead of latent factors was smaller sample size compared to the earlier study, and the increased number of paths to be estimated due to the additional background variables. This must be kept in mind when comparing the results of this study to the earlier study, as well as the fact that the sixth grade items were not identical to those used in the ninth grade test version the year before.

Test score was calculated for all the pupils in two stages. First the scores of the two reading comprehension tasks were summarised into one score, consisting of the scores of the 20 items together. The verbal reasoning tasks were summarised into one task consisting of 10 items. Then all the task-specific scores for each task were transformed to percentages of correctly solved items in order to bring all the measures together in the same scale. In the last stage an average percentage of correctly solved items was calculated based on the task-specific percentages. The reliability of the separate tasks is presented in Table 6.1. In the table the reliability of the additional cognitive task of third grade analogical reasoning skills, used in controlling for prior ability, is also displayed. All the reliabilities were acceptable.

### 6.2.2 Learning-related attitudes

In this study learning-related attitudes were understood to consist of mastery and detrimental attitudes which are not necessarily opposite extremes of the same continuum (see Kupiainen et al., 2014). Mastery

attitudes were measured by three scales: Mastery Extrinsic Orientation (Achievement Orientation); Agency: Effort; and Importance of School, which are all described in Chapter 3. The descriptions of the three detrimental scales – Means-ends: Ability; Means-ends: Chance; and Self-handicapping are also in Chapter 3. The reliability of the scales is presented in Table 6.1. For mastery attitudes the reliability was very good and for the detrimental attitudes acceptable.

**Table 6.1.**  
Reliability of the cognitive tasks and the attitude scales

Scale	Number of items	$\alpha$
<b>Sixth grade LTL test</b>	51	.83
Reading comprehension	20	.63
Arithmetical Operations	7	.60
Mathematical Concepts	8	.63
Verbal proportional reasoning	10	.55
Control of variables	6	.69
Grade point average (GPA)	4	.88
Analogical reasoning third grade	8	.78
<b>Mastery Attitudes</b>	9	.89
Mastery: Extrinsic	3	.87
Agency: Effort	3	.79
Importance of School	3	.86
<b>Detrimental Attitudes</b>	9	.75
Means-ends: Ability	3	.60
Means-ends: Chance	3	.67
Self-handicapping	3	.68

In regard to the test score, also the attitude scales were used as manifest variables in path modelling instead of latent factors in SEM. The mastery and detrimental attitude scores were calculated by first averaging the three items for each scale, and then averaging the three average scores for each scale for mastery and detrimental attitudes.

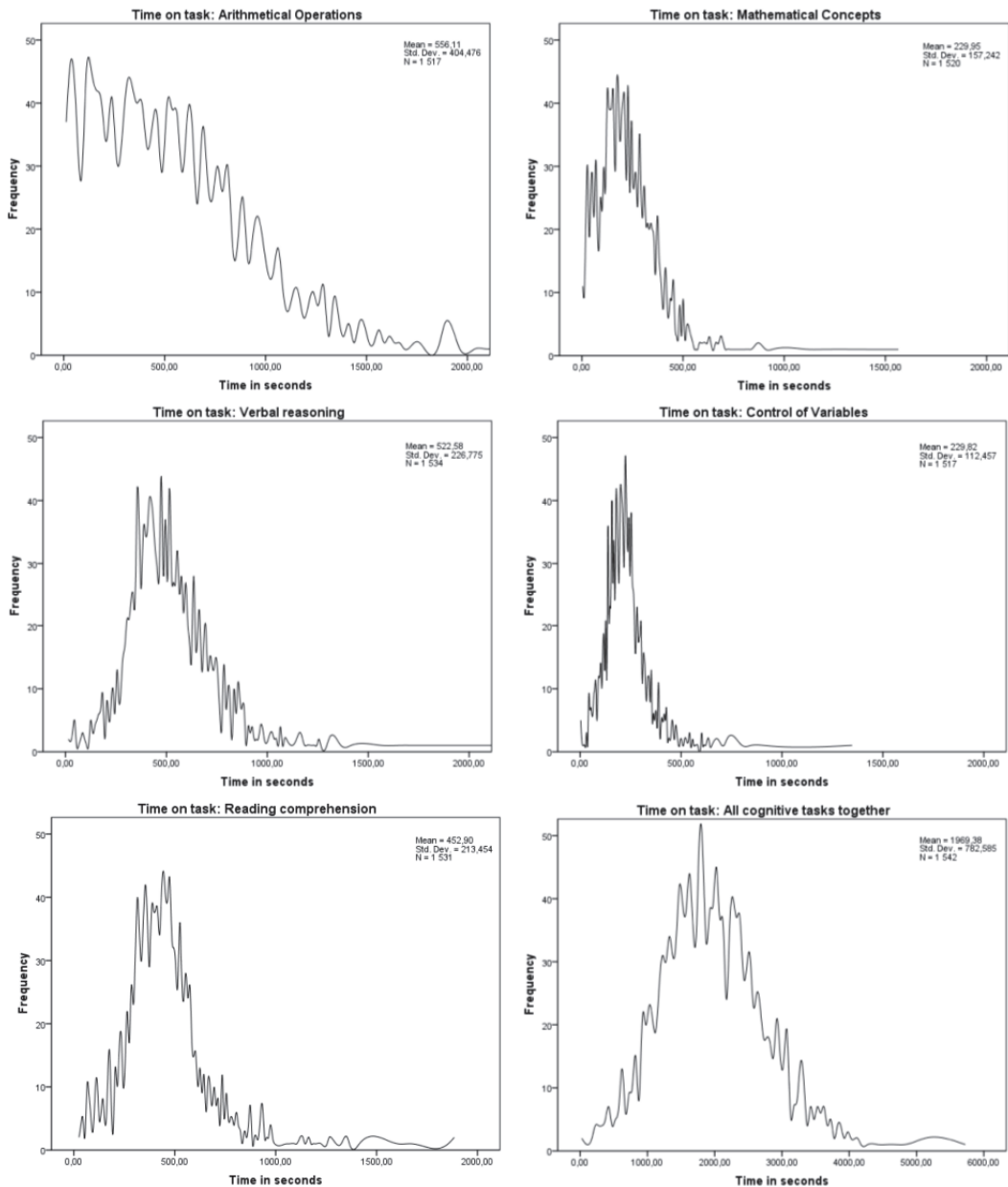
### 6.2.3 Time on task

Time on task was extracted from the log files of CBA. The system registered the time each pupil opened a task and the time of submitting it. Time was counted in seconds for the whole task comprising the instructions, one or more pre-solved example items and all the items to solve. Thus, the definition of time on task here and in the earlier study by Kupiainen et al. (2014) differs for instance from the item-specific one of Goldhammer et al. (2014) and the notion of response time of Wise and colleagues (see Wise, 2006) – all of them also concentrating on log file analyses. As with other measures, no latent factor for time on task was specified like in Kupiainen et al. (2014), and the task-specific time variables were simply summarised to a Time on task variable for all cognitive tasks together. Before summarising the scores, possible outliers were examined by graphical inspection (see Kupiainen et al., 2014) as there was a small possibility of pupils forgetting to submit the task before moving to the next one and later returning to do the submission, but no clear outliers were detected in any of the tasks. Moreover, the time distributions were somewhat different from what is usual for time variables (which is a high peak in the beginning and a long tail with longer response times, see for instance Wise & Kong, 2005), and even though the task-specific time variables were not normally distributed the total time for all the cognitive tasks was. Therefore, unlike in Kupiainen et al. (2014), no logarithmic transformations were needed. The distributions of the time variables are presented in Figure 6.1.

### 6.2.4 Background variables

For the first model using grade point average (GPA), pupils' success in the most important school subjects was extracted from the teacher evaluations as described in Chapter 3. For testing hypotheses two to four pupils' self-reported gender and teacher-reported support needs were used in the models. Gender was used in the models as a dummy coded variable in which 0= Boy (N=784) and 1= Girl (N=759). The teachers reported whether each pupil had received intensified or special support, and for the purposes of this study those two categories were recoded into one dummy-variable with 0= No support needs (N=1184) and 1= Support

needs (N=221). The information regarding support needs was missing for 138 pupils (9 %).



**Figure 6.1.** The distributions of time on task for Arithmetical Operations, Mathematical Concepts, Verbal reasoning, Control of Variables, Reading comprehension, and for all cognitive tasks together.

### 6.3 Statistical methods

SPSS18 was used for performing basic statistical analyses and for studying the distributions of the time variables. For all the other analyses, structural equation modelling (SEM) was performed in AMOS21. The deviation from normality of all variables in this substudy was small (skewness and kurtosis between -1 and 1), and maximum likelihood estimation was used (see Kline, 2005). The models were considered as having a good fit with CFI and TLI > .95 and RMSEA < .08. Also reported are  $\chi^2$  values, but due to the large sample size significant p-values were to be expected. Therefore they are not considered as an absolute criterion for model fit. Besides analysing the direct effects as in earlier substudies, the mediation hypotheses were tested by studying indirect effects. According to Zhao, Lynch & Chen (2010) mediation can be equated with an indirect effect: if the direct effect is then not significant, the mediation is full. If the direct and the indirect effects are statistically significant and they both are positive/negative, the mediation is partial (Zhao et al., 2010). In that case the direct effect between the independent and dependent variable decreases after the mediator variable is added into the model (MacKinnon et al., 2000). Testing the significance of the indirect effects would have required producing confidence intervals with a bias-corrected bootstrap method (see Cheung & Lau, 2008), but that was unfortunately not possible to test with Amos21 when having data with occasional missing values. Therefore, the interpretation of the results of the indirect effects must be done with reservations

### 6.4 Results

The descriptive statistics for all the variables used in path modelling are presented in Table 6.2. and by groups in Tables 6.3. and 6.4. The statistics are calculated only for pupils who completed the CBA version of the LTL test.

**Table 6.2.**

Descriptive statistics of the variables used in structural equation modelling

Variable	N	Min	Max	M	Sd
LTL Test score	1540	2.50	85.20	39.95	15.50
Mastery attitudes	1530	1.00	7.00	5.33	0.98
Detrimental attitudes	1515	1.00	7.00	3.40	0.93
Grade point average (GPA)	1435	4.75	10.00	7.93	1.01
Analogical reasoning	1303	0.00	100.00	40.19	29.88
Time on task: All cognitive tasks	1542	19.00	5707.00	1969.38	782.58

N= Number of responses, Min= minimum value, Max=maximum value, M=Mean, Sd=Standard deviation

**Table 6.3.**

Descriptive statistics by gender for the variables used in the models

Variable	Girls		Boys	
	M	Sd	M	Sd
LTL Test score	41.34	14.84	38.58	16.00
Mastery attitudes	5.42	0.93	5.24	1.01
Detrimental attitudes	3.37	0.89	3.43	0.96
Grade point average (GPA)	8.07	0.98	7.79	1.01
Analogical reasoning	41.46	29.55	38.96	30.16
Time on task: All cognitive tasks	2066.06	751.41	1875.94	801.04

**Table 6.4.**

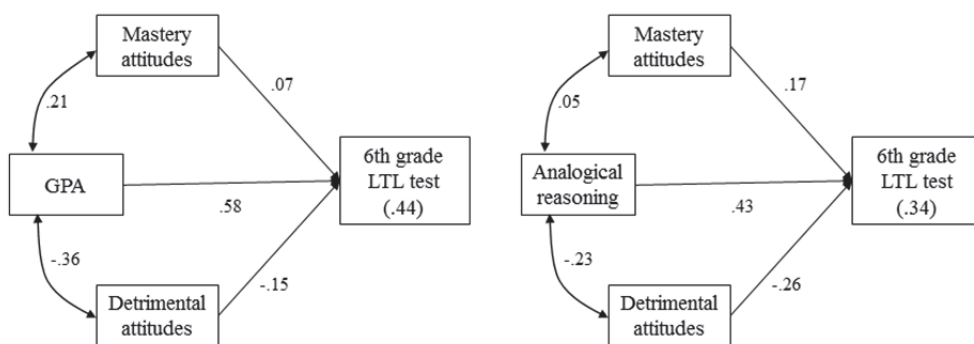
Descriptive statistics by support needs for the variables used in the models

Variable	No support needs		Support needs	
	M	Sd	M	Sd
LTL Test score	42.39	14.93	28.13	11.63
Mastery attitudes	5.37	0.94	5.13	1.17
Detrimental attitudes	3.32	0.90	3.86	0.94
Grade point average (GPA)	8.13	0.89	6.79	0.86
Analogical reasoning	42.56	30.21	25.27	21.55
Time on task: All cognitive tasks	2016.20	732.72	1707.27	806.31



H 3.1.: The first hypothesis of this substudy was that time on task would mediate the effects of sixth graders' attitudes and prior ability in the same way as it did for ninth graders in the earlier study of Kupiainen et al. (2014). More specifically, it was expected that the indirect effect would be stronger for detrimental attitudes than for mastery attitudes or prior ability. It was also hypothesised that a measure of general cognitive competence instead of GPA would make the interpretation of the results easier as an external measure is not expected to be affected by attitudes in the same way as GPA. Therefore, at this stage, two versions of the models were compared: in the first set of models prior ability was indicated by GPA as in the earlier study of Kupiainen et al. (2014). In the second set of models GPA was replaced by an external measure of prior ability; the analogical reasoning test score from the third grade.

First the effects of attitudes and prior ability on the sixth grade test score were studied without taking time on task into account. Two simple path models were specified; one in which test score was predicted by mastery and detrimental attitudes and GPA, and another in which GPA was replaced by an analogical reasoning test score from the third grade. The models are presented in Figure 6.2. The fit indices for both models were good (CFI=.995, TLI=.955, RMSEA=.057,  $\chi^2=6.088$ ,  $df=1$ ,  $p=.014$ , and CFI=.992, TLI=.917, RMSEA=.059,  $\chi^2=6.377$ ,  $df=1$ ,  $p=.012$ , respectively).



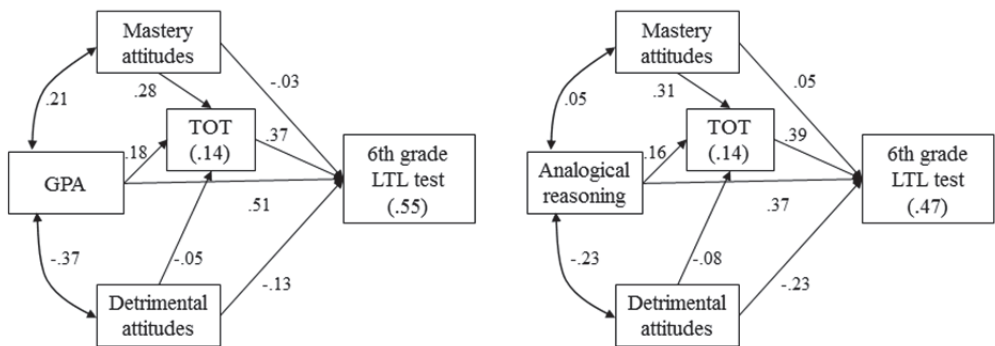
**Figure 6.2.** Predicting the sixth grade test score with prior ability and mastery and detrimental attitudes. On the left side GPA has been used as the indicator of prior ability, and on the right side the third grade analogical reasoning test score was used. Numbers in parentheses indicate the share of accounted for variance.

Figure 6.2. shows that the sixth grade LTL test score could be relatively well predicted by prior ability and mastery and detrimental attitudes. When comparing the model with GPA with that of Kupiainen et al. (2014), it can be seen that GPA was as strong a predictor for sixth graders' test scores as it was for 9th graders. However, the role of attitudes was much weaker than it was for ninth graders, possibly reflecting younger pupils' limited self-awareness (Demetriou & Kazi, 2006). Thus, the share of explained variance was smaller than it was in the earlier study (44 % vs. 61 %). When comparing the two models in Figure 6.2., it can be seen that just as expected, both mastery and detrimental attitudes had stronger partial correlations with GPA than with analogical reasoning skills. Kupiainen et al. (2014) suggested that pupils would get rewarded for having positive learning-related attitudes in their school grades, which would explain the correlations which were in that study even stronger than here. Figure 7.2. shows that at least for mastery attitudes this seems to be the case as they were moderately related to GPA ( $r=.21$ ,  $p<.001$ ) but not at all to analogical reasoning ( $r=.05$ , ns.). However, detrimental attitudes correlated also with analogical reasoning ( $r=-.23$ ,  $p<.001$ ), but the correlation was weaker than with GPA ( $r=-.36$ ,  $p<.001$ ). A possible explanation for this is that pupils with a high level of detrimental attitudes in the sixth grade had some issues with attitudes already in the third grade, and therefore they did not try their best in the analogical reasoning test. Contrary to the expectations, mastery and detrimental attitudes did not correlate with each other in any of the two models, and the path was therefore removed.

Despite the negative correlation with detrimental attitudes it could be concluded that an external measure, in this case the analogical reasoning test, brought the independent role of attitudes visible in explaining the test score better. This can be seen in the path coefficients of the sixth grade test score: In the left model with GPA mastery attitudes predicted the test score only weakly ( $\beta=.07$ ,  $p<.001$ ) while in the right model the effect was somewhat stronger ( $\beta=.17$ ,  $p<.001$ ). The same observation was made in the case of detrimental attitudes, which were a slightly stronger predictor of the test score in both models ( $\beta=-.15$ ,  $p<.001$  vs.  $\beta=-.26$ ,  $p<.001$ ). However, GPA was a slightly better predictor of test score than the third grade analogical reasoning skills ( $\beta=.58$ ,  $p<.001$  vs.  $\beta=.43$ ,

$p < .001$ ), probably because it covers a broader spectrum of competences instead of a single skill, just like the LTL test does.

The next step was to add time on task in the models as a mediator. This was done for both models of Figure 6.2., that is, for the one with GPA as an indicator of prior ability, and for the other model using analogical reasoning test scores from the third grade. The models with time on task are presented in Figure 6.3. Both models fitted the data well (CFI=.997, TLI=.953, RMSEA=.059,  $\chi^2=6.313$ ,  $df=1$ ,  $p=.012$ , and CFI=.995, TLI=.931, RMSEA=.060,  $\chi^2=6.567$ ,  $df=1$ ,  $p=.010$ , respectively).



**Figure 6.3.** Predicting the sixth grade test score with prior ability, mastery and detrimental attitudes and time on task (TOT) as a mediator between the attitudes, prior ability and the test score. On the left side GPA has been used as the indicator of prior ability, and on the right side the third grade analogical reasoning test score was used. Numbers in parentheses indicate the share of accounted for variance.

Figure 6.3. shows that, just as expected, time on task was a meaningful predictor of sixth graders' test score even when prior ability was taken into account, regardless of the measure of prior ability ( $\beta=.37$ ,  $p < .001$  and  $\beta=.39$ ,  $p < .001$ ). The effect was not as strong as for ninth graders in the study of Kupiainen et al. (2014) but still much stronger than the role of attitudes was in the models without TOT. Moreover, TOT was predicted by the other variables in the models even if the share of explained variance of it was not as large as in the study of Kupiainen and colleagues (14 % vs. 38 %). The most important difference can be found in the role of detrimental attitudes: While in the earlier study TOT mediated the effects of detrimental attitudes on LTL test score, here the indirect effects were only  $\beta=-.02$  and  $\beta=-.03$  for the left and the right

model, and the direct effects of detrimental attitudes on TOT were  $\beta = -.05$  ( $p < .05$ ) and  $\beta = -.08$  ( $p < .001$ ). Even though it was not possible to produce confidence intervals for the indirect effects with the statistical programme used here, it is likely that these indirect effects were not statistically significant. Thus, TOT did not mediate the effects of detrimental attitudes on the sixth grade LTL test score, and the direct effects on test score were moderate even when having TOT in the model ( $\beta = -.13$ ,  $p < .001$  and  $\beta = -.23$ ,  $p < .001$  for the left and the right model).

For mastery Attitudes it was just the opposite: While in the study of Kupiainen et al. (2014) the mediation was weak and competitive (after the mediating variable the direct effect was negative even though the indirect effect was positive, see MacKinnon et al., 2000), in this study the indirect effects of mastery attitudes on LTL test score were stronger than for detrimental attitudes ( $\beta = .10$  and  $\beta = .12$  for the left and the right model). Accordingly, the direct effects decreased to  $\beta = -.03$  (ns.) and  $\beta = .05$  ( $p < .01$ ). In other words, in the left model with GPA as the measure of prior ability mastery attitudes predicted the test score only indirectly via GPA and TOT, and also in the right model with Analogical reasoning mastery attitudes explained very little variance in addition to the other variables in the model. The role of mastery attitudes in explaining TOT was clear, however, and it was the best predictor of time investment in both models of Figure 7.2. ( $\beta = .28$ ,  $p < .001$  and  $\beta = .31$ ,  $p < .001$ ).

Just like in the earlier study with 9th graders, time on task was also predicted by prior ability regardless of the measure of it ( $\beta = .18$ ,  $p < .001$  and  $\beta = .16$ ,  $p < .001$  for GPA and analogical reasoning). There were also small indirect effects through TOT ( $\beta = .07$  and  $\beta = .06$ ) which corresponded with that of Kupiainen et al.'s (2014) study.

The comparison of the four models showed that the role of time on task in explaining pupils' test score in a low stakes assessment was more or less similar for sixth grade pupils as it was for 9th grade pupils in the earlier study of Kupiainen et al. (2014) with approximately the same measures. However, TOT mediated the effects of mastery attitudes instead of the detrimental ones, which may be due to the limited self-awareness and self-evaluation skills of 12-year-old pupils. Demetriou and Kazi (2006) showed that pupils' self-awareness increases dramatically from the age of 11 to 15, and for 12-year-olds it can be easier to evaluate

oneself through positive statements instead of negative ones. This may also be the explanation for the effects of attitudes being, in general, weaker than in the earlier study. Comparison to the earlier study must however be done with reservations, as instead of using latent factors only manifest variables were used in this study in the models. Moreover, the sixth grade LTL test was not identical on item-level with the ninth grade version used in the other study. Nevertheless, it can be concluded that H 3.1. was at least partially supported regarding the mediating role of time on task, even though the mechanism of the effects of attitudes seemed to be slightly different for younger pupils. It could also be concluded that using an external measure of cognitive competence instead of GPA makes the results easier to interpret as GPA is clearly influenced by both mastery and detrimental attitudes. Therefore, even if the share of explained variance of test score was slightly smaller in the models with third grade Analogical reasoning scores instead of GPA, the model with Analogical reasoning was chosen for further analyses for testing the hypotheses 4.2.- 4.4.

H 3.2.: In the second hypothesis it was assumed that girls would have more mastery attitudes and less detrimental attitudes, and a higher sixth grade test score even though there should be no gender difference in girls' and boys' general cognitive competence. This was assumed to be due to increased investment of time, which is related to positive attitudes. To test this hypothesis gender was added in the model with Analogical reasoning as the measure of prior ability. The gender variable was dummy-coded with 0 = Boys and 1 = Girls.

At the first stage it was studied how gender was related with performance in general. A simple path model was tested, in which sixth grade performance was predicted by analogical reasoning skills and gender only. Just as expected, gender turned out to be unrelated with third grade Analogical reasoning skills. It, however, predicted sixth grade performance weakly but statistically significantly ( $\beta=.07$ ,  $p<.001$ ). The model fit was good (CFI=.996, TLI=.979, RMSEA=.029,  $\chi^2=2.314$ ,  $df=1$ ,  $p=.128$ ).

Next, the attitude variables and time on task were added to the model. Contrary to expectations, gender was not related to the level of detrimental attitudes. However, being a girl predicted both mastery

attitudes and Time on task weakly but statistically significantly ( $\beta=.09$ ,  $p<.001$  for both variables), and there were also small indirect effects both on TOT via mastery attitudes ( $\beta=.03$ ) and on the test score ( $\beta=.05$ ). Moreover, the direct effect of gender on the sixth grade test score decreased to  $\beta=.02$  (ns.). Thus, the effect of gender, which was clear when the test score means were compared by simpler statistical methods ( $M=41.36$  vs.  $38.58$ ,  $t=-3.54$ ,  $p<.001$ ), was almost completely mediated by Time on task and mastery attitudes. The model fit was good (CFI=.993, TLI=.971, RMSEA=.033,  $\chi^2=13.476$ ,  $df=5$ ,  $p=.019$ ). Even though it was not possible to test the significance of the indirect effects, they nevertheless removed the direct effects of gender on test score completely. Therefore, it could be concluded that H 3.2. was supported except for the non-significant relationship between gender and detrimental attitudes. The effect of gender was so small, however, that the share of explained variance of the test score did not increase from the 47 % of the model without gender.

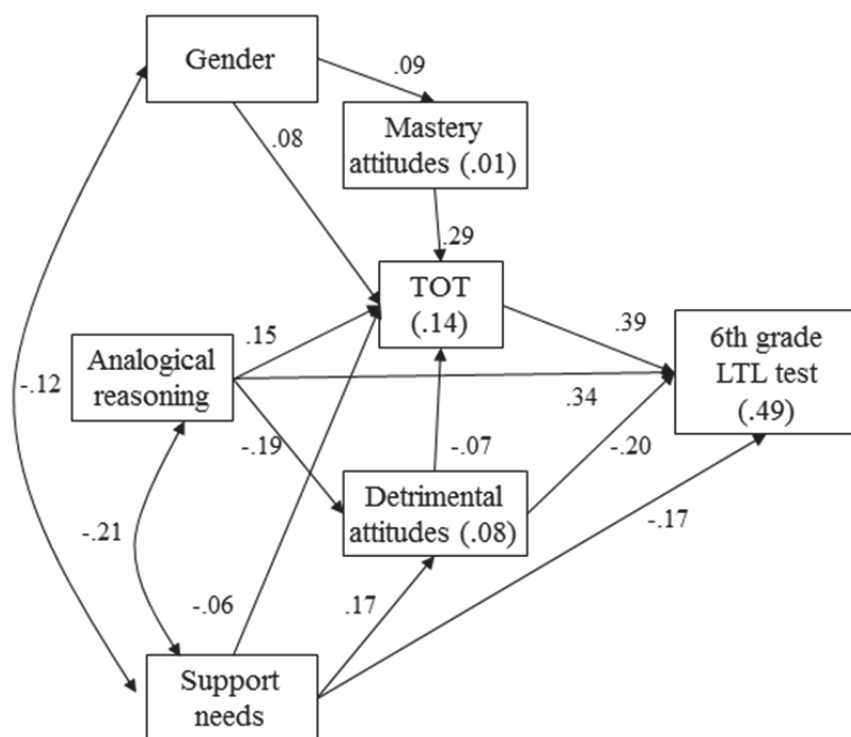
H 3.3.: The third hypothesis was that support needs would increase the time needed and they would therefore be positively related to TOT. Support needs were also expected to be related to lower prior ability but not directly to sixth grade test score when prior ability was taken into account. However, support needs were expected to be related to more negative attitudes, and an indirect effect through attitudes was expected.

Support needs were indeed related to a lower analogical reasoning test score ( $r=-.21$ ,  $p<.001$ ), and they predicted detrimental attitudes positively ( $\beta=.18$ ,  $p<.001$ ). They also decreased the relationship between analogical reasoning and detrimental attitudes slightly to  $\beta=-.19$ ,  $p<.001$ , analogical reasoning to test score to  $\beta=.34$ ,  $p<.001$  and detrimental attitudes to test score to  $\beta=-.20$ ,  $p<.001$ . The changes were small but they all indicated that when considering the effects of detrimental attitudes on performance, support needs are an important background factor to be taken into account. However, contrary to H 3.3., support needs predicted TOT negatively ( $\beta=-.07$ ,  $p<.01$ ), and they also had a direct negative effect to the test score ( $\beta=-.17$ ,  $p<.001$ ). Support needs were not related to mastery attitudes. Thus, despite lower cognitive ability – which would according to Carroll (1963) predict an increase in time needed – the pupils in need of support actually spent less time on the tasks than the

others. They had more detrimental attitudes, and in addition to a small indirect effect on test score ( $\beta=-.07$ ) support needs predicted test score also directly – indicating that pupils in need of support performed worse than others with similar levels of prior ability, time investment and attitudes. Thus, H 3.3. was supported only partially in regard to the relationships with prior ability and detrimental attitudes. The model fit was acceptable (CFI=.986, TLI=.928, RMSEA=.056,  $\chi^2=23.111$ ,  $df=4$ ,  $p<.001$ ).

H 3.4.: At the last stage gender and support needs were included in the model simultaneously to see how these background factors together predicted the other variables in the model. It was assumed that more boys than girls would be identified as having support needs, so a negative correlation between support needs and gender was expected. Regardless of that, both background variables were expected to have an independent contribution in explaining the other variables. The final model is presented in Figure 6.4. The model fitted the data well (CFI=.989, TLI=.956, RMSEA=.038,  $\chi^2=22.660$ ,  $df=7$ ,  $p=.002$ ).

Figure 6.4. shows that, just as expected, support needs and gender correlated weakly but statistically significantly ( $r=-.12$ ,  $p<.001$ ), so more boys than girls received support for their studies. Having both background variables in the model simultaneously did not change the other path coefficients with more than one decimal from what was reported in H 3.2. and 3.3., so both background variables had an independent – albeit weak – contribution in explaining directly or indirectly the variance of the sixth grade LTL test score, of which 49 % was explained with the final model. The share of explained variance of time on task did not change from 14 % when having the background variables in the model. In the final model also mastery and detrimental attitudes were endogenous variables, and it could be seen that of detrimental attitudes 8 % of variance was explained by support needs and analogical reasoning. For mastery attitudes, gender, the only predictor in the model, explained only 1 % of the variance of it. In summary, Figure 6.4. shows that H 3.4. was fully supported.



**Figure 6.4.** Predicting the sixth grade test score with prior ability, mastery and detrimental attitudes and time on task (TOT) as a mediator between the attitudes, prior ability and the test score. Gender and support needs are included in the model as dummy-variables (1= Girl and Support needs). Numbers in parentheses indicate the share of accounted for variance.

### 6.5 Conclusions of Substudy 3

The third substudy showed the importance of learning-related attitudes and effort as measured by time investment on task in explaining test performance in a low stakes assessment. Time on task had a relatively strong effect on the test scores, and it also shed light on the mechanisms of how attitudes influence performance for 12-year-old pupils. The substudy showed that mastery attitudes predicted greater effort as measured by time on task, which in turn predicted better performance.

The main findings of this substudy are however related to explaining group-level differences based on gender and support needs. The study showed that girls' slightly more positive attitudes and greater effort explained the observed gender difference completely. This finding is of high importance, as it can partially explain also why gender differences



during later school years increase further. Nevertheless, the effect of attitudes on performance is only beginning to increase when the pupils are at the age of the participants of the present study. However, more research is needed with older age groups to understand the development of gender differences better. This study also showed that, to some extent, the reason that pupils with support needs do not perform as well as expected based on their earlier competences is due to their higher level of detrimental attitudes and reduced time investment. Unlike with gender, the mediation was far from complete, which indicates that this study failed in addressing other important factors, which explain the performance of pupils with support needs.



## **7 Substudy 4: School, class and peer effects on the development of learning to learn skills**

The last substudy ties the results of the three other substudies together, deepening the understanding of the findings presented separately in each substudy. The aim is to examine to what extent the phenomena discussed earlier occur at an individual level and how big a role schools, classes and possible subgroups within classes play in explaining the differences and changes. To do that, social networks within classes are first studied qualitatively to see if it would be possible to define close-enough approximations of realistic peer groups to be used as the lowest level in multilevel modelling when analysing large-scale assessment data. After that, variance components modelling is applied to all the variables used in the other substudies to find systematic school-, class and peer group - level effects that would partially answer some of the concerns stated above in the introduction. That is, to find possible explanations for the growing gender differences and the decrease in assessment results and to define concrete targets for interventions.

The research questions of the last substudy are:

Q 4.1.: How are pupils' social networks built within classes during the first 6 years of basic education? Can peer group be easily used as the lowest level in multilevel modelling of large-scale assessment data?

Q 4.2.: Are the differences in pupils' performance in the beginning of basic education partly explained by the school and class they attend, or the performance of peers in the same classes?

Q 4.3.: Do schools, classes and peer groups explain more of the variance of the cognitive assessment results in the beginning of the fourth grade and at the end of the sixth grade?

Q 4.4.: Are there systematic school, class or peer effects in explaining the variance of learning-related attitudes and the change of them from the fourth to sixth grade? Does the same apply to the change of task interest?

Q 4.5.: Can the findings of Q 4.3. and 4.4. be generalised across cities with different policies regarding school choice? Are the effects different when learning-related attitudes are understood to consist of both mastery and detrimental attitudes?

Q 4.6.: Are there systematic school, class or peer effects in explaining the variance of time investment in the assessment tasks?

Q 4.7.: Are there important gender differences in variables used in answering research questions 4.3. – 4.6.?

The following hypotheses were set:

H 4.1.: Boys of 7 to 12 years attending the same class all play together, and it is easy to use the boy group as an approximation of a real peer group. Girls however tend to play with fewer good friends (Rose & Rudolph, 2006), so girl subgroups have to be defined based on sociometric nominations.

H 4.2.: Systematic school and peer group level effects on first grade results are almost non-existent: There has not been enough time for them to produce effects after only a month of schooling, and in the beginning of first grade most pupils attend their local school. There may be systematic class-level effects which are partly explained by differences in the instructions in the assessment situation: With illiterate first graders the tasks are by necessity more teacher-dependent, and regardless of detailed written instructions some variation in the situation is possible. However, some of the class-level effects may be due to the non-random assignment of first graders in their classes (cf. Willms, 2010).

H 4.3.: In the beginning of the fourth grade there are more differences explained by schools and classes in cognitive tasks than in the first grade, and the shares of explained variance increase by the end of the sixth grade. Peer groups have some explanatory value too, but the effects are smaller than for schools and classes. The gender difference in the fourth grade mathematical tasks in Substudy 1 in the Helsinki sample should be seen here as a “boy group” effect.

H 4.4.: Learning-related attitudes develop in interaction with peers (Ryan, 2000), and the same is expected to happen in the case of task interest too. The effects of schools and classes are smaller but not equal to zero.

H 4.5.: Regarding cognitive tasks, the effects of schools are smaller and the effects of classes bigger in Vantaa, which has a more restricted policy of school choice and less differentiated schools (Varjo & Kalalahti, 2011). The peer influences on attitudes are expected to be found regardless of municipality.

H 4.6.: There is either a class or peer group effect on time investment: Even though pupils are expected to work on the assessment tasks independently, in a school computer lab it is not possible to entirely prevent the pupils from following other pupils' progress. Even the randomisation of the order of tasks for each pupil separately does not keep slower pupils from rushing towards the end of the assessment when the others are already finished.

H 4.7.: Boys are more vulnerable to the influences of school- and classmates than girls are (Mähönen, Jasinskaja-Lahti & Liebkind, 2011), which would partially explain their underperformance in external assessments.

## **7.1 Participants**

In this substudy almost all data used in other substudies were used, and the descriptions of the participants can be read in Chapter 3. Only the Vantaa follow-up data from the first to third grade were not included in the analyses as PBA data comparable to the Helsinki data were available only for a random sample of 20 % of pupils. Making valid conclusions about systematic class- or subgroup-level effects would not be possible with so few pupils representing each class and subgroup.

Social networks of the classes were analysed only based on the Helsinki sample. The Vantaa data could not be used here as there were no data available about the social relationships in the classes.

## **7.2 Measures**

### **7.2.1 Cognitive tasks, attitude scales and measures of task interest and time investment**

This substudy aims at deepening the understanding of the phenomena studied in the three other substudies. Therefore, all the variables analysed in those three studies were taken into account in this substudy. Thus, all the cognitive tasks and attitude scales described in Chapter 3 and the reading skills, task interest and time on task measures described in Chapters 4 to 6 were included in the analyses.

### **7.2.2 Social relationships in school classes**

At the end of the second and fifth grade the pupils in the Helsinki sample filled out short questionnaires of their social relationships within the class. They were asked to name a maximum of 5 classmates who they preferred to work with on school tasks, who they played with during the breaks and who they played with after school. The method was adapted from traditional sociometric literature (e.g. Terry & Coie, 1991), but negative nominations were not asked for due to ethical reasons. In this study they were not needed either as the nominations of play time playmates were only used for drawing sociograms for each class for defining peer subgroups in order to estimate their variance components.

### **7.3 Statistical methods**

Besides qualitative analyses of social networks in classes, Variance Components models were run in the Advanced Statistics option of SPSS 18. As the deviation from normality was small for all the variables used in this Substudy (skewness and kurtosis between -1 and 1), maximum likelihood estimation could be used. Only the main effects of the random factors were analysed: Interactions could not be included in modelling because the statistical programme could not handle the large number of subgroups within classes.

### **7.4 Results**

As this Substudy deepens the analyses of the variables used in the other three substudies, the descriptive statistics for all but the sociometric variables are presented in the Result sections of Chapters 4 to 6.

#### **7.4.1 Social networks of school classes**

H 4.1.: The first hypothesis of the last substudy was related to the structure of the social networks of school classes as for testing the other hypotheses of the substudy the classes needed to be divided in smaller subgroups. It was expected that during the first 6 years of basic education boys attending the same class would play all together, but girls would prefer spending time with fewer good friends. It was assumed that in

further analyses “boys in a class” could easily be used as a valid enough approximation of the real peer group while for girls, sociometric information would be necessary in determining the subgroups.

Of the three questions used for collecting sociometric nominations, the second one regarding playmates during breaks was selected for this study. It was considered to best reflect social preference and friendships instead of other possible factors: Preference on who to work with on school assignments may be biased by school success, and after-school playmates may be determined more by distance from home to school than actual preference. In both the second and fifth grade the maximum number of playmate nominations was restricted to 5, so the pupils named 0 to 5 classmates as playmates during breaks. The descriptive statistics of the nominations received from other pupils in the class are displayed in Table 7.1.

**Table 7.1.**

Descriptive statistics of sociometric nominations received by second and fifth graders

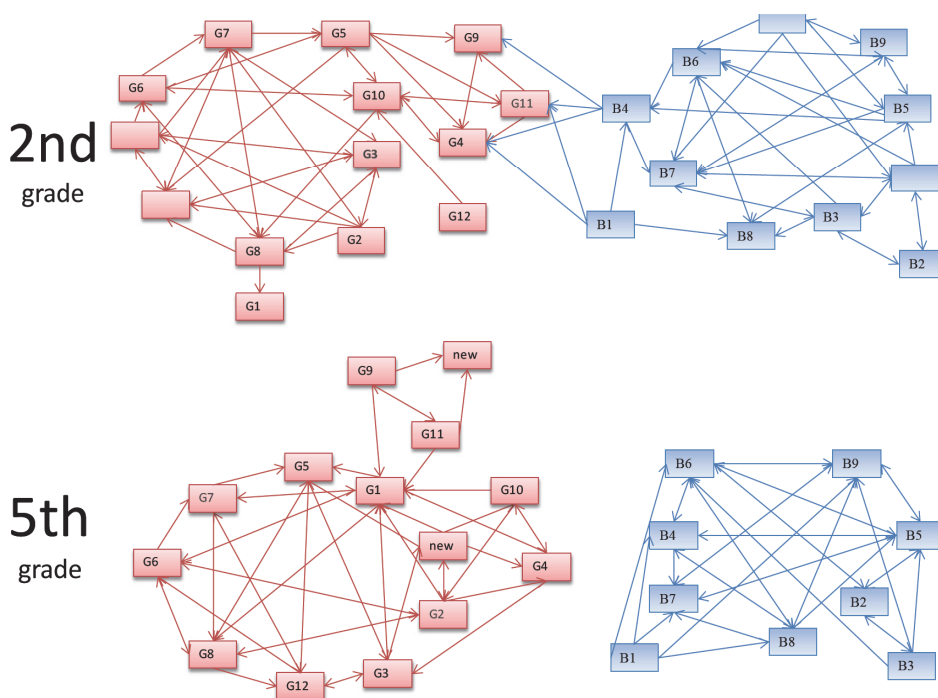
Variable	N	Min	Max	M	Sd
Number of received nominations second grade	754	0	12	3.56	1.93
Number of received nominations fifth grade	921	0	10	3.70	2.15

N= Number of responses, Min= minimum value, Max=maximum value, M=Mean, Sd=Standard deviation

Table 7.1. shows that in both the second and fifth grade there were pupils who did not receive any nominations at all and other pupils who received nominations from approximately half of their classmates. On average, each pupil was nominated by a little bit less than four classmates in both grades. The number of received nominations in the second and fifth grade correlated moderately ( $r=.31$ ,  $p<.001$ ). Data from both grades were available for 632 pupils.

After calculating basic descriptive statistics, the nominations were used for drawing sociograms for each class separately. The sociograms were drawn manually, with a two-tailed arrow indicating a mutual relationship and a one-tailed arrow indicating a non-mutual nomination. Sufficient data for drawing sociograms were available for 38 second grade classes and 45 fifth grade classes. An example of sociograms from

two different grades for the same class is given in Figure 7.1. It is to be noted, however, that in the sample there were very few classes which had remained approximately the same over the transition from second to third grade. Nevertheless, the basic structures of girls' and boys' social relationships were more or less similar in all classes regardless of whether they had been mixed in between or not.



**Figure 7.1.** An example of social relationships of a school class which has – unlike most other classes in the study – not been mixed between the second and fifth grade. Codes in the boxes are unique names given to the participants (G= Girl, B=Boy), so changes in pupils' social positions can also be followed. Pupils without codes were no longer in the sample in the 5<sup>th</sup> grade but they were of the same gender as their peers. Two-tailed arrows indicate mutual relationships. The structure of the social networks of boys and girls presented here were typical also of classes which had been reorganised in the beginning of the third grade.

The first observation from the sociograms was that while in second grade there were some – even though they were few – relationships between girls and boys in many classes, in the fifth grade cross-gender nominations were basically nonexistent. It could be therefore concluded that girls and boys formed more or less separate social networks in all the classes of this study.



Studying the sociograms revealed that, just as expected in H 4.1., boys formed only one big group in all classes. There could be “outsiders” who apparently wanted to be a member of the group but who received very few if any nominations from their classmates, like the boy B1 in Figure 1. Nevertheless, there were clearly no separate boy subgroups either in the second or fifth grade, and it could be concluded that H 4.1. was supported in respect to the boys. Analysing the stability of peer relationships and social status over the years – also visible in Figure 7.1. – was beyond the scope of this study and will therefore not be discussed further here.

Girls' sociograms were somewhat more surprising. Contrary to the expectations it was very difficult to observe subgroups based mainly on mutual nominations: In second grade it could be seen a little clearer that girls tended to have a few good friends, but by the end of fifth grade the networks had become exactly like boys' big peer groups. There could be some girls who were not quite “inside” the group but still, “girls in the class” seemed to form a relevant peer group for most girls in any class of this study. So, for girls H 4.1. was not supported.

The next step was originally to assign each child to his or her real-life peer group for further analyses. Based on the sociograms it was concluded that this could be done more simply than expected: peer groups were coded in the data by combining the categorical variables of gender and class, and these gender groups within classes were then used in variance components models as approximations of peer groups.

#### **7.4.2 Variance components modelling**

H 4.2.: In the second hypothesis it was expected that there would not be significant school or peer group level effects on first grade cognitive results, as the children had only gone to school for approximately one month at the time of the assessment. However, some class-level effects were expected because the tests were teacher-administered and despite detailed written instructions there is always a risk for individual interpretations of them.

The results of the variance components model of the Helsinki first grade data are presented in Table 7.2. The variances calculated by maximum likelihood estimation are converted to percentages to make the results comparable across measures.

**Table 7.2.**

Variance components of the first grade variables of Substudy 1 (see Figure 4.1.)

Level	Analogical reasoning	Visuo-spatial memory	Following instructions	Reading skills at school start
School	1.6 %	4.2 %	2.7 %	0.0 %
Class	6.9 %	14.4 %	3.1 %	4.9 %
Gender group in class	0.0 %	4.1 %	0.0 %	0.0 %
Individual	91.5 %	77.3 %	94.2 %	95.1 %

Percentages indicate the share of variance explained by the level

Table 7.2. shows that just as expected, there were systematic class-level effects that explained the variance of all the first grade cognitive tasks. The effect, a little surprisingly, was the smallest in the Following instructions task (3.1 %) even though it was based on the teacher dictating the path to be drawn. Apparently, the teachers were carefully following their instructions in how to administer this task. The effect was the largest for the visuo-spatial memory task (14.4 %). A possible cause for this is that some teachers may have kept the stimulus visible a little longer than the instructed three seconds, which may have made the tasks easier for their pupils. In this task there was also a gender group effect (4.1 %), while in other tasks gender group did not have any systematic effects on the results. There was a class-level effect on reading skills at school start (4.9. %), which most likely reflected the differences in teachers' expectations and the use of the evaluation scale. The somewhat stronger class effect in analogical reasoning skills (6.9 %) – the task that was the least dependent on teachers' instructions – may tell about systematic differences between classes: In Finland, children are not completely randomly assigned in first grade classes if there is more than one class in the same school. There is a lot of information available from daycare (Ahtola et al., 2011) which is used also in forming balanced classes and groups in addition to organising support. Thus, some of the class level differences in the beginning of basic education are actually produced consciously by the schools.

Contrary to H 4.2., there were small school-level effects on all the cognitive tasks. Even though the effects were not strong (1.6 – 4.2 % of explained variance), they showed that there were systematic differences

between schools already in the very beginning of basic education. As at this stage most children still go to their local schools even in municipalities with more possibilities for school choice, these differences can be interpreted as indicators of differentiation of residential areas (cf. Bernelius, 2013). Indeed, controlling for mothers' education decreased the school-level effects to 0.7 – 2.9 %.

To study this finding further variance components of the same variables were estimated also for the whole cohorts of first graders in Vantaa, in which residential areas are less differentiated. As all the first graders took the PBA version of the first grade test in 2010, the full data were used here even though only the random sample completing the PBA version of the third grade test was selected in Substudy 1. The variance components are presented as percentages in Table 7.3.

**Table 7.3.**

Variance components of the first grade variables of Substudy 1 in Vantaa

<b>Level</b>	<b>Analogical reasoning</b>	<b>Visuo-spatial memory</b>	<b>Following instructions</b>	<b>Reading skills at school start</b>
School	0.0 %	0.0 %	4.3 %	0.1 %
Class	1.1 %	15.1 %	5.6 %	2.6 %
Gender group in class	5.9 %	0.0 %	0.0 %	3.9 %
Individual	92.9 %	84.9 %	90.1 %	93.4 %

Percentages indicate the share of variance explained by the level

As can be seen in Table 7.3., school-level effects were smaller in the municipality with less differentiated residential areas and, as a result of it, local schools. The stronger school-level effect on Following instructions (4.3 %) – the task requiring comprehension of verbal instructions – was most likely due to a high number of immigrants in some schools in Vantaa. Controlling the mother tongue of the pupils decreased the school-level effect to the same level as it was in Helsinki without controlling for mothers' education, and controlling for mothers' education diminished it further.

For some reason the gender group level effect in Vantaa was relatively strong for analogical reasoning skills (5.9 %), whereas the class effect was small (1.1 %). This was interpreted as a sign of even less random assignment of pupils in classes than in Helsinki, most likely related to the

imbalance between genders regarding support needs. This, however, is beyond the scope of this substudy and will not be discussed further here. It was concluded that H 4.2. was supported regarding the most central assumptions, but there were a lot of nuances in the results that should be studied further.

H 4.3.: The third hypothesis was that school and classes would explain more of the variance of the cognitive tasks in the beginning of the fourth grade than they did in the first grade, and by the end of the sixth grade the shares of variance explained by schools would increase further. As boys turned out to be better in mathematical thinking in the fourth grade, a small boy group effect was expected here too. Regarding other variables, it was expected that peer groups would have some explanatory value too, but the effects would be smaller than for schools and classes.

As the fourth grade cognitive results were discussed in both Substudy 1 and 2, variance components were calculated separately for the cognitive variables used in the studies. In Substudy 1, reading comprehension, mathematical thinking and reasoning skills were treated as separate dimensions of the fourth grade LTL skills, so the variance components for the three subscores are presented separately in Table 7.4. Substudy 2 used only a single test score for the fourth grade performance, but analogical reasoning scores were used in controlling general cognitive competences. The variance components of these scores are presented in Table 7.5., which contains also variance components for the sixth grade test score of the Helsinki sample.

**Table 7.4.**

Variance components for Helsinki of the fourth grade cognitive tasks of Substudy 1 (see Figure 4.1.)

<b>Level</b>	<b>Reading comprehension</b>	<b>Mathematical thinking</b>	<b>Reasoning skills</b>
School	8.8 %	2.6 %	3.4 %
Class	13.9 %	6.9 %	4.9 %
Gender group in class	0.0 %	8.5 %	1.9 %
Individual	77.4 %	82.0 %	89.8 %

Percentages indicate the share of variance explained by the level

Table 7.4. shows that most of the group-level variance in the fourth grade was still on the class- or peer group-level even though schools explained a relatively high proportion of the variance of reading comprehension (8.8 %). Reading comprehension scores were also quite strongly explained by classes (13.9 %), while gender groups within classes did not provide any added value regardless of the gender difference in favour of girls. In mathematical thinking, however, the gender difference in favour of boys did in fact seem to depend partly on systematic boy group effects as gender groups within classes explained 8.5 % of the variance of them. For reasoning skills, the results were as expected: Both schools and classes explained some variance, but the effects were not very high.

**Table 7.5.**

Variance components of the cognitive tasks of Substudy 2 (see Figure 5.1.)

Level	Fourth grade LTL test	Analogical reasoning	Sixth grade LTL test
School	5.9 %	4.5 %	11.0 %
Class	4.9 %	2.1 %	3.9 %
Gender group in class	5.4 %	3.6 %	1.5 %
Individual	83.8 %	89.8 %	83.6 %

Percentages indicate the share of variance explained by the level

In Table 7.5. where the fourth grade results are presented as a single score, which is comparable to the sixth grade results in the same table, an interesting phenomenon is observed. While in fourth grade class and peer group effects were together stronger than school effects, by the end of the sixth grade the school effects had increased and they now explained 11 % of the variance of the test score in Helsinki, with class and peer group effects being much smaller. It was concluded that H 4.2. regarding the increase of school effects over years was supported, even though it has to be noted here that the Vantaa sixth grade results were different in this respect. They will be presented in detail in H 4.5.

H 4.4.: The fourth assumption was that learning-related attitudes and task interest would develop in interaction with peers and that schools and classes would not explain as much variance of them as gender group in class does. To study that, the variance components of all the variables used in the second substudy were estimated, first separately for the

attitude scale means in the fourth and sixth grade (Tables 7.6. and 7.7.) and then for all the change variables (Table 7.8.).

**Table 7.6.**

Variance components of the fourth grade attitude scales of Substudy 1 (see Figure 4.1.)

Level	Learning orientation	Achievement orientation	Agency: Effort
School	0.0 %	0.0 %	0.0 %
Class	5.5 %	0.0 %	0.0 %
Gender group in class	0.0 %	7.6 %	2.7 %
Individual	94.5 %	92.4 %	97.3 %

Percentages indicate the share of variance explained by the level

Table 7.6. shows that in the fourth grade, schools did not explain any variance of the attitude scale means, and class had systematic effects only on Learning orientation (5.5 %). Gender groups within classes explained Achievement orientation the most (7.6 %), and they also had a small effect on Agency: Effort (2.7 %).

**Table 7.7.**

Variance components of the attitude scales: The classes of Substudy 1 in the beginning of the fourth grade

Level	Learning orientation	Achievement orientation	Agency: Effort
School	0.2 %	1.2 %	2.7 %
Class	0.0 %	0.0 %	0.0 %
Gender group in class	6.6 %	3.3 %	4.9 %
Individual	93.1 %	95.5 %	92.4 %

Percentages indicate the share of variance explained by the level

Gender groups within classes explained the greatest shares of variance of the attitude scale means also in the sixth grade. Now class did not have an effect on any of the scale means, but schools explained their variance a little (0.2 % - 2.7 %). Even though Learning orientation was in the fourth grade partly explained by class effect, in the sixth grade only gender groups within classes were of significance (6.6 %). They also explained some of the variance of Achievement orientation and Agency: Effort (3.3 % and 4.9 %, respectively).

**Table 7.8.**

Variance components of the change variables of Substudy 2 (see Figure 5.1.)

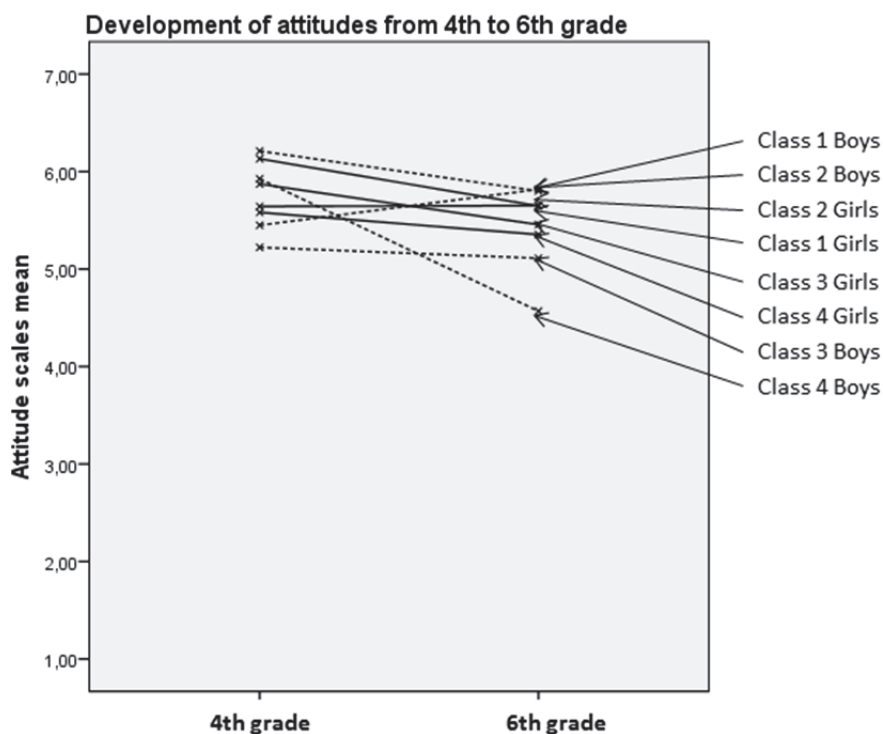
<b>Level</b>	<b>Learning orientation</b>	<b>Achievem. orientation</b>	<b>Agency: Effort</b>	<b>Task interest: Arit. Op.</b>	<b>Task interest: Mental Arit.</b>	<b>Task interest: Reasoning</b>
School	0.9 %	0.2 %	1.5 %	5.3 %	4.2 %	0.0 %
Class	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	2.9 %
Gender group in class	2.6 %	1.6 %	4.0 %	3.4 %	2.2 %	0.0 %
Individual	96.5 %	98.2 %	94.5 %	91.4 %	93.6 %	97.1 %

Percentages indicate the share of variance explained by the level

For the longitudinal change variables, the shares of variance explained by systematic school, class or peer group effects were in general smaller than when the same scale means were looked at cross-sectionally. The changes in Learning orientation and Achievement orientation had happened mainly at an individual level, but gender groups within classes explained a small proportion of their variance (2.6 % and 1.6 %, respectively). For Agency: Effort – the children's evaluations of the effort they put into schoolwork – the gender group effect was stronger (4.0 %), and schools explained some of its variance too (1.5 %). Thus, changes in children's self-reported effort were to a small extent dependent on how their peers' self-reported effort had changed; in addition, also school culture may have played a small role. Since the results of the changes in all the attitude scale means pointed in the same direction, even if their magnitude varied, it could be concluded that there was a peer group effect in the attitude change from the fourth to sixth grade. Thus, H 4.4. was supported.

This interpretation was further confirmed by drawing separate plots for each school of the Repeated measures GLM results presented in the results section of Substudy 2, which shows all the within-class gender groups of the school in the same plot. It could be seen clearly that attitudes could have developed in opposite directions for girls and boys in the same class, or for girls/boys in two different classes of the same school. An example of this is given in Figure 7.2. A relatively large school was selected to be presented here to demonstrate the great peer group level variation in one school only. For demonstration purposes, the three

attitude scale means were averaged to a single score for learning-related attitudes in the fourth and sixth grade. In this, the results are not completely identical to those presented in substudy 2, as GLM was performed only for the separate attitude scales, and the latent mean used in the other analyses was calculated by using confirmatory factor analysis in SEM.



**Figure 7.2.** An example of attitude development in the four classes of one school from the fourth to sixth grade. The lines represent within-class gender groups of the school. Boys are represented with dotted lines.

H 4.5.: In the fifth hypothesis it was expected that in Vantaa the systematic school effects would be smaller and class effects bigger than those presented in H 4.3. for Helsinki due to the differences in school choice policies in the two cities. No municipal differences were expected to be found in attitudes as peer group influences should not depend very much on schools. The variance components of the variables used in the third substudy with sixth graders from Vantaa are presented in Table 7.9.



**Table 7.9.**

Variance components of the variables of Substudy 3 (see Figure 6.4.)

Level	Analogical reasoning	Mastery attitudes	Detrimental attitudes	Time on task	Sixth grade LTL test
School	2.9 %	3.7 %	1.2 %	7.3 %	2.3 %
Class	4.8 %	3.9 %	4.7 %	13.9 %	15.9 %
Gender group in class	0.6 %	5.6 %	0.0 %	6.0 %	4.8 %
Individual	91.7 %	86.8 %	94.2 %	72.8 %	77.1 %

Percentages indicate the share of variance explained by the level

It can be seen in the last column of the table that quite a large proportion of variance of the sixth grade test scores were explained by group-level variation, but just as expected, most of the variance was at a class-level (15.9 %) and very little at a school-level (2.3 %). Even though the results are not directly comparable to those presented in Table 7.5., as also reading comprehension and verbal reasoning were included in the test scores in Vantaa, it can be concluded that in Vantaa the variation seems to be placed in classes instead of schools and not in a random way. This can also be seen in the third grade analogical reasoning scores in the left column of the table even though the effects were smaller. Like in the first grade results presented in Table 7.3., also for sixth graders the within-class gender group effects in the cognitive tasks were stronger in Vantaa than in Helsinki. This gives more support to the assumption that the classes may have been consciously balanced by placing high-achieving girls together with boys in need of support, or vice versa.

For mastery attitudes there was a small school-level effect (3.7 %) and unlike in Helsinki, also classes explained a unique share of the variance (3.9 %) even when gender groups within classes were included in the analysis. Peers – classmates of the same gender – explained 5.6 % of the variance of the mastery attitudes. For detrimental attitudes the effects were smaller and peer groups did not seem to play any role in explaining the variance of them. However, there were significant gender differences in the variance components of detrimental attitudes, which will be presented in detail in H 4.7. Here it can be concluded that H 4.5. was supported.

H 4.6.: The sixth hypothesis was that there would be either a class or a peer group effect on time investment in CBA. There is very little earlier

literature on this new area of interest, and the hypothesis was based on observations during the data collection of Substudy 3. Even though the pupils were expected to work on the assessment tasks independently, in a school computer lab it was not possible to entirely prevent the pupils from following other pupils' progress at least superficially. Even the randomisation of the order of tasks – which was done also in the present study – was not expected to suffice in keeping slower pupils from rushing towards the end of the assessment session when the others were already finished.

Table 7.9. shows that individual differences explained only 72.8 % of the variance of Time on task, which is the lowest percentage of all the variables in this substudy. This means that 27.2 % of the variance of TOT was explained by group-level factors. As expected, there was a gender group effect (6 %) which can be interpreted as a peer effect, and class was as expected the strongest level (13.9 %) in explaining the variance. Also school had a moderate systematic effect (7.3 %) on TOT. It was concluded that H 4.6. was supported.

H 4.7.: In the last hypothesis it was expected that boys would be more vulnerable to the influences of school- and classmates, assuming this would partly explain their underperformance in many assessments. This was studied by calculating variance components of all the variables used in substudies 2 and 3 separately for boys and girls. Gender group could naturally not be included in these analyses as a level when the data were split by gender. The variance components of the cognitive tasks of the second substudy are presented in Table 7.10. Girls' and boys' results are separated by vertical bars, always showing girls' results first.

**Table 7.10.**

Variance components by gender of the cognitive tasks of Substudy 2 in Helsinki (see Figure 5.1.)

Level	Analogical reasoning	Fourth grade LTL test	Sixth grade LTL test
School	7.7   5.0 %	1.1   13.6 %	13.6   12.3 %
Class	6.8   0.0 %	13.6   2.8 %	3.0   1.8 %
Individual	85.5   95.0 %	85.4   84.1 %	83.4   85.9 %

The variance components of Girls | Boys are separated by vertical bars

Percentages indicate the share of variance explained by the level

Table 7.10. shows that when reading comprehension and verbal reasoning were not included in the test scores, there were no big gender differences in how much of the variance was explained by other than individual level. In analogical reasoning schools explained both girls' and boys' performance (7.7 % and 5.0 %, respectively) and for girls an additional 6.8 % of variance was explained by class. The most interesting observation here was that in the fourth grade systematic group-level variation was for girls in classes, but for boys it was already at the school-level like it was later for both genders in the sixth grade test. Thus, in Helsinki, school seemed to play a bigger role in fourth grade boys' performance while for girls classmates of same gender were of greater importance. In this, H 4.7. was not supported by the results.

Next, similar analyses were performed for all the variables used in the third substudy with Vantaa's sixth graders. Table 7.11. presents the percentages of explained variance.

**Table 7.11.**

Variance components by gender of the variables of Substudy 3 (see Figure 6.4.)

Level	Analogical reasoning	Mastery attitudes	Detrimental attitudes	Time on task	Sixth grade LTL test
School	4.5   1.8 %	0.2   6.7 %	1.0   2.7 %	9.3   8.7 %	1.6   4.8 %
Class	5.1   4.6 %	8.5   7.7 %	0.0   4.3 %	13.4   17.5 %	14.8   18.7 %
Individual	90.4   93.6 %	91.4   85.5 %	99.0   93.0 %	77.3   73.8 %	83.6   76.5 %

The variance components of Girls | Boys are separated by vertical bars

Percentages indicate the share of variance explained by the level

It can be seen that for third grade analogical reasoning, the results were not very different from the results of Helsinki, even though for girls the percentages of variance explained by group levels were slightly lower and for boys a little higher. However, for boys, classes explained more variance than schools, whereas for girls schools' effects were almost as strong as classes'. As already seen in Table 7.9., classes explained relatively big shares of variance of sixth grade test scores, and this applied both to girls and boys even if for boys the effect was even stronger than for girls (14.8 % vs. 18.7 %). For boys, also schools were of some importance (4.8 %), whereas for girls their effect was very small. Regarding time on task, the class effects discussed above in H 4.6. were

somewhat stronger for boys, but the school effects were of equal magnitude.

Table 7.11. shows also variance components by gender for mastery and detrimental attitudes for Vantaa's sixth graders. The results indicate some important gender differences in how children can influence each other's attitudes. Girls' mastery attitudes were not explained by school, but classmates of the same gender had a significant effect (8.5 %). Girls' detrimental attitudes were only very weakly related to their schoolmates' (1.0 %) and not at all to their classmates' detrimental attitudes. For boys, however, the effects of both schools and classes were in general higher. Mastery attitudes were explained by classes (7.7 %) but also by schools (6.7 %). Boys' detrimental attitudes were explained both by school (2.7 %) and class levels (4.3 %). Thus, the results of Vantaa's sixth graders clearly supported the hypothesis of boys' greater risk of being influenced by their peers' task behaviour and attitudes, which can then affect their results.

To study the gender differences in the development of attitudes further, variance components were calculated by gender also for the change variables of the second substudy. Table 7.12. presents the results.

**Table 7.12.**

Variance components by gender of the change variables of Substudy 2 in Helsinki (see Figure 5.1.)

Level	Learning orientation	Achievem. orientation	Agency: Effort	Task interest: Arit. Op.	Task interest: Mental Arit.	Task interest: Reasoning
School	3.2   0.0 %	2.3   0.0 %	0.0   0.0 %	0.0   8.5 %	0.7   6.2 %	0.0   1.2 %
Class	0.0   3.6 %	0.0   1.9 %	1.4   8.6 %	8.1   0.0 %	2.9   2.4 %	3.4   0.0 %
Individual	96.8   96.4 %	97.7   98.1 %	98.6   91.4 %	91.9   91.5 %	96.3   91.3 %	96.6   98.8 %

The variance components of Girls | Boys are separated by vertical bars

Percentages indicate the share of variance explained by the level

Table 7.12. shows that many of the changes in attitudes and interest had happened mainly at an individual level both for girls and boys. Interestingly, the little share of variance that was explained by group-level factors followed a pattern which is difficult to interpret. Girls' attitude changes were in general slightly explained by schools and interest changes by classes, whereas for boys it was just the opposite. The result that boys' interest changes regarding mathematical tasks were

relatively strongly explained by how other boys' interest had changed in the same school (8.5 % and 6.2 %). In addition, the fact that boys' fourth grade performance varied at the school level but not much at a class level, supports the interpretation of the results of the first substudy that in Helsinki there seems to be school cultures which encourage boys in mathematics during the first few grades. Regarding interest in mathematical tasks, it is possible that boys are initially more similar to other boys in the same school, and therefore also the changes in their interest are related. This interpretation would however require additional analyses which were not performed here.

Possibly the most important gender difference in Table 7.12. is in the changes of self-reported effort in schoolwork (Agency: Effort). For girls, classmates of the same gender explained only very little variance of the evaluations (1.4 %), whereas for boys they played a much bigger role (8.6 %). Thus, changes in boys' self-reported effort depended in part on the changes in their peers' effort in schoolwork. In Table 7.11. the same was seen as a higher percentage of variance when classmates explained actual time investment in the assessment.

The results presented in Tables 7.10.-7.12. indicate that while boys' performance is not necessarily directly explained by group-level factors much more than girls' performance, their attitudes and effort clearly are. Therefore, it can be concluded that based on the present study boys seem to be more vulnerable to the influences of their peers.

## **7.5 Conclusions of Substudy 4**

This substudy proved through sociogram analyses that girls and boys form separate peer groups in school classes but at the same time that there were no clear subgroups of boys or girls. Therefore, within-class gender groups could be used in variance components models as approximations of peer groups. Modelling showed that there were small school-level and somewhat stronger class-level effects which explained the variance of first grade cognitive competences in Helsinki. This indicated that children were not randomly assigned in their classes and that schools' pupil populations were different from the beginning, mostly because of different educational levels of mothers in different residential areas. In Vantaa, these initial between-school differences were as

expected somewhat smaller, but pupils' mother tongue produced small effects between schools. In Vantaa it also appears that the assignment of children in classes was even less random in that there were systematic gender group effects in pupils' reasoning skills.

As expected, the systematic school effects increased by the end of the sixth grade in Helsinki, whereas in Vantaa the variation stayed mostly between classes in schools. However, in Helsinki there was a gender difference in that boys' performance was more strongly explained by schools than classes already in the fourth grade; this possibly reflects a traditional "boys and mathematics" -culture in some schools, which was also seen in school-level changes in boys' task interest. In Vantaa, boys' performance was in general more influenced by school- and especially classmates' performance compared to girls.

Learning-related attitudes and task interest had developed and changed mainly at an individual level, but peer groups – classmates of the same gender – played a role in this development especially for boys. The boy effect was the strongest for self-reported effort in schoolwork and how it changed from the fourth to sixth grade, and for detrimental attitudes in the sixth grade. Boys' task behaviour, measured by their time investment on assessment tasks, was also more strongly depending on their peers' task behaviour. This was so even though also generally the school and class effects were stronger than expected, which possibly tells about differences in how the assessment situation was organised in different schools and classes.

As a general conclusion of this substudy it can be said that schools and classes play a role in how children's performance develops during the first six years of basic education and also how much effort they put into the assessment. Learning-related attitudes, in contrast, develop mainly individually and in gender groups within classes. It is not unusual that in the same class girls' attitude development goes systematically in the opposite direction to boys' attitude development. It can also be concluded that based on this study it is very likely that "the boy problem" seen in assessment results is at least partly explained by attitudes, effort and the more negative peer group culture regarding them.

## 8 Discussion

This study provided through four complementary substudies a look from an assessment perspective at the development of learning to learn skills during the first six years of basic education. The aim was to find reasons for why children's learning outcomes seem to develop differently in different schools, classes and peer groups, and especially for how children become different in how much they are willing to show what they can in external assessments. As the increasing gap between girls and boys has been of a great concern also in international assessments, this study had a special emphasis on the development of gender differences. Another target of special interest were pupils in need of support for their studies as the Finnish educational legislation has recently been changed to better follow the principles of inclusion, prevention and early intervention.

Since the present study is based on the Finnish learning to learn framework (Hautamäki et al., 2002), and it utilised municipal longitudinal learning to learn assessment data sets, the development of the learning to learn skills was accordingly understood to comprise the development of cognitive competences and learning-related attitudes. Moreover, as the focus of the study was not only on how these components of learning to learn develop but also on how they are demonstrated in an assessment situation, situational factors, which are related to the outcomes of the assessment, were also looked at. More specifically, task interest and time investment in tasks were studied, starting from the basic assumption that while they partly depend on more stable individual characteristics such as cognitive competences and learning-related attitudes, they are also partly affected by the situation or the tasks themselves. Nevertheless, they were expected to provide additional value when trying to understand how the differences in pupils' performance in external assessments develop.

The results of the four substudies are discussed here together, summarising first the findings related to the development of performance in the cognitive assessment tasks without the other factors which were

expected to affect them. After that, the development of learning-related attitudes is discussed in relation to the results regarding performance in the cognitive tasks. Finally, the results regarding task interest and time investment are presented, and it is discussed whether they are relevant factors when trying to understand what assessment results really tell about pupils' competences.

## **8.1 Development of performance in cognitive assessment tasks**

The development of performance in the cognitive assessment tasks was addressed in all four substudies, even though only in the second substudy and in the relating analyses in the fourth substudy were identical items used with the same children more than once. The first and third substudies tried instead to find predictors for children's later performance from their success in age-appropriate tasks during earlier school years.

### **8.1.1 From first grade to the turn of the third and fourth grade**

In the first study it was evaluated how largely children's performance in a low-stakes assessment at the turn of the third and fourth grade depended on their general cognitive competences and reading skills in the beginning of the first grade. The role of attitudes was also examined, but those results will be discussed in detail in the section about the development of learning-related attitudes as they were not relevant for understanding the cognitive results.

Regarding cognitive competences, the results of the first substudy supported the hypothesis that both cognitive competences and reading skills as demonstrated during the first months of basic education are good predictors of children's performance in the cognitive learning to learn tasks at the turn of the third and fourth grade (cf. Duncan et al., 2007). On the other hand, the findings also supported the assumption that reading comprehension, mathematical thinking skills and reasoning skills are not predetermined by individual differences in the beginning of school as more than half of the variance of mathematical thinking and reasoning skills, and more than 80 % of the variation of reading



comprehension were explained by other factors than the measured cognitive competences. Even though it is likely that reading comprehension would have been explained much better if more accurate measures for children's prior verbal abilities and language proficiency had been used (cf. Leppänen et al., 2008), the results still indicate that education has a significant role in fostering children's cognitive competences also at a more general cross-curricular level (cf. Adey et al., 2007).

Even though first grade cognitive competences explained to some extent later reading comprehension, early reading skills were not a particularly good predictor of later reading comprehension in the Helsinki schools even though much higher dependence was expected based on earlier studies (e.g. Leppänen et al., 2008). Of course, this may be partly due to the inaccurate measure of initial reading skills – the use of teacher evaluations instead of an objective measure – but still they were expected to be more related to later reading comprehension. What is more surprising, in Vantaa these two things were hardly at all related to each other when children's general cognitive competences were taken into account. A possible explanation for this could be found in the difference between initial reading skills: according to the teachers, the estimates of initial reading skills were on average higher in Helsinki than in Vantaa, and this may have made it easier for the teachers in Helsinki to evaluate pupils' reading skills accurately. The present study cannot answer, however, as to why children's initial reading skills were evaluated as higher in Helsinki. It is possible that teachers' expectations are systematically biased depending on the municipality, but it is just as likely that there are systematic municipal-level differences in how reading skills or prerequisites for them are emphasised in pre-school education. This finding should not depend on children's background as the educational level of children's mothers who participated was equal in both municipalities, there were no municipal-level differences in children's first grade cognitive competences and at a general level there were no differences in their performance in the assessment tasks at the turn of the third and fourth grade. However, when looking at the different domains of the assessment tasks separately, it can be seen that the Helsinki children still have a slight advantage in reading

comprehension and also reasoning – which may in part be explained by the use of partially verbal reasoning tasks – while in mathematical thinking the Vantaa children perform slightly better. Even though the performance of children with an immigrant background is not in the scope of the present study, it has to be noted here that the samples from the two cities did not differ from each other in this respect, so the differences in teacher-evaluated reading skills are not explained by that.

Early reading skills predicted slightly later performance in mathematical tasks in both cities, but they were not related to performance in the reasoning tasks. They were, however, moderately related to first grade cognitive competences which were strong predictors of both mathematical thinking and reasoning tasks both in Helsinki and in Vantaa. Thus, regarding reading skills it can be concluded that when the individual differences in cognitive competences are taken into account, better initial reading skills as evaluated by teachers are related to a slightly better achievement at the turn of third and fourth grade at least when achievement is measured by external assessment tasks. This may be explained by attitudes – which are difficult to measure in the first grade – as early readers had in this study slightly more positive attitudes in the third grade. This, however, would require further research with more rigorous measures of young children's motivational attitudes (cf. Lepola, 2004).

### **8.1.2 From the turn of the third and fourth grade to the end of the sixth grade**

In the second substudy, the Helsinki sample was followed from the beginning of the fourth grade to the end of the sixth grade. This substudy was really addressing the development of the cognitive learning to learn competences as only tasks and items which were identical in both test versions were used. Even though the main emphasis of the study was on changes in learning-related attitudes and task interest, both of which will be discussed later, the development of cognitive LTL competences in the whole sample will be presented here shortly.

During the follow-up period of three years, the children improved their performance in all the three tasks – of which two measured mathematical thinking and one Piagetian reasoning – by solving on

average 64 % of the tasks, compared to only 44 % in the fourth grade. The statistical significance of this effect was quite strong ( $t=-29.486$ ,  $p<.001$ ), and it can be concluded that children's mathematical thinking and reasoning skills develop relatively a lot at this age (cf., Csapó, 1997; Demetriou et al., 1991). As expected, fourth grade performance was a good predictor of sixth grade performance, and adding an external measure of general cognitive competences in the beginning of the fourth grade increased the share of explained variance of the sixth grade test score to 34 %. Thus, even though there was a clear relationship between earlier performance in the same tasks and general cognitive competences, more than 60 % of the variance was still explained by other factors. The small effects of changes in task interest and learning-related attitudes will be presented in the next section. In addition, there were systematic school-level effects in the cognitive tasks which increased during the follow-up period. This will be discussed in detail in the last part of this section, as even though it, at the same time, tells about education influencing children's performance in cognitive tasks (Adey et al., 2007), it also tells about differentiation of the schools regarding children's performance in the assessment (cf. Bernelius, 2013).

Sixth graders' performance in the cognitive learning to learn tasks was assessed also in Vantaa, but unlike in Helsinki, the only cognitive measure from three years earlier was the analogical reasoning task. Thus, the analyses were mainly cross-sectional, only controlling for the initial cognitive competences. Moreover, the sixth grade test score consisted of more tasks as in addition to those used in longitudinal analyses in Helsinki also verbal reasoning tasks and reading comprehension were used in the analyses. This means that the Helsinki and Vantaa results in the sixth grade are not directly comparable and only brief conclusions about Vantaa are made here as the main points of that study – the effects of attitudes on time investment – have not yet been discussed. Also in Vantaa, third grade analogical reasoning skills were a good predictor of sixth grade learning to learn test score but whether they provided any added value compared to performance in the LTL tasks cannot be answered here. In contrast, the Vantaa study shows clearest the importance of learning-related attitudes and task behaviour in explaining performance, which will be discussed later in this chapter. It also shows

how much more school grades are influenced by attitudes than external measures of achievement are. Therefore, school grades are covered briefly also here even though they were not the focus of the present study.

### ***School grades***

Whereas school grades can be understood as reflecting the cognitive competences which have been fostered during education and function as an initial state of future learning (Gustafsson & Carlstedt, 2006; Snow, 1996), they are not a pure measure of childrens' cognitive performance especially in lower grades (National Board of Education, 2004). Kenney-Benson and colleagues (2006) suggested that girls get better school grades because they have a higher level of mastery attitudes, which foster beneficial learning strategies leading to better grades. In the Finnish context it seems that pupils' get also directly rewarded or punished for their learning-related attitudes in regard to school grades even in the ninth grade (Kupiainen et al., 2014), and the results of the third substudy regarding the grades of the Vantaa sixth graders point also in the same direction. When school grades were used as the measure of initial cognitive competences in the analyses, the role of attitudes was small in explaining performance even though they correlated with school grades. Performance in the external reasoning task was much weaker in relation to later attitudes (cf. Kenney-Benson et al., 2006), even though later detrimental attitudes were not completely independent of it. Therefore, it was concluded that when possible, using an external measure of prior cognitive competence as a control variable instead of school grades is recommended in assessment studies as the results are then easier to interpret. On the other hand, school grades clearly paint a richer picture of pupils' competences than any single external measure as they were nevertheless stronger predictors of later performance in the learning to learn assessment. Regardless of that, school grades will be not discussed further in this study.

#### **8.1.3 Gender differences in the development of performance**

Based on their general cognitive competences, girls and boys should be able to perform equally well both in school subjects and in educational assessment studies. Traditionally, girls have had an advantage in reading

and boys in mathematics, which has been partially due to biological and developmental differences in the underlying cognitive abilities (Halpern, 2000), or using the language of the theory of the architecture of the mind by Demetriou and others (2011), in specialised structural systems. However, in the Western cultures girls have for decades got better school grades than boys (e.g. Kimball, 1989), and especially lately they have increased the gap from boys in external international assessment studies, not only in reading but also in areas which have traditionally been boys' strengths (OECD, 2013a). Therefore, the development of gender differences was of particular interest in the present study.

The first substudy started from an assumption that while it was expected that there are more girls than boys who can read already when they come to school, there are otherwise no differences in girls' and boys' cognitive competences, especially when they were measured by mainly visuo-spatial tasks. Girls got slightly better evaluations of their reading skills in the beginning of first grade both in Helsinki and in Vantaa. Even though the measure was only the teachers' estimated competence level, this corresponds with the results of earlier studies conducted with the same age group in Finland (e.g. Leppänen et al., 2008). Girls' small advantage in reading was seen also in their third grade reading comprehension test scores even when the initial differences were controlled for. Therefore, regarding reading, it can be concluded that the gap between girls and boys begins to grow already during the first three years of basic education (cf. Panula, 2013).

The results regarding gender differences in mathematical thinking are more difficult to interpret. First, when the analyses were performed for the both samples together, it looked like boys would be slightly better in mathematical tasks in the third grade even when initial cognitive competences were taken into account (cf. Halpern, 2000). In closer analyses it was however noticed that this gender difference applied only to Helsinki, and it was much stronger than girls' advantage was in reading. Thus, boys in Helsinki performed better than girls in mathematical assessment tasks in the beginning of the fourth grade, but substudy 2 showed that girls closed this gap by the end of the sixth grade. In Vantaa, on the contrary, there was no gender difference in the third grade mathematical tasks and in general the girls in Vantaa performed

better than boys in the sixth grade. However, when their initial cognitive competence – in which there was no gender difference – was controlled for, the “girl-effect” was completely mediated by more positive attitudes and increased effort in the assessment as measured by time investment in the task. This will be discussed further in the next sections of this chapter.

There is a lot of research about gender differences in mathematics, and with young children the results have been partially controversial (e.g. Aunio, Hautamäki, Sajaniemi & Van Luit, 2009). Nevertheless, as older boys have traditionally outperformed girls in mathematics (Halpern, 2000), it has been suggested that the development of these differences is partially social, being related to the different expectations and attributions of teachers (Jones & Dindia, 2004) and parents (Eccles, 2011; Kenney-Benson et al., 2006). One possible explanation to the municipal-level difference observed in the present study could be that for some reason the more traditional gender biases still prevailed in the Helsinki schools whereas in Vantaa this would not be an issue. Indeed, in Helsinki a greater proportion of group-level variation of the fourth grade mathematical tasks compared to other tasks was explained by gender groups within classes, which indicates that even within the same classes girls’ and boys’ mathematical thinking skills develop in a different way. It has to be noted that even though girls’ increased the gap compared to boys in reading, the same kind of gender group level effect within classes cannot be seen. In reasoning skills there were no gender differences in either of the two cities.

#### **8.1.4 Support needs and the development of performance**

In Finland, children who need support in their studies have since the beginning of the comprehensive education system been quite advantaged compared to children in many other countries (Graham & Jahnukainen, 2011; Sabel et al., 2011). This has also been seen in international comparative assessment studies like PISA (OECD, 2013a), in which the weakest pupils have in general performed much better than weak performers in other countries. The support system was however adjusted during 2007-2011 to meet the increasing support needs and to better follow the principles of prevention and early intervention. According to

the new three-tiered support model (see Thuneberg et al., 2013), children are usually not classified as special education pupils when they first come to school. Instead, during the first years of basic education general and intensified support should receive much stronger emphasis, and only if this is not sufficient, should special support be provided. Thus, in the present study – especially when most special education classes were excluded from the assessment – there were not many children who were officially recognised as having support needs in the beginning of first grade – even if the Helsinki first grade data were collected already in 2007. By the turn of the third and fourth grade, however, there were children who were receiving intensified support and also those who had been officially identified as having special education needs.

In the first substudy in which first graders' performance was followed to the turn of the third and fourth grade, the effectiveness of the support system – and on the other hand the validity of the first grade measures in identifying children in need of support – was evaluated by looking at how children who had by the beginning of the fourth grade been identified as having either intensified or special support needs were performing in the assessment. This was done by controlling the effects of the initial cognitive competences and reading skills; so the results tell about how they were doing compared to pupils with similar initial performance but who were not receiving intensified or special support. The basic assumption was that the differences between children who received support and other children should not increase during the follow-up period.

As expected, support needs were in strong negative relation with cognitive competences and reading skills in the beginning of the first grade for both samples. Thus, the pupils in need of support seemed to have been adequately identified at school. On the other hand, this can be interpreted as evidence of the validity of the first grade test from a practical perspective – it managed to identify the same pupils who were identified by other measures at school.

Receiving intensified or special support was related to children's background, which was measured through the educational level of their mothers. A mother's lower education increased the possibility of a child being identified as having support needs. Support needs were also

correlated with gender even though the relationship was not equally strong. This means that more boys had been identified as having support needs – a result that can be easily observed in any school also in Finland (cf. Henning-Stout & Close-Conoley, 1992; Thuneberg, 2007). During the first three school years this may be explained by boys' weaker reading skills in the beginning and the relatively strong connection between reading difficulties and support needs (cf. Halpern, 2000), but the results of this study still leave open the question about whether the support needs of girls are always adequately identified. Some of the gaps in girls' knowledge and skills may not be easily visible to teachers as girls seem to cover up some of the disadvantages they have in learning with positive attitudes and effort, as will be discussed here later. Also, based on other studies they do not externalise their problems as often as boys do (e.g., Leadbeater, Kuperminc, Blatt & Hertzog, 1999). Therefore, they may not always get the support they would need (cf., Panula, 2013).

Support needs did not directly predict reading comprehension or mathematical thinking at the turn of the third and fourth grade, so the initial differences had not increased in any of the two samples. Just the opposite, children who had received support for their studies even performed better than expected in the reasoning tasks. In other words, they had to some extent actually managed to close the gap with other pupils. As the reasoning items were partially overlapping with the items of the first grade test, this can be interpreted as an indicator of the effectiveness of the support system during the first three school years in both cities. However, even though the first graders of the two cities were not different in regard to their background, initial competences except for reading skills or performance at the turn of the third and fourth grade, there were twice as many children who were recognised as having at least mild support needs in Vantaa. This may tell more about the different policies of the municipalities in determining support needs than about actual differences between children – a phenomenon which has been visible in the official statistics of education for a long time ([www.stat.fi](http://www.stat.fi)).

While the first substudy reported about the progress of performance of children who had received support during the first three years of basic education, the results of the two other substudies with third to sixth graders were not equally positive. In the second substudy the gap



between pupils who had received support during grades 4 to 6 and others increased significantly when exactly the same tasks were used both in the beginning of the fourth grade and at the end of the sixth grade. That is, even though also those who had received support improved their performance, they did not improve it as much as the others did. The same could be seen when comparing the 25 % of children with the lowest school grades to their higher-achieving schoolmates even though the difference in improvement was not as big as it was for children with support needs. Even though small differences could be also observed in attitudes, they did not explain this kind of development in performance. These results fit well together with findings from international studies where the differences between pupils were found to increase over time, with higher performers improving their results more and lower performers getting even weaker (e.g., Shaywitz et al., 1995).

The results of sixth graders in Vantaa were very similar in this respect. Support needs were related to cognitive competences in the third grade, but even though the initial differences were taken into account, receiving support predicted significantly lower performance in the sixth grade learning to learn test. Even though the effect was partially mediated by attitudes and effort, they did not completely explain why sixth graders with support needs did not reach the level of performance which could have been expected based on their cognitive competences. This implies that even though the present study manages to cover many phenomena which are related to how children's learning to learn competences develop during the first six years of basic education, it fails to address some important – most likely psychological and social – factors which affect the performance of children with support needs.

### **8.1.5 Socio-economic background and the development of performance**

The effects of children's socioeconomic background, measured by the educational level of mothers, were examined only in the first substudy. Just as expected, there were differences in children's performance already in the beginning of the first grade due to mothers' education (cf. Caro et al., 2009), but unlike in studies using only school-level aggregates (see Kuusela, 2010), mothers' education explained less than 10 % of the

variance of first grade cognitive competences and even less of teachers' evaluation of reading skills in the beginning of the first grade. Mothers' education was also related to support needs.

Mothers' education did not at all predict performance at the turn of the third and fourth grade in any of the two samples when the differences in the first grade performance were controlled for. In other words, the initial differences did not grow over time at least when measured with the learning to learn test. This was also the expectation, as Caro and colleagues (2009) have earlier shown with a large-scale Canadian data that the socio-economic background-related differences begin to accelerate first when children are about 11 years old. It is also a positive finding regarding educational equity, especially as it applied to both Helsinki and Vantaa. In Helsinki it is more common for highly educated parents to choose schools in which most of the families have higher socio-economical status (Bernelius, 2013; Varjo & Kalalahti, 2011), and some of the fourth graders of the present study had in fact selected another school after the second grade unlike in Vantaa where almost everyone had stayed in their original schools. Without analysing the effects of any individual school the results of this study suggest that, in general, a more differentiated school choice policy does not seem to help children with higher educational background to gain more than other pupils, at least not during the first three years of compulsory education. In the future these analyses should be extended to higher grade levels, and the performance of children in the "magnet schools" should be examined separately.

### **8.1.6 School, class and peer group effects on the development of performance**

The last substudy concentrated on the group-level effects on performance, attitudes and the changes in them. The initial assumption was that systematic school effects would be stronger in Helsinki already in the beginning of basic education and that they would increase over time (cf. Bernelius, 2013). In Vantaa the systematic variation was expected to be mainly at a class-level throughout the basic education. In addition, within-class gender group level effects were expected to

increase by age, reflecting possible peer influences in learning and performance.

In the first grade, the school effects were indeed relatively small in Helsinki and even smaller in Vantaa. Even though the effects were not strong, they nevertheless showed that there were systematic differences between schools already in the very beginning of basic education especially in Helsinki. As at this stage most children still go to their local schools, even in municipalities with more possibilities for school choice, these differences can be interpreted as indicators of differentiation of residential areas (cf. Bernelius, 2013). Indeed, controlling for mothers' education decreased the school-level effects in Helsinki by a half, which means that the systematic school-level variation was mainly explained by the differences in pupil populations of the schools. As residential areas, and as a result of it local schools, are less differentiated in Vantaa the school-level effects were accordingly smaller, even though in one of the first grade tasks – the only one in which there was any school effect at all in Vantaa – the effect was in fact stronger than in Helsinki. This task required comprehension of verbal instructions, and one possible explanation of a stronger school effect in it can be the high number of immigrants in some schools. Controlling for the mother tongue of the children decreased the school-level effect in this task to the same level as it was in Helsinki without controlling for mothers' education, and controlling for mothers' education diminished it even further. Thus, also this school effect was explained by differences in pupil population.

As expected, in both Helsinki and in Vantaa there were systematic class-level effects that explained the variance of all the first grade cognitive tasks which were administered by teachers according to written instructions. However, the task which was most dependent on teachers, the graphic dictation, produced relatively small class effects. This was interpreted as a sign that the teachers were carefully following their instructions regarding how to administer this task. The biggest class effects (about 15 % of the variance explained) were in the visuo-spatial memory task, which was most likely telling about small differences in the length of time the stimuli were visible for pupils even though every teacher had received exactly the same instructions with the same time limits.

A little surprisingly, there was a moderate class effect in analogical reasoning skills in Helsinki (6.9 %) even though the task should have been the least dependent on teachers' instructions. This may tell about systematic differences between classes: In Finland children are not completely randomly assigned to first grade classes if there is more than one class in the same school. There is a lot of information available from daycare (Ahtola et al., 2011), which is used also in forming balanced classes and groups in addition to organising support. Thus, some of the class level differences in the beginning of basic education are actually produced consciously by the schools. This may be the reason for the observed class-level effect on initial reading skills too, which, however, can also tell about differences in teachers' expectation levels and their use of the evaluation scale.

The Vantaa results revealed something more about the differences between the first grade classes. The within-class gender group effect was in Vantaa relatively strong for analogical reasoning skills (5.9 %) while the class effect was in fact small (1.1 %). This may tell about an even less random assignment of pupils in classes in Vantaa and is most likely related to the imbalance between genders regarding support needs. In other words, the results may indicate that placing boys with support needs in a class has been compensated for by having higher-performing girls in the same class, or vice versa. This kind of difference between the two cities was visible also in the sixth grade results even though in Vantaa it tells about another age cohort.

When the Helsinki schools were followed from the beginning of the fourth grade to the end of the sixth grade, it was noticed that the school effects had increased already during the first three years, but in the beginning of the fourth grade they were not yet much stronger than the class- or peer group effects. By the end of the sixth grade most of the systematic group-level variation of the cognitive results was at a school-level. Using the words of Willms (2010), vertical segregation had increased during the follow-up period even though the same children were assessed at both times. It cannot be concluded, however, that the gap between higher- and lower-performing schools would have grown during the follow-up as additional analyses showed that there were initially lower-performing schools which improved their performance

relatively more while some initially higher-performing schools did not quite meet the expectations set based on children's fourth grade test scores. Thus, systematic Matthew-effects (see Bast & Reitsma, 1997) were visible only at an individual level: higher initial cognitive competences as measured by an external reasoning test helped in improving test performance more than average, and the gap between higher and lower achievers increased over time.

In Vantaa the sixth grade results were different. The systematic group-level variation of the cognitive results was still mostly between classes and their importance in explaining the results had even increased. The role of schools in explaining variation of the test scores was in contrary very small. In this, the results were close to the between-school differences at a national level (e.g. Hautamäki et al., 2013; Kupari et al., 2013), and they most likely reflect the different school choice policy of Vantaa compared to Helsinki (Bernelius, 2013; Varjo & Kalalahti, 2011). This pattern with smaller school- than class-level differences has been the most common in international comparisons even in countries with much more segregative educational systems (Willms, 2010), so the Helsinki results were somewhat surprising in this respect. Therefore, further research is needed in order to understand how initial class-level effects transform into school-level effects over the years.

Gender differences brought an interesting addition to this surprising observation in Helsinki. As already mentioned above, girls reached boys in mathematical thinking skills during the follow-up time, and there was no gender difference in reasoning or mathematical thinking in the sixth grade. In the beginning of the fourth grade, girls' results still followed the more common pattern in which their performance was partially explained by class-level effects, but schools explained very little of the variation of the results (Hautamäki et al., 2013). For boys, already at this age schools explained 13.6 % of the variation of the test scores – which in the sixth grade was approximately the case both for girls and boys. This may reflect the same phenomenon which was suggested as an explanation for the municipal-level differences presented above: There is something in the school culture, which is not dependent on classes or peer groups, which encourages boys in respect to mathematical tasks (cf. Kenney-Benson et al., 2006) still in the beginning of fourth grade. By the

end of the sixth grade, the girls catch up, but since at the same time the share of variance of girls' performance explained by schools increases dramatically from non-existent to 13.6 %, it looks like the improvement of girls' performance is accordingly partially a school-level phenomenon.

Also here, the Vantaa results from the sixth grade tell a somewhat different story. Even though systematic school-level effects were slightly stronger for boys than they were for girls, they were nevertheless very small compared to the Helsinki results. In contrast, the class-level effects were relatively strong for both boys and girls. Yet, from the perspective of the development of gender differences it is interesting to note that almost 24 % of the variance of boys' results was explained by either school or class-level effects while for girls the total percentage was only 16.4. Therefore, boys seemed to be more vulnerable to the effects of their environment at least when the assessment includes reading comprehension, which was not used in the Helsinki study. Also the results regarding attitudes, which will be presented in the next section, support this interpretation.

### **8.1.7 Summary of the development of performance in the cognitive tasks**

The present study shows that children's performance in the cognitive learning to learn tasks at the turn of the third and fourth grade can quite well be predicted by their initial cognitive competences and readings skills in the beginning of the first grade. However, more than half of the variation of the performance of mathematical thinking and reasoning, and even more of the variance of reading comprehension, was explained by other factors than initial individual differences. The same observation was done also when the development of the cognitive learning to learn competences was followed from the turn of the third and fourth grade to the end of the sixth grade: Earlier performance and cognitive competences played an important role in explaining later performance, but quite a lot of variation remained unexplained by them. Therefore, it was concluded that the development of more general thinking skills can clearly be influenced by means of education. There were systematic school, class and peer group effects on the development of performance in the cognitive learning to learn assessment tasks. In Vantaa, the group-

level differentiation happened mostly within schools whereas in Helsinki the schools – which were only slightly more differentiated in the beginning – seemed to differentiate more from each other during the first six years of basic education. However, systematic Matthew-effects could be observed only at an individual level.

Reading skills as evaluated by teachers in the beginning of the first grade were related to a slightly better achievement at the turn of the third and fourth grade even when general cognitive competences were taken into account. The effects were stronger in Helsinki where for some reason first graders were, according to their teachers, better readers than their agemates in Vantaa when they came to school. The gap between girls and boys in reading began to grow already during the first three years of basic education in both cities. In both Helsinki and in Vantaa, girls improved their performance slightly more also later in other domains. In Helsinki, however, they performed lower than boys in mathematical tasks in the beginning of the fourth grade, but they closed the gap by the end of the sixth grade. The results on attitudes, time investment and effort, which will be discussed later, indicate that the development of gender differences on performance in assessments depend at least to some extent on them. They also show that boys are more easily influenced by their school- and classmates of the same gender.

Mothers' education was related to children's initial cognitive competences, but the differences did not increase from the beginning of the first grade to the end of the third grade. A mother's lower education also increased the possibility of a child being identified as having support needs. There were more boys than girls who received support, and more pupils in Vantaa than in Helsinki who had been identified as having support needs at the turn of the third and fourth grade. As expected, support needs were in a strong negative relation with cognitive competences and reading skills in the beginning of the first grade for both samples. However, children who had received support for their studies even performed better than expected in the reasoning tasks at the turn of the third and fourth grade, so they had to some extent actually managed to close the gap with other children. This was interpreted as a sign of the effectiveness of the support system during lower grades. Later, the results were unfortunately not equally positive: During grades 4 to 6

the gap increased again, and even though the effect was partly mediated by attitudes and effort, these did not completely explain why sixth graders with support needs did not reach the level of performance which could have been expected based on their cognitive competences. The present study clearly did not manage to address some important – most likely psychological and socio-emotional – factors which could partially explain why the differences increase over time. Nevertheless, it shows how important it is to evaluate the effectiveness of the individually designed support by attitude-related measures in addition to the perhaps more common achievement-related or behavioural measures.

## **8.2 Learning-related attitudes: Their development and effects on performance**

As there is a firm understanding that attitudes affect learning and educational outcomes (e.g. Eccles & Wigfield, 2002; Demetriou et al., 2011; Harackiewicz et al., 2002; Kenney-Benson et al., 2006; Mangels et al., 2006; Patrick et al., 1999; Snow, 1996), and they form a central part of the Finnish learning to learn framework (Hautamäki et al., 2002), it was expected that some of the individual and group-level differences in the performance in cognitive assessment tasks would be explained by differences in the development of learning-related attitudes. As answering to the attitude questionnaires requires moderate reading skills and a capacity for self-evaluation, the development of attitudes was followed only from the turn of the third and fourth grade until the end of the sixth grade. Even then it is questionable as to whether 10-year-olds are mature enough to evaluate themselves (Demetriou & Kazi, 2006; Harter, 1999), so the aims of this study were twofold. On the other hand, the purpose was to examine at which stage children's self-evaluations begin to play a role in explaining performance in the context of Finnish learning to learn assessments. On the other hand – regardless of the relationships with performance – the aim was to understand how attitudes develop in lower grades: to what extent the well-documented decrease of attitudes and interest (see Ainley et al., 2002b) is due to increasing self-awareness and self-evaluation capacity (Demetriou &



Kazi, 2006), and to what extent it is a group phenomenon in which classmates and peers play an important role (see Ryan, 2012).

In the first substudy, which tried to find predictors for children's performance in the cognitive learning to learn tasks at the turn of third and fourth grade, it was expected that besides earlier competences and reading skills, learning-related attitudes would explain some variance of their performance as well. Children were in general very positive in their self-evaluations in all three measured areas – studying in order to learn things and for getting good grades, and working hard at school – and the relationships of attitudes with actual performance was very weak. That is, also lower-performing children reported high levels of positive attitudes towards learning and going to school. This was however to be expected due to the young age of the children (cf. Demetriou & Kazi, 2006). Of the cognitive domains of learning to learn, attitudes were related only to mathematical thinking in which more positive attitudes predicted slightly higher test scores. Early reading skills, as reported by teachers, were also related to learning-related attitudes positively but very weakly. Regarding attitudes, there were no differences between children in Helsinki and Vantaa.

The second substudy – and another study which used the full six-year Helsinki data (Vainikainen, Wüstenberg, Kupiainen, Hotulainen & Hautamäki, submitted) – shed more light on the development of these positive attitudes and their increasing relationships to test performance. Substudy 2 shows that in general all learning-related attitudes decreased significantly during the follow-up period, but at the same time their value in explaining test performance increased (Vainikainen et al., submitted). The magnitude of change however did not directly explain performance, but it was related to changes in task-specific interest, which provided some added value in understanding the children's performance in the sixth grade assessment (cf. Renninger & Hidi, 2011).

The results support in general the findings of earlier studies about the development of learning-related attitudes of the children of the age of the participants of the present study (e.g. Kenney-Benson et al., 2006) even though there are not many studies which manage to partial out the effects of prior cognitive competences or school achievement from the effects of attitudes. This is however important as prior competences often explain a

part of the variation of attitudes, and when the effects of competences are taken into account, the explanatory power of attitudes and other affective factors usually decreases (see Ainley et al., 2002a). In the present study and the other study using the same data, however, fourth grade attitudes (Vainikainen et al., submitted) or the change in them by the end of the sixth grade (Substudy 2), were not related to prior cognitive competences. This tells about the validity of the measures as they do not too much depend on actual competence level of children – in substudy 2 the change scales were also confirmed to be measurement invariant for children with different levels of school achievement or support needs. What is more important, the results suggest that regardless of achievement level, targeting the changes in children's learning-related attitudes by means of interventions could at this age lead to higher interest in school assignments or assessment tasks or greater effort in form of time investment (cf. Kupiainen et al., 2014; Substudy 3). They in turn could enhance performance both in everyday school assignments and external assessments (cf. Renninger & Hidi, 2011). To do this, the social aspects of attitude development need to be understood better, and regarding the results of the present study this will be discussed further in the last part of this section.

The third substudy introduced detrimental attitudes – believing in the role of luck or abilities in explaining performance and self-handicapping strategies – in addition to the mastery attitudes which are above referred to simply as positive attitudes. The results showed that they both played an independent role in explaining sixth grade performance even though the role of attitudes was much weaker than it was for ninth graders in the study of Kupiainen and others (2014), thus possibly reflecting younger pupils limited self-awareness (Demetriou & Kazi, 2006). The results confirmed also the assumption that attitudes have a much stronger connection to school grades than to external assessment tasks (cf. Kenney-Benson et al., 2006), which was already discussed shortly above when the use of school grades as a measure of prior cognitive competences was evaluated. Detrimental attitudes – even if measured in the sixth grade – were however related also to general cognitive competences in the third grade, which may be explained by the fact that children with a high level of detrimental attitudes in the sixth grade had

some issues with attitudes already in the third grade and therefore they did not try their best in the third grade assessment either. Nevertheless, despite the negative correlation with detrimental attitudes, it could be concluded that an external measure, in this case the analogical reasoning test, made the independent role of attitudes in explaining test score more visible in later analyses.

A little surprisingly, mastery and detrimental attitudes were in this study not related to each other at all even though in earlier studies with older participants they have correlated negatively (Hautamäki et al, 2013; Kupiainen et al., 2014). This confirms the interpretation that even in the sixth grade, the 12-year-olds are not very mature in their self-evaluations (cf. Demetriou & Kazi, 2006) even though their relationships with performance were much clearer than it was for the fourth graders in the second substudy. Regardless of the limitations, there were some important subgroup-level differences in how attitudes were related to test performance in the sixth grade, and they will be presented in detail in the following parts of this section and especially when the effects of time investment on performance is discussed.

### **8.2.1 Gender differences in the development of learning-related attitudes**

One of the central hypotheses of this study was that the increasing gender differences in school achievement and assessment results could be explained by gender differences in the development of learning-related attitudes, interest and effort. Earlier research has shown that boys who have low prerequisites for reading can undergo negative motivational changes already as early as in pre-school age (Lepola, 2004), and it is likely that it influences their later task behaviour and performance.

Indeed, the results of the first substudy showed that already at the turn of third and fourth grade, when attitudes had little to do with actual performance, girls' attitudes were nevertheless more positive than boys' both in Helsinki and Vantaa (cf. Kenney-Benson et al., 2006). Interestingly, taking gender into account also increased slightly the effect of learning-related attitudes on mathematical thinking skills, which was the only competence area where attitudes played any role at this age. This suggests that girls could to some extent cover up the disadvantage they

had in mathematical tasks – at least in Helsinki – with positive attitudes and possibly also greater effort which was not measured in the first substudy. This interpretation is however in contradiction with much of the literature regarding gendered educational choices and expectancy values (Eccles, 2011), which start from an assumption that girls' lower self-perceptions and subjective task values regarding mathematics causes them to make less ambitious choices in mathematics in the later years of education. Even though girls' later choices regarding mathematics have been seen as a problem in Finland too, in the fourth grade, girls' learning-related attitudes – if they have any effect at all – can enhance their performance at least in the mathematical tasks of an external assessment study. This could be an explanation for older girls' superiority in most other assessed areas too, and it made it important to study the development of learning-related attitudes, task interest and effort closer in the following substudies.

The results of the second substudy are less straightforward to interpret. While girls' attitudes in the beginning of the fourth grade in Helsinki were more positive than boys' as described above, they also decreased more than boys' attitudes by the end of the sixth grade, and the gender difference in attitudes was no longer statistically significant. As the cognitive tasks used in this substudy did not include reading comprehension, there was no gender difference in the test scores either, and the effects of changes in learning-related attitudes and task interest were similar for both genders. It has to be noted, however, that even though girls' attitudes had decreased more, their performance in the cognitive tasks had nevertheless improved relatively more as they had now reached boys in mathematical thinking. This pattern is not quite what was expected considering the results of earlier studies regarding gender differences in the relationships of mastery attitudes and mathematical test performance (e.g. Kenney-Benson et al., 2006), and based on the results of the second substudy further research was needed to understand the mechanisms of how attitudes and test performance are related at this age.

Some answers were obtained already in the next substudy in which it was investigated how attitudes influence test performance through time investment. Gender difference in the sixth graders' mastery attitudes in

Vantaa was small but statistically significant, and there was an equally weak gender difference in time investment in the assessment tasks. Girls were slightly more positive than boys in their attitudes, and they also spent statistically significantly more time on tasks than boys did (cf. Ainley et al., 2002b). When these effects were taken into account, gender was no longer directly related to performance even though girls clearly received higher test scores than boys. In this substudy, their superiority in the assessment tasks was completely explained by attitudes and effort when support needs – which were related to performance negatively – were controlled for. Support needs were more strongly associated with gender, performance, time investment and detrimental attitudes, and they will be discussed below after presenting first how this development begins from the turn of the third and fourth grade.

### **8.2.2 Support needs and the development of learning-related attitudes**

In this study, the negative relationship between support needs and learning-related attitudes was visible both in Helsinki and Vantaa already when attitudes were measured for the first time when the children were about 10 years old. This could have been partly due to how one of the subareas of attitudes, achievement orientation, was measured – for children on a very low performance level it may not be a realistic goal to get good school grades – but motivation to learn new things and to try as hard as possible were not expected to be directly related to the initial level of performance. However, the second substudy, which took a closer look at the attitudes in Helsinki, proved this assumption partly incorrect. When the fourth graders were grouped according to their school achievement, the lowest performers reported indeed lower achievement orientation and also lower effort, and these differences compared to higher achievers persisted until the end of the sixth grade. Support needs, in contrast, were only related to learning orientation: children who had been identified as having support needs were less motivated than others to learn and understand new things. In the sixth grade this difference was no longer visible as the other children's learning orientation had decreased relatively more, and their attitudes were now on the same level as the attitudes of children receiving support.

Regarding positive learning-related attitudes – or mastery attitudes – the results of the sixth graders in Vantaa were very similar. While support needs were related to lower levels of positive attitudes for the three years younger age cohort, in the sixth graders' cohort there were no differences in this respect. Regarding detrimental attitudes, which were analysed only in the third substudy, the results were however unambiguous. In addition to being directly related to lower test performance and time investment, support needs predicted significantly higher levels of detrimental attitudes, which also influenced the test scores of these children. That is, sixth graders who had been identified as having support needs, had stronger beliefs that school success and learning depends on factors – abilities and chance – which are out of their own control. They also had more self-handicapping strategies like giving up easily in front of more demanding tasks. The third substudy show clearly that sixth graders who have been identified as having support needs perform significantly worse than their general cognitive competences would suggest, and this is partly due to the combination of reduced effort and a higher level of detrimental attitudes. This finding fits well together with the earlier results of Thuneberg (2007), where it was found that special education pupils belong much more often than average pupils to the pupil profile type who have particularly low motivation in addition to low achievement.

Together the results of the second and the third substudies indicate that whereas younger children with support needs differ from their agetates in that they have slightly less positive attitudes, this pattern changes during the three years so that they develop more negative attitudes despite still having quite high levels of positive attitudes too. As these negative or detrimental attitudes have been shown to be directly related to task behaviour and therefore to task performance during higher grades (Kupiainen et al., 2014), it would be important to try to stop this development during lower grades when it still may be possible (Hamm et al., 2012). Thus, the support system should systematically address also the development of learning-related attitudes and not only concentrate on following the progress and effectiveness of support by achievement- or behavioural-related measures.

### **8.2.3 Socio-economic background and the development of learning-related attitudes**

Children's socio-economic background and the development of learning-related attitudes were addressed in this study only very superficially. Even though mothers' education and test performance have been shown to be related in higher grades both in earlier learning to learn studies (e.g. Hautamäki et al., 2013) and subject-specific assessments (e.g. Kärnä et al., 2012), in the present study these relationships were analysed only in the first substudy regarding mastery attitudes at the turn of the third and fourth grade in Helsinki and Vantaa. The results showed that 10-year-olds' attitudes were not related to the educational level of their mothers. Considering the very high average level of attitudes and the almost non-existing relationship between attitudes and actual performance, it can only be concluded that children's self-evaluation skills are at this age so limited (Harter, 1999) that self-reported attitudes do not (yet) reflect the possible differences in the parental support for schoolwork.

### **8.2.4 School, class and peer group effects on the development of learning-related attitudes**

Probably the most central hypothesis in the last substudy was that the development of learning-related attitudes partly depend on how school- and classmates' and smaller peer groups' attitudes develop at the same time. Identifying early signs of problematic group-level attitude development is highly important because at this age it is still possible to successfully influence the development of effort- and achievement – oriented peer group culture by means of interventions in school settings (Hamm et al., 2012).

In the fourth grade the Helsinki schools did not have any systematic effects on the very positive attitudes, but Learning orientation – having learning of new things as a goal itself without external rewards – was to some extent (5.5 %) explained by classmates' self-evaluations. Peer groups – classmates of the same gender – explained 7.6 % of the variance of Achievement orientation and some variance of self-reported effort in schoolwork. This means that already in the fourth grade there could be separate girl or boy group norms in the same class regarding school

grades or performance in exams, and about effort (cf., Hamm et al., 2012). In the sixth grade also learning-orientation had gone from a partly class-level phenomenon to a partly gender-group level phenomenon, and in addition to these peer influences small school effects could also be seen.

Even more interesting is the change of attitudes in schools, classes and peer groups, and particularly the gender differences in them. When girls and boys were looked at together, it first seemed that little of the change had happened at a group level: Gender groups within classes explained some of the variance of the change of Learning and Achievement orientation and a little bit more of the variation of the changes in self-reported effort in schoolwork, but the effects were not strong. When gender differences were looked at more closely it was noticed that some of the changes in girls' attitudes had in fact happened at a school and not at a class-level. However, schools still explained only two to three per cent of the variance. For boys, in contrast, a somewhat stronger class-level effect was found for the changes in self-reported effort in schoolwork: 8.6 % of the variation of boys' self-reported effort changes was explained by the changes of the other boys in the class. Thus, self-reported effort change was really partially a group-phenomenon for boys. In Vantaa this was then visible in the form of boys' stronger peer effects on actual effort as measured by time investment, which will be discussed later.

For Vantaa's sixth graders the results regarding mastery attitudes – achievement orientation, importance of school and effort in schoolwork – were somewhat different from the Helsinki results. Systematic school effects were slightly stronger but still quite weak (less than four per cent), but unlike in Helsinki there was also a class-level effect of the same magnitude. Nevertheless, classmates of the same gender were also in Vantaa influencing each others' positive attitudes the most. When the effects were looked at separately for boys and girls, something interesting was noticed, however. The school-level effect seemed to apply only for boys' who were moderately influenced by other boys in the whole school (6.7 % of the variance explained). Classmates, however, were of equal importance both for girls' and boys' mastery attitudes (about 8 % of the variance explained). The hypothesis about boys' greater influence on each



others' attitudes (cf. Mähönen et al., 2011) was further confirmed when detrimental attitudes were taken into analyses. While in the whole data the variation of them seemed to be mostly within classes and slightly in schools, but gender groups did not explain it at all, in separate analyses it was seen that these effects were only telling about boys: The variation of girls' detrimental attitudes was 99 % individual. Boys, in contrast, were influenced by their schoolmates (2.7 %) and classmates (4.3 %) of the same gender.

### **8.2.5 Summary of the development of learning-related attitudes**

The present study followed the development of learning-related attitudes from the turn of the third and fourth grade to the end of the sixth grade. The results showed that in general children's attitudes were very positive when they were first measured but that they decreased by the end of the sixth grade, becoming at the same time more related to their actual performance. In this, the results support earlier literature about the development of children's capacity for self-evaluation. Even though learning-related attitudes at the turn of the third and fourth grade were only related to performance in the domain of mathematical thinking, some important group-level differences were observed in the development of these attitudes during the three following years. In Helsinki, girls who performed slightly weaker than boys in mathematical thinking in the beginning of the fourth grade could in fact compensate to some extent with more positive attitudes. However, their attitudes decreased to boys' level by the end of the sixth grade even though they improved their mathematical performance relatively more. In Vantaa, the gender difference in positive attitudes was visible both for the third and sixth graders and in the sixth grade its influence on test performance was unambiguous: Girls' higher mastery attitudes and increased effort as measured by their time investment on tasks explained completely their better test scores. Having support needs was at the turn of the third and fourth grade related to a lower level of mastery attitudes – especially to the motivation to learn new things having that as a goal in itself – but by the end of the sixth grade other children's mastery attitudes decreased to the same level. In the sixth grade, however, children with support needs

had significantly more detrimental attitudes, which also predicted lower performance than would have been expected based on their earlier performance.

The results also show that attitude development is partially a school-, class- and peer group level phenomenon. Especially boys are vulnerable to the influence of the other boys in their class and to some extent of the other boys of the same age in the whole school. This was particularly clear in regard to the development of self-reported effort in schoolwork and detrimental attitudes in the sixth grade, which for girls were mainly individual phenomena. The evidence from all three substudies with children from both Helsinki and Vantaa leads to one conclusion regarding attitudes: while boys' performance is not necessarily directly explained by group-level factors much more than girls' performance, their attitudes and effort – both self-reported and more objectively measured – clearly are. The results strongly indicate that boys' development of attitudes and effort – which already during lower grades happen partially in interaction with classmates of the same gender – can explain why they do not perform as well as girls do in educational assessment studies even though based on their general cognitive competences they should be able to. This of course can be reflected also in their daily schoolwork and school achievement, which can then have cumulative consequences in their later school career.

### **8.3 Task interest and performance in the assessment**

Task interest – both individual and situational (see Renninger & Hidi, 2011) – has been associated with enhanced performance (e.g. Van Yperen, 2003). What makes it particularly important for the present study is that girls have been found to put more effort also into tasks they find uninteresting while boys give up or rush through tasks more easily (Ainley et al., 2002b). Thus, also task interest and changes in it may play a role when gender differences in educational assessment studies are sort to be understood. Therefore, the second substudy concentrated on changes in children's learning-related attitudes and task-specific interest from the fourth to sixth grade by examining the effects of changes on performance in different subgroups in the learning to learn assessment at the end of the sixth grade. In the last substudy the understanding of the

results was deepened by analysing whether the changes were in part explained by systematic school, class or peer group level effects as they would possibly be easier targets for school-based interventions.

In the second substudy it was first examined how children's interest towards the two mathematical and one reasoning task used in the study had changed from the beginning of the fourth to the end of the sixth grade. As expected, fourth graders evaluated the tasks as more interesting than sixth graders. Since the same children were evaluating the tasks at both times, it could be concluded that task-specific interest seems to decrease by age (Renninger & Hidi, 2011) just like more general learning-related attitudes do. The change also happened regardless of the contents of the tasks. It has to be remembered, however, that reading comprehension tasks were not included in the present analyses. The changes in task interest were not related to the children's earlier cognitive competences or performance in the fourth grade assessment, so the changes did not reflect the fact that some children experienced the tasks as being more difficult for themselves than others did. This was further examined by dividing children into four groups based on their school achievement. The results will be discussed below together with support needs.

As expected, changes in task interest predicted children's test performance in the sixth grade. The effect was not very strong – like they never are when the effects of affective factors are evaluated by first controlling for prior competences (e.g. Ainley et al., 2002a). Nevertheless, the results showed clearly that children's performance in learning to learn assessment tasks change in the same direction as their self-reported interest in the tasks. That is, if the change in task interest was negative, children improved their performance in the tasks slightly less than what their earlier performance and cognitive competences suggested they would do. If the change was positive, their later performance was slightly better than expected. The same effect was not found for more general learning-related attitudes: even if they in another study and using the same data (Vainikainen et al., submitted) were found to be related to sixth grade performance, the change in them from the fourth grade did not explain the changes in performance. The changes of more general learning-related attitudes and task-specific interest were

however quite strongly related to each other (cf. Renninger & Hidi, 2011), so they can indirectly affect performance through interest and task behaviour (cf. Ainley et al., 2002b). This hypothesis could unfortunately not be tested in the present study as task behaviour analyses would have required log data of computer-based assessment and the Helsinki data were collected entirely on paper.

### **8.3.1 Gender differences in task interest**

When gender differences were examined in this substudy, it was expected, based on earlier literature, that changes in task interest would affect boys' performance more as they have in earlier studies been found to put less effort into tasks they find uninteresting (Ainley et al., 2002b). On the other hand, it was assumed that girls' – whose general learning-related attitudes had in Helsinki decreased to the same level of boys' attitudes by the sixth grade – performance would be more influenced by these attitude changes. Both these assumptions proved to be incorrect. The only gender difference in task interest was found in the water level reasoning task, which girls found more interesting both times. However, girls' and boys' interest had decreased equally and the relationships between this decrease and test performance in the fourth and the sixth grade were similar for both genders. Thus, without taking into account task behaviour as measured by time investment in tasks, the results of this study contradict Ainley's and colleagues' study (2002b) in which interest affected boys' time investment which in turn affected their performance. Therefore, more research on the computer-based learning to learn data is needed before conclusions can be made about gender differences in task interest. Based on this study it can be concluded that the decrease of self-reported interest in tasks seems to be equally harmful for both girls and boys, and it must be understood if this decrease is partly a group-level phenomenon.

### **8.3.2 Support needs and task interest**

As already discussed above, both academic achievement and support needs were in this substudy related to the level of performance. There were significant differences between the four school achievement groups

and children in need of support compared to the others in their fourth grade test performance, and unfortunately the differences also increased slightly over time. It was therefore examined whether changes in attitudes and task interest would partly explain this growth. As reported above, school achievement groups differed from each other on some of the attitude subscales, but there were no differences in the magnitude of the decrease of them. The differences in attitudes of children with support needs were even smaller. Regarding task interest there were no group differences whatsoever. Task interest had developed in a similar way in all school achievement groups and for children receiving support for their studies, and their relationships with performance was equal in all groups. In other words, changes in task interest and learning-related attitudes correlated, but only task interest changes predicted later performance over and above fourth grade performance and general cognitive competences. The only difference between children with support needs and others was a little surprising: In both the fourth and sixth grade the pupils with support needs reported higher levels of interest than the others in the Arithmetical Operations task, in which the operators (+, -, \*, /) were replaced by letters, and they had to find out which operator produces the given result. A possible explanation for this may be that the first items of the task are relatively easy and also low performers may feel more successful in them compared to many other task in which the instruction is already more complex. Based on this result, it would be recommendable to always include some easier items in the beginning of assessment tasks to give also weaker performers a chance to feel successful. This of course increases the testing time, but in the future adaptive solutions in computer-based assessment can be of help here. In an adaptive test higher performers could pass the easy items relatively quickly while weaker performers would not even need to try to advance to highly demanding items, which they would not be able to solve anyway. This in turn could help in preventing the development of detrimental attitudes of the children in need of support: As seen in the third substudy, sixth graders who needed support for their studies believed stronger than others that success is a matter of luck, and that high-level abilities are needed in order to succeed in school; they also reported more self-handicapping strategies. Differentiating teaching is

one of the most central means of support (Thuneberg et al., 2013), and now the next step should be the differentiation of assessment as well, which is nowadays possible with the modern computer-based assessment platforms.

### **8.3.3 School, class and peer group effects on task interest**

The last substudy showed that unlike the development of performance, many of the changes in attitudes and interest had happened mainly at an individual level both for girls and boys, and less than 10 % of the variance was explained by group-level factors. Interestingly, the little that had happened at a group-level followed a pattern which is difficult to interpret but which may be highly useful information when designing interventions for enhancing girls' and boys' attitudes and interest. The results showed that girls' general attitude changes were slightly explained by schools while task interest changes had happened in classes, following to some extent the development of interest for other girls in the class. Therefore, if girls' attitudes towards specific tasks or assessment situations are to be influenced by means of interventions, it is necessary to go to the peer group level to try to identify group norms that may lead to unwanted development of interest (cf. Hamm et al., 2012). However, for boys the pattern was just the opposite: Boys' interest changes regarding mathematical tasks were relatively strongly explained by how other boys' interest had changed in the whole school, and the role of their own classmates of the same gender was not important. General attitude changes especially regarding effort in schoolwork, in contrast, had happened in interaction with boys in their own class, and it will also be discussed in the next section how boys' time investment – a more objective measure for effort – depended more strongly on classmates' task behaviour. Thus, for boys it would be extremely important to address effort-related group norms at a peer group level so as to avoid possible underperformance in the assessments. On the other hand, the results that interest development and also fourth grade performance were for them partly explained by school-level instead of class-level factors and support the interpretation of the results of the first substudy that in Helsinki there seems to be school cultures which encourage boys to mathematics during the first few grades. It is possible that regarding

interest in mathematical tasks boys are initially more similar to other boys in the same school, and therefore also the changes in their interest are related. This interpretation would however require additional analyses which were not performed in the present study.

### **8.3.4 Summary on task interest in relation to performance in the assessment**

To summarise the results regarding task interest, it can be concluded that changes in task interest predict later performance over and above earlier performance and analogical reasoning skills. The changes in task interest are related to changes of the more general mastery attitudes, which however do not directly explain performance changes further. The levels of both task interest and learning-related attitudes decline quite evenly in all subgroups – girls and boys, children on different school achievement levels and children who need support for their studies – from the fourth to the sixth grade, and the relationships between the changes of them and performance is similar for all subgroups. However, there are gender differences in how task interest seems to develop in interaction with other children: For boys interest develops partly at a school level, indicating that school-level programmes for enhancing interest and task behaviour would be recommendable for enhancing their performance in the assessments. For girls, however, interest development seems to happen mainly in interaction with peer groups – other girls in the same class – which makes it important to address group norms regarding interest too.

The results of the present study support Hulleman and colleagues' (2008) assumption that mastery attitudes are related to interest and interest to performance – or in the case of the present study the changes in them. However, the mediation hypothesis of Hulleman and colleagues did not get much support in this study with 10 to 12 year olds due to the weak relationship between any affective construct and performance. Regardless of the weak connections to performance, the general decrease in both interest and attitudes from the fourth to the sixth grade is in compliance with the findings of earlier studies with school-age children (see Renninger & Hidi, 2011, for a review), and this may be partly

explained by children's increasing capacity for self-evaluation at this age (Demetriou & Kazi, 2006).

#### **8.4 Time investment and performance in the assessment**

Computer-based assessment provides opportunities for analysing children's task behaviour in ways which are not possible when the assessments are conducted traditionally on paper. There is an increasing body of evidence from log file analyses that task behaviour – for example strategy use and time investment – can have relatively strong effects on outcomes (e.g. Ainley et al., 2002b; Goldhammer et al., 2014; Kupiainen et al., 2014; Wise & Kong, 2005). However, relatively little is known about the factors which affect task behaviour especially during the lower grades. Kupiainen and colleagues (2014) showed, with a nationally representative data of Finnish ninth graders, that detrimental attitudes made pupils invest less time on task which affected their performance negatively when their prior competences were taken into account. Mastery attitudes, on the other hand increased time investment, but the effect was much smaller. In the present study it was first examined whether the same patterns could be found with three years younger children, and after that the emphasis was on gender differences and the effects of support needs which were not addressed in the earlier study.

The results showed that time investment was an important predictor of sixth graders' test performance: The more time the children spent on doing the tasks the better their results were. The magnitude of the effect was more or less the same as it was for ninth grade pupils in the earlier study of Kupiainen and others (2014) with approximately the same measures. Both prior school achievement and general cognitive competences as measured by an analogical reasoning task three years earlier were positively related to time investment. That is, children with higher cognitive competences or better school achievement spent slightly more time on tasks. This does not quite fit together with the expectations deriving from Carroll's (1963) model regarding lower achievers' increased need for time. On the other hand, the result supports the findings of Goldhammer and colleagues (2014) that in more complex problem-solving situations time investment is usually related to better outcomes while in more basic curricular tasks the relationship is negative. That is,



higher performers can do basic tasks quickly and still correctly, but in problem-solving they also need to use more effort in order to get good outcomes. Still, lower performers should not spend any less time on doing the tasks when putting sufficient effort on the tasks. This will be discussed when the results regarding support needs are presented.

Just like in the study of Kupiainen and colleagues, the effects of attitudes on performance were at least partly mediated by time investment in the assessment tasks. However, for sixth graders' this happened mostly with mastery attitudes instead of detrimental ones, which may be due to the limited self-awareness of 12-year-old pupils. Demetriou and Kazi (2006) showed that pupils' self-awareness increases dramatically from the age of 11 to 15, and for 12-year-olds it can be easier to evaluate themselves through positive statements instead of negative ones. This may also be the explanation as to why the effects of attitudes were in general weaker than in the earlier study. Comparison with the earlier study must however be done with reservations, as instead of using latent factors only manifest variables were used in this study in the SEM models. The composition of the sixth grade LTL test was at the item-level also not identical with the ninth grade version used in the other study. Nevertheless, it can be concluded that time on task mediates the effects of attitudes also for 12-year-olds even though the mechanism of the effects of attitudes seem to be slightly different for younger pupils.

#### **8.4.1 Gender differences in time investment**

Attitudes and interest have been shown to influence time investment (Ainley et al., 2002b; Kupiainen et al., 2014), and in earlier studies girls have been found to have more mastery attitudes and to some extent also less detrimental attitudes (Kenney-Benson et al., 2006; see also Huesmann et al., 2009; Olweus, 1979). Therefore, it was expected that there would be gender differences in how attitudes affect test performance through time on task too. It was hypothesised that this would in part explain girls' better assessment results even though there should be no gender difference in girls' and boys' general cognitive competence (Halpern, 2000).

As already discussed above, no gender differences could be found in Vantaa's sixth graders general cognitive competences three years earlier,

but girls showed clearly better performance in the sixth grade assessment tasks. Being a girl was also weakly but significantly related to a higher level of mastery attitudes. Gender was also related to time investment in the assessment so that girls spent on average slightly more time on tasks than boys did. Together these completely explained girls' superiority in the sixth grade LTL test: When mastery attitudes and time investment were taken into account, gender did not explain any additional variance of the sixth grade test scores. This finding is of high importance as it indicates that also more generally girls' higher performance in assessments – which has lately become an increasing concern also internationally (e.g. OECD, 2013a) – may be simply explained by their task behaviour which reflects their more general attitudes towards learning at school. If this is true, “the boy problem” could be addressed by intervention programmes targeted at enhancing boys' effort in all schoolwork. And since the development of attitudes and also the effort demonstrated in the assessment situation are partly group-level phenomena, it would be crucial to develop tools for teachers for better identifying early signs of the development of detrimental group norms among boys in the class. However, before it is possible to suggest more detailed plans for intervention development, more research on the relationships between attitudes, effort and task behaviour is needed. Further analyses of the learning to learn log data could provide many answers, and this kind of research should also be extended to utilising other computer-based large-scale assessment data, for example the PISA-databases.

#### **8.4.2 Support needs and time investment**

In Carroll's (1963) classic model on the relationship between time on task and learning time needed depends on the initial competences of the child. According to Carroll, learning is determined by the ratio of the time needed and the time spent. Based on this, children in need of support for studies should use more time than others for doing the tasks as their initial cognitive competences are significantly lower than others' as shown also by the first substudy. Putting it the other way around, it may be possible that the growth of the difference in performance between children in need of support and others from the fourth to the sixth grade

in the second substudy is partly related to insufficient effort which in computer-based assessment could have been checked from the log data. Therefore, it was important in the third substudy to analyse in detail how support needs were related to sixth graders' attitudes and time investment.

Indeed, Carroll's assumption of increased need of time was not met: Children who had been identified as having support needs spent slightly less time on tasks than others with similar prior cognitive competences – and earlier it was already concluded that children with lower cognitive competences invested generally less time in tasks. This of course may be partly due to the last items of each task being simply too difficult for them to even try to solve them, but nevertheless, both lower cognitive competences and support needs were related to slightly reduced time investment.

Even more important is that support needs predicted higher levels of detrimental attitudes which in turn predicted clearly lower test performance and also slightly reduced time investment. In addition, support needs were still directly related to lower test performance even though all these other factors were taken into account. As in other substudies and earlier literature (e.g. Henning-Stout & Close-Conoley, 1992; Thuneberg, 2007), support needs were also related to gender. Thus, even if the first substudy showed that during the first three school years the support system manages to keep the differences between children relatively stable, the two other substudies showed that this is unfortunately not the case any more from the turn of the third and fourth grade to the end of the sixth grade (cf. Caro et al., 2009). The results presented here indicate that this growth may partly be due to the negative development of detrimental attitudes and reduced effort at least in the assessment situation. As shown above, the development of detrimental attitudes and self-reported effort was partly a peer group-level phenomenon especially for boys, and before making further conclusions it needs to be discussed how time investment depends on school- and classmates' task behaviour in the assessment situation.

### **8.4.3 School, class and peer group effects on time investment**

When the variation of children's time investment was in the last substudy divided into school, class and peer group levels, it was noticed that the percentages of explained variance of each level were somewhat different from the other variables of this study. Individual differences explained only slightly more than 70 % of the variation of time investment which means that almost thirty percent of the variation was explained by systematic group-level effects. Considering the strong direct relationship between time investment and performance, this result is quite alarming, and it can in fact explain some of the systematic school- and class-level effects observed in test scores too. Thus, the systematic effects of schools and classes on test performance can partly tell about systematic differences in the assessment situation and not about the more general outcomes of education.

The strongest effects were found at the class-level (13.9 %). This corresponded with the class-level effects in test performance in Vantaa, but in time investment there was also a relatively strong school-level effect (7.3 %), which was not visible in the test scores. These two effects on time investment together makes one suspect that there were systematic differences between schools and classes in how the assessment situation was organised timewise even though there was a predefined time limit and detailed instructions for the assessment session. It is possible that placing the assessment session at the end of the school day, or right before a sports lesson for instance, will lead to more children trying to do the tasks as fast as possible with reduced effort. This should be studied more closely in a separate study as it is a serious issue which is compromising the reliability of all low-stakes assessment results. Only in very rare instances it is possible to allocate a whole school day for an external assessment, and the increased use of computer-based assessment sets additional constraints on the timing of the assessment sessions. Typically, school computer labs cannot accommodate all the participants simultaneously, and they are needed for other purposes as well, so scheduling the assessment sessions can be quite a challenge for schools. However, more accurate instructions are clearly needed to

diminish the possible and even probable systematical biases in the assessment results.

Not all the group-level variation was explained by schools and classes, but also within-class gender groups played a role (6 %) in how much time sixth graders spent on the assessment tasks. This can be interpreted as a peer group effect, especially when there seemed to be a gender difference in how classmates of the same gender influenced each others' task behaviour. Boys' time investment was more strongly dependent on their peers' time investment (17.5 % of the variance explained) compared to girls (13.4 %), so the additional four percent of explained variance was most likely telling about boys who were rushing through the last tasks when they understood their male peers were already finished. This interpretation, however, would require more detailed analyses of the time investment of the separate tasks and on the effects of task order which were not addressed in the present study. Anyhow, based on the present study it can be concluded that even if the children work on the assessment tasks individually, it seems that in a school computer lab it is very hard to prevent children from following each others' progress at least superficially and to be influenced by others' task behaviour. Boys who in general seem to be more vulnerable to the effects of situational factors and disturbances, may suffer from this more than girls do even though the use of computers may also enhance their task motivation more (cf. Halldórsson, McKelvie & Björnsson, 2009).

#### **8.4.4 Summary on the effects of time investment on performance**

The present study shows clearly that when the aim is to understand children's performance in educational assessment studies, time investment is an important factor which needs to be taken into account. The results show that just like with 15-year-olds in an earlier study (Kupiainen et al., 2014), also sixth graders' time investment is partly explained by their attitudes even though the mechanisms are somewhat different when children are younger, and that time investment is a strong predictor of performance in the test. In addition, the present study provides new information about how girls' higher performance in the assessment can be even entirely explained by time investment and

slightly stronger mastery attitudes which affect performance through time investment. Children in need of support – who are more often boys than girls – spend less time on assessment tasks, partly because of their higher levels of detrimental attitudes. This can partially explain the increase of the gap in the test scores of children with support needs compared to the others over time. As both detrimental attitudes and effort as measured by both time investment and self-reports are partly group-level phenomena, they would be excellent targets for school-based interventions.

This study also shows that log data of computer-based assessment may reveal systematic school- and class-level differences in how the assessment situation is conducted regardless of detailed instructions. In a school computer lab it may be more difficult to control the situation than in traditional paper-based assessment settings, and the use of computers changes also the assessment situation when it is usually not possible to accommodate more than about 20-25 children in the computer lab, making the scheduling of the assessment demanding. The present study shows that more research is clearly needed before the results of paper-based assessment and computer-based assessment can be fully compared (cf. Hautamäki et al., 2013). At the same time, the log data of computer-based assessment provides possibilities to understand children's task behaviour much better than earlier (cf. Greiff et al., 2013), and only a small amount of these possibilities have been so far utilised in Finnish large-scale assessments.

The next step from here could be to collect longitudinal computer-based data in order to evaluate how much individual children's task behaviour varies from time to time. Children's approach to novel tasks at different time points is not necessary very stable. Siegler and colleagues (see Siegler, 2005) have shown in multiple studies that children change between multiple problem-solving strategies even on two occasions close in time. Computer-based data collection would enable, besides analysing the effects of time investment, examining the strategy use in the assessment tasks (e.g. Greiff, Wüstenberg, Holt, Goldhammer & Funke, 2013), and this would be of great importance when designing interventions for enhancing thinking and problem-solving skills.

## 8.5 Limitations of the study

Conducting large-scale educational assessment studies requires a lot of effort from the research centre responsible for the implementation, the education departments of the municipalities and individual schools with hundreds of teachers and thousands of children. Factors compromising the validity of the results can never be eliminated entirely, and the present study makes no exception in this respect. In fact, one of the main aims of the present study was to address potential threats to the reliability and validity of large-scale assessments in general, so many of the issues presented above have already brought many limitations of the study into discussion. The most severe limitation is that when stakes are low, there is no way to force all the participants to do their best in the assessment situation. The results of the present study, however, can help in understanding how attitudes, interest and effort can influence performance and the development of it over time, and it may serve as a basis for school-level interventions for diminishing the problems related to it in the future.

There are other limitations of the study as well, which do not depend on the children or schools participating in the assessments. As the study utilised three different data sets, which were collected at different time points in two municipalities, the results are comparable across samples only to a certain extent. In this study, compromises had to be made to achieve a sufficient level of comparability and yet having measures, which were reliable enough to justify the conclusions. The greatest concern was that the first grade learning preparedness test had been modified at the item-level between the data collections in Helsinki in 2007 and in Vantaa in 2010 as the distributions of the scales were not normal in the full Helsinki data. This made the number of common items in both test versions quite low, which then influenced the reliability of the scales when only common items were selected for the analyses. To some extent, the same problem applied also to the third/fourth grade scales used in the first substudy as at the time of the Helsinki data collection the third/fourth grade test version was still under development. Due to these reasons, the results of Helsinki and Vantaa were really compared only in the first substudy and the corresponding parts of the last substudy, and the second and the third substudy utilised the best available data from

one municipality only. Therefore, all the substudies should be replicated with at least one new longitudinal data set, using a larger number of common items with the Vantaa study. The same phenomena – the influences of attitudes, task interest and time investment on the development of test performance, and the systematic school, class and peer effects on them – should also be studied with other measures than the Finnish learning to learn scales. This would be particularly important with regard to the development of gender differences, and to the increasing gap between pupils with support needs and others, as the results of the present study can be highly relevant for intervention planning. This, however, is true only if the results are not dependent on specific measures.

Another limitation is the lack of measures of pupils' language proficiency and skills. This would be particularly important when using young children's self-evaluations as it is very difficult to know whether the weak or non-existent connections of attitudes and performance depend more on the limited capacity for self-evaluation than on their limited understanding of the questionnaires. However, in another study using the 6-year follow-up data from Helsinki (Vainikainen et al., submitted) the attitude scales used in the present study were concluded as being measurement invariant across time, which supports the use of them already in the fourth grade and the interpretation that the problem lies in children's self-evaluation skills and not in the understanding of the language. Yet, also many of the cognitive tasks of the Finnish LTL scales require language skills, and in the future the initial differences in them should be better taken into account.

Pupils with support needs were especially focused on here. Yet, the study revealed particular limitations for the interpretation of their results from the turn of the third and fourth grade to the end of the sixth grade. The results showed that they were slowly falling behind compared to the others, and this relative decrease could only to a very limited extent be explained by the measures used in the present study. Thus, the study clearly did not manage to address some important – most likely psychological and socio-emotional – factors which could explain why the differences increased over time. Furthermore, the dichotomous categorisation of pupils into two groups (No support needs and Support



needs) may have masked some differences between pupils receiving intensified support and special support, and it would have been extremely useful if it had been possible to study the differences between pupils with different reasons for their support needs (for instance learning difficulties vs. behavioural problems). However, such data were not available in the present study. Nevertheless, the results showed that pupils with support needs had more detrimental attitudes and that they clearly did not invest a sufficient amount of time and effort in the sixth grade assessment, so their performance told more about what they were willing to show in the assessment than what they may have been able to do. Of course, this applies most likely to their everyday school work as well, and more detailed analyses of their time investment could have helped in planning interventions for enhancing their performance. This, however, was not possible in the present study as due to technical reasons, time on task was available only on a task-basis, not at an item-level.

Log file analyses of time on task revealed also further limitations, which most likely do not apply only to the present study but to all educational assessment studies, which are administered by teachers according to written instructions. Regardless of the pre-defined lengths of the assessment sessions, there were relatively strong systematic class effects on time investment, which indicate that there were systematic differences in how the assessment situation was conducted in practice. It is possible that placing the assessment session at the end of the school day, or right before lunch for instance, results in pupils' doing the tasks as fast as possible with reduced effort. The quite strong peer group effects, especially for boys, point in the same direction: it is possible that many boys were rushing through the last tasks when they understood their male peers were already finished. It seems that in a school computer lab it is very hard to prevent children from following each others' progress at least superficially and to be influenced by others' task behaviour regardless of strict written instructions to the teachers. This interpretation, however, would require more detailed analyses on time investment of the separate tasks and on the effects of task order, which were not addressed in the present study.

There were also limitations in how attitudes and task interest were measured. Measuring primary school children's attitudes is always difficult due to their limited capacity for self-evaluation. In fact, one of the findings of the present study was that also in the context of Finnish learning to learn assessments, attitudes begin to be gradually related to performance at the age of 10 to 12, just like previous literature from different contexts suggest. In the Finnish learning to learn assessment package there are many other attitude scales, which were not used here, and somewhat stronger (or even weaker) connections could have been found if for example academic self-concepts or self-efficacy had been included in the analyses. However, all the available measures could not be utilised in one single study, and the selection was therefore made on a theoretical basis. Since the results were in this respect not surprising considering the earlier literature, the selection of the scales was not questioned at any later point either.

Whereas for other attitudes there were numerous scales in the questionnaire to be selected from, for task interest the situation was just the opposite. Task-specific interest was asked with a single question in connection to each cognitive task, and to use it as a reliable measure the questions had to be regressed on a latent factor measuring the interest in the whole test. To study the effects of different task types or task characteristics on interest would have required more rigorous measures, and it was therefore not possible in the present study. It is also to be noted, that the tasks comprised of more items than the ones used in this study as there were also easier items in the fourth grade test version and more difficult items in the sixth grade test version. The questions referred always to the task as a whole, not specifically to the items used in this study, and the items excluded from the present analyses may have influenced pupils' evaluations of the interestingness of the tasks too.

Finally, the restrictions considering the samples, the missing data and the dropout rates need to be mentioned here. The samples were unusually large and representative, and the attrition rates were relatively low due to the fact that the data were collected as a part of normal school work. Therefore, the results of the study are generalisable at least within the capital area, in which both the municipalities are located. As on any school day, about 5 % of the pupils were absent at the time of the

assessments, but otherwise all the sampled pupils participated. Also their parents were active in returning their questionnaires as their response rates were over 80 % on almost all occasions. However, due to a large number of measures and the low stakes nature of the assessments, there were some item- or task-level missing data. In this study, the issue of missing data was solved by using maximum likelihood estimation when possible as it utilises the best available data without inputting any values. A different approach could have been chosen, but this was considered as sufficient in this case.

As the pupils of the three samples were followed for from three to six years, dropout needs to be discussed, too. Since the studies were conducted on assignment from the municipalities, according to their interests the pupils were followed only within the municipalities, not if they moved away from there. Therefore, in Helsinki, individual pupils were searched for and assessment materials delivered to them even if they had moved to schools which were located in Helsinki but that were not a part of the school sample. However, data were not obtained from all of these movers. In Vantaa, the follow-up was easier to organise as all the schools participated in the assessment on both occasions. Therefore, the dropout rates were lower and consisted only of pupils who had moved away or were absent from the second assessment due to illness.

When the sample size, the length of the follow-up period, and the number of measures increase, it necessarily increases the limitations of the study regarding its implementation – there are always things, which are not under control of any individual researcher. On the other hand, the advantages of a large-scale longitudinal approach – generalisability, statistical power and prediction instead of only correlation – are so obvious that limitations on other areas need to be accepted. The present study provides information about the development phenomena, which are usually considered as limitations of large-scale low stakes educational assessments. In the future, hopefully, the results can be utilised not only in understanding the development of pupils' performance and enhancing educational practices, but also in developing educational assessment methods further.

## 8.6 General conclusions and practical implications

The general aim of the present study was to understand the development of the performance in educational assessments, which usually is analysed only cross-sectionally. Based on numerous assessment studies, we know for instance that girls tend to perform better than boys (e.g. Hautamäki et al., 2013; OECD, 2013), that in Finland the between-school differences are small but between-class differences relatively large (Hautamäki et al., 2013; Yang Hansen, Gustafsson, & Rosén, 2014) and that pupils with support needs get lower scores, both in regard to their performance and their motivation (Thuneberg, 2007). However, there is little large-scale longitudinal research, which is comparable to the present study, which shows the following: that the later observed gender differences develop slowly over time during primary school and can to a large extent be explained by differences in attitudes and effort; that the large differences between classes are at least in a more typical Finnish municipality produced by the schools through a non-random assignment of pupils in classes already at the school start; or that pupils with support needs start lower but manage to keep up with the others for the first three years of basic education before beginning to slowly fall behind. Therefore, the most important practical implications of the study are related to the possibilities for evidence-informed early intervention and prevention.

The results showed that whereas girls were evaluated by their teachers as being slightly better readers already when they came to school, there was no gender difference in pupils' performance in the learning preparedness test. Girls, however, gained slightly more in reading comprehension during the first three years of basic education. Boys in Helsinki outperformed girls in mathematical thinking in the beginning of the third grade, but girls closed the gap by the end of the sixth grade. Also more generally, from the end of the third grade to the end of the sixth grade girls improved their performance slightly more than boys in both municipalities. All this implies that already from the first grades girls better adapt to the demands of compulsory education, and they manage to better utilise their potential later in their school career. This is seen also as boys' higher support needs even though there are no differences in girls' and boys' initial learning preparedness. The log data analyses of the sixth grade computer-based assessment revealed

that in the present study the girls' advantage could be completely explained by their more positive attitudes and greater effort as measured by their time investment in the tasks. The results also showed that changes in attitudes and task interest happened to some extent in classes and peer groups, and boys – who were also identified as having support needs more often than girls – seemed to be more vulnerable to the influences of their boy classmates both regarding their attitudes and task behaviour in the assessment situation. In general, learning-related attitudes declined by age, but this change was unrelated to the changes in performance. Changes in task interest, however, were a meaningful predictor of later performance.

These results emphasise the importance of creating a school and class atmosphere, which supports the positive development of effort and persistence of all learners. This is particularly important for boys: the development of peer group cultures, which have harmful influences on attitudes and the development of performance, should be addressed by means of early intervention and prevention already in primary school as their effects are based on this study visible already in the sixth grade. The same applies to pupils with support needs: based on this study the support system works very well during the three first years of basic education, but after that these pupils' detrimental attitudes begin to have a greater influence on their performance, and at least in the assessment situation they do not put as much effort into the tasks as the others do on average. Therefore, the effectiveness of the provided support should not be evaluated on performance-related measures only, but more attention should be paid to the positive development of effort and attitudes. Differentiating teaching is one of the most central means of support (Thuneberg et al., 2013), and now the next step should be the further differentiation of assessment, both in terms of assessing the effectiveness of the provided support but also more generally when the effectiveness of education is evaluated in assessment studies.

Between-school differences slightly increased during the six years of follow-up in Helsinki, but in Vantaa the pattern was much more typical for Finnish schools, having the variation between classes in schools (cf., Yang Hansen et al., 2014). This tells most likely about the emerging phenomenon of differentiation of the capital schools (cf., Bernelius,

2013), which was now shown also longitudinally by following the same pupils. However, the longitudinal analyses showed that even though performance seemed to develop in different ways in different schools, there were no school-level Matthew-effects visible. That is, some schools also started lower but managed to lift the performance relatively more than the others, whereas some schools did not quite meet their own expectations of especially high performance. In this respect, the development of performance was much more even in Vantaa, in which the systematic school effects remained more or less stable, being very small all the way. This implies that in a typical Finnish municipality with few differentiated or profiled schools, pupils are usually in a quite equal position regarding how their development and attitudes develop during the primary school years. Of course, this does not yet tell anything about how different the trajectories can become within the same schools based on class assignment, and the results of the present study indicate that the same phenomenon of differentiation, which in Helsinki is visible at the school-level, can be found within each school in Vantaa as a result of selection. Thus, more research is clearly needed on this area before making any firm conclusions.

The present study has brought into discussion many issues, which may partially explain both the decrease of Finnish pupils' assessment results and the increase of differences between pupils, peer groups, classes and schools. This work should now be continued by launching more systematically coordinated longitudinal assessment programmes with nationally representative samples. Only then can we really understand the role of education in enhancing pupils' competences and keep the quality high of the Finnish basic education also in the future.

## References

- Adey, P., Csapó, B., Demetriou, A., Hautamäki, J., & Shayer, M. (2007). Can we be intelligent about intelligence? Why education needs the concept of plastic general ability. *Educational Research Review*, 2, 75-97. doi:10.1016/j.edurev.2007.05.001
- Adler, P.A., Kless, S.J., & Adler, P. (1992). Socialization to gender roles: Popularity among elementary school boys and girls. *Sociology of Education*, 65 (3), 169-187.
- Aho, E., Pitkänen, K., & Sahlberg, P. (2006). Policy development and reform principles of basic and secondary education in Finland since 1968. Washington, D.C: The World Bank.
- Ahtiainen, R., Beirad, M., Hautamäki, J., Hilasvuori, T., Lintuvuori, M., Thuneberg, H., Vainikainen, M-P, & Österlund, I. (2011). *Tehostettua ja erityistä tukea tarvitsevien oppilaiden opetuksen kehittäminen 2007–2011: Kehittävän arvioinnin loppuraportti. [Developing the instruction of students in need of intensified and special support in 2007-2011. The final report of developing evaluation]*. Research reports of the Ministry of Education and Culture, 2012:5. Helsinki: Opetus- ja kulttuuriministeriö. 74 p.
- Ahtola, A. (2012). Proactive and Preventive Student Welfare Activities in Finnish Preschool and Elementary School: Handling of Transition to Formal Schooling and National Anti-Bullying Program as Examples. Turku: Finland. University of Turku.
- Ahtola, A., Silinskas, G., Poikonen, P.-L., Kontoniemi, M., Niemi, P., & Nurmi, J.-E. (2011). Transition to formal schooling: Do transition practices matter for academic performance? *Early Childhood Research Quarterly*, 26 (3), 295-302.
- Ahtola, A. & Vainikainen, M.-P. (in press). Koulun ja oppilaitoksen psykologin työ sisäisen palvelun ja verkostotyön konteksteissa: Raportti työnkuvakyselystä. [The work of school psychologists in the contexts of internal services and networks: A report of a questionnaire about job descriptions of psychologists]. *Psykologia*.
- Ainley, M., Hidi, S., & Berndorff, D. (2002a). Interest, learning, and the psychological processes that mediate their relationship. *Journal of Educational Psychology*, 94(3), 545-561. doi:10.1037//0022-0663.94.3.545

- Ainley, M., Hillman, K., & Hidi, S. (2002b). Gender and interest processes in response to literary texts: Situational and individual interest. *Learning and Instruction, 12*, 411-428.
- Aunio, P., Hautamäki, J., Sajaniemi, N., & Van Luit, J.E.H. (2009). Early numeracy in low-performing young children. *British Educational Research Journal, 35*(1), 25-46. DOI:10.1080/01411920802041822
- Aunola, K., Leskinen, E., & Nurmi, J.-E. (2006). Developmental dynamics between mathematical performance, task motivation, and teachers' goals during the transition to primary school. *British Journal of Educational Psychology, 76*, 21-40.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* Vol. 8, 47-89. New York: Academic Press.
- Bast, K., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex model with structured means. *Multivariate Behavioural Research, 32*(2), 135-167.
- Berndt, T. J., & Keefe, K. (1995). Friends' influence on adolescents' adjustment to school. *Child Development, 66*, 1312-1329.
- Bernelius, V. (2013). *Eriytyvät kaupunkikoulut - Helsingin peruskoulujen oppilasohjan erot, perheiden kouluvalinnat ja oppimistuloksiin liittyvät aluevaikutukset osana kaupungin eriytymiskehitystä*. City of Helsinki Urban Facts, Research reports 2013:1. <http://urn.fi/URN:ISBN:978-952-272-530-1>
- Bond, T.G. (1995). *BLOT: Bond's Logical Operations Test*. Townsville, Australia: James Cook University. (Original work published 1976).
- Bond, T.G., & Fox, C.M. (2012). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Second edition. New York/East Sussex: Routledge.
- Brown, B.B., & Lohr, M.J. (1987). Peer group affiliation and adolescent self-esteem: An integration of ego-identity and symbolic interaction theories. *Journal of Personality and Social Psychology, 52*, 47-55.
- Burns, M.K. & Ysseldyke, J.E. (2005). Comparison of existing Response-to-Intervention models to identify and answer implementation questions. *The California School Psychologist, 10*, 9-20.
- Byrne, B.M., & Stewart, S.M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*(2), 287-321.
- Caro, D.H., McDonald, J.T., & Willms, J.D. (2009). Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education, 32*(3), 558-590.



- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Cheung, G. W. & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables. Bootstrapping with structural equation models. *Organisational Research Methods*, 11(2), 296-325. doi: 10.1177/1094428107300343
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227-268. doi: 10.3102/00346543066003227
- Crick, N.R. & Dodge, K.A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115(1), 74-101.
- Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioral Development*, 20 (4), 609-626.
- Csapó, B. (2007). Research into learning to learn through the assessment of quality and organisation of learning outcomes. *The Curriculum Journal*, 18(2), 195-210.
- Csapó, B. & Molnár, Gy., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, 106 (3), 639-650. doi: 10.1037/a0035756
- Deakin Crick, R. (2007). Learning how to learn: The dynamic assessment of learning power. *The Curriculum Journal*, 18(2), 135-153.
- Demetriou, A., & Kazi, S. (2006). Self-awareness in *g* (with processing efficiency and reasoning). *Intelligence*, 34, 297-317.
- Demetriou, A., Pachaury, A., Metallidou, Y., & Kazi, S. (1996). Universals and specificities in the structure and development of quantitative-relational thought: A cross-cultural study in Greece and India. *International Journal of Behavioural Development*, 19 (2), 255-290. doi: 10.1080/016502596385785
- Demetriou, A., Platsidou, M., Efklides, A., Metallidou, Y., & Shayer, M. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction*, 1, 19-43.
- Demetriou, A., Spanoudis, G., & Mouyi, A. (2011). Educating the developing mind: Towards an overarching paradigm. *Educational Psychology Review*, 23(4), 601-663. doi:10.1007/s10648-011-9178-3

- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental psychology, 43*(6), 1428-1446. doi:10.1037/0012-1649.43.6.1428
- Dweck, C., Chiu, C., & Hong, Y. (1995). Implicit theories: Elaboration and extension of the model. *Psychological Inquiry, 6*(4), 322-333.
- Eccles, J. (2011). Gendered educational and occupational choices: Applying the Eccles et al. model of achievement-related choices. *International Journal of Behavioural Development, 35*(3), 195-201.
- Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology 53*(1), 109-132.
- Gagné, F., & St. Père, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*, 71-100.
- Gazzaniga, M.S., Ivry, R.B., & Mangun, G.R. (1998). *Cognitive neuroscience: The biology of the mind*. New York: Norton.
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106* (3), 608-626.
- Graham, L.J. & Jahnukainen, M. (2011). Where art thou, inclusion? Analysing the development of inclusive education in New South Wales, Alberta and Finland. *Journal of Education Policy, 26*(2), 263-288. DOI: 10.1080/02680939.2010.493230
- Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts – something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364-379. doi:10.1037/a0031856
- Greiff, S., Wüstenberg, S., Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189-213.
- Greiff, S., Wüstenberg, S., Holt, D., Goldhammer, F. & Funke, J. (2013). Computer-based assessment of complex problem solving: Concept, implementation, and application. *Education Tech Research Dev*. DOI 10.1007/s11423-013-9301-x.

- Gustafsson, J. -E. & Carlstedt, B. (2006). Abilities and grades as predictors of achievement: The Encapsulation theory. Paper presented at the annual meeting of the American Psychological Association, New Orleans, August 10–13.
- Halldórsson, A.M., McKelvie, P., & Björnsson, J.K. (2009). Are Icelandic boys really better on computerised tests? In T. Matti (Ed.). *Northern Lights on PISA 2006. Differences and similarities in the Nordic Countries*. TemaNord 2009:547. Copenhagen.
- Halpern, D. F. (2000). *Sex Differences in Cognitive Abilities. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hamm, J.V., Hoffman, A., & Farmer, T.W. (2012). Peer cultures of academic effort and achievement in adolescence: Why they matter, and what teachers can do about them. In A.M Ryan & G.W. Ladd (eds.) *Peer Relationships and Adjustment at School*, pp. 219-250. Charlotte, NC: Information Age Publishing, Inc.
- Hanushek, E.A., Kain, J.F., Markman, J.M., & Rivkin, S.G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18, 527-544.
- Hanushek, E.A., Kain, J.F., & Rivkin, S.G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *The Review of Economics and Statistics*, 84(4), 584-599.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: necessary and illuminating. *Journal of Educational Psychology*, 94 (3), 638–645. doi:10.1037//0022-0663.94.3.638
- Harris, D.N. (2010). How do school peers influence student educational outcomes? Theory and evidence from economics and other social sciences. *Teachers College Record* 112(4), 1163-1197.
- Harter, S. (1999). *The construction of self. A developmental perspective*. New York: The Guilford Press.
- Hautamäki A., Hautamäki J., & Kupiainen S. (2010) Assessment in Schools – Learning to Learn. *International Encyclopedia of Education* 2010(3), 268–272.
- Hautamäki, J. (1984). *Peruskoululaisten loogisen ajattelun mittaamisesta ja esiintymisestä. [Measuring and the occurrence of logical thinking among basic school students]*. Joensuun yliopiston yhteiskuntatieteellisiä julkaisuja 1. Joensuu: Joensuun yliopistopaino.

- Hautamäki, J., Arinen, P., Hautamäki, A., Lehto, J., Lindblom, B., Kupiainen, S., Outinen, K., Pekuri, M., Reuhkala, M., Scheinin, P. (2001). *Ensiaskleet – oppimisen edellytykset: Luokanopettajille tarkoitettu seulan toimivuus Helsinki-aineiston perusteella. [First steps – Prerequisites of learning: The functionality of the screening test for class teachers based on Helsinki-data]*. City of Helsinki, Publication series A17: 2001.
- Hautamäki, J., Arinen, P., Bergholm, B., Hautamäki, A., Kupiainen, S., Kuusela, J., Lehto, J., Niemivirta, M. & Scheinin, P. (1999). *Oppimaan oppiminen ala-asteilla. [Learning to learn in primary schools]*. National Board of Education, Evaluation 3/1999.
- Hautamäki, J., Arinen, P., Hautamäki, A., Ikonen-Varila, M., Kupiainen, S., Lindblom, B., Niemivirta, M., Rantanen, P., Ruuth, M. & Scheinin, P. (2000). *Oppimaan oppiminen yläasteilla. [Learning to learn in lower secondary schools]*. National Board of Education 7/2000.
- Hautamäki, J., Arinen, P., Hautamäki, A., Kupiainen, S., Lindblom, B., Mehtäläinen, J., Niemivirta, M., Rantanen, P., Ruuth, M. & Scheinin, P. (2003) *Oppimaan oppiminen yläasteella 2. Tilanne vuonna 2001 ja muutokset vuodesta 1997. [Learning to learn in lower secondary schools. The situation in 2001 and the changes since 1997]*. National Board of Education, Evaluation 6/2003.
- Hautamäki, J., Arinen, P., Hautamäki, A., Kupiainen, S., Lindblom, B., Niemivirta, M., Rantanen, P. & Scheinin, P. (2002b) *Oppimaan oppiminen toisen asteen koulutuksessa. [Learning to learn in upper secondary schools]*. National Board of Education, Evaluation 2/2002.
- Hautamäki, J. & Kupiainen, S. (2014). Learning to learn in Finland. Theory and policy, research and practice. In Ruth Deakin Crick, Cristina Stringher & Kai Ren (Eds.) *Learning to Learn. International perspectives from theory and practice*. Routledge.
- Hautamäki, J., Kupiainen, S., Arinen, P., Hautamäki, A., Niemivirta, M., Rantanen, P., Ruut, M., Scheinin, P. (2005). *Oppimaan oppiminen ala-asteella 2. Tilanne vuonna 2003 ja muutokset vuodesta 1996. [Learning to learn in primary schools. The situation in 2003 and the changes since 1996]*. National Board of Education, Evaluation 6/2005.
- Hautamäki, J., Kupiainen, S., Arinen, P., Hautamäki, A., Niemivirta, M., Rantanen, P., Scheinin, P. (2006). Learning-to-learn assessment in Finland – versatile tools to monitor and improve effectiveness and equity of the education system. *Research-based teacher education in Finland – reflections by Finnish teacher educators*. Kasvatusalan tutkimuksia; 25, pp. 189-202. Turku: Finnish Education Research Association.

- Hautamäki, J., Arinen, P., Eronen, S., Hautamäki, A., Kupiainen, S., Lindblom, B., Niemivirta, M., Pakaslahti, L., Rantanen, P., Scheinin, P. (2002a). *Assessing Learning-to-learn: A Framework*. Helsinki: National Board of Education, Evaluation 4/2002.
- Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M.-P., & Hotulainen, R. (2013). *Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne vuonna 2012 ja muutos vuodesta 2001 [Learning to learn at the end of basic education: The situation in 2012 and the change from 2001]*. University of Helsinki, Department of Teacher Education, Research Reports 347. Helsinki: Unigrafia.
- Henning-Stout, M., & Close-Conoley, J. (1992). Gender: A subtle influence in the culture of the school. In F.J. Medway & T.P. Cafferty (eds.), *School psychology: A social psychological perspective* (pp.113-135). Hillsdale, Nj: Lawrence Erlbaum Associates.
- Hidi, S., & Renninger, K.A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111-127.
- Hirvonen, K. (2012). *Onko laskutaito laskussa? Matematiikan oppimistulokset peruskoulun päättövaiheessa 2011 [Are pupils' counting skills declining? Ninth graders' performance in mathematics in 2011]*. Koulutuksen seurantaraportit 2012:4. Helsinki: National Board of Education.
- Hosenfeld, B., van den Boom, D.C., & Resing, W.C.M. (1997). Constructing geometric analogies test for the longitudinal testing of elementary school children. *Journal of Educational Measurement, 34*(4), 367-372.
- Hoskins, B., & Fredriksson, U. (2008). *Learning to Learn: What is it and can it be measured?*. JRC Scientific and Technical Report. European Commission, Centre for Research on Lifelong Learning.
- Huesmann R. L., Dubow E.F. & Boxer, P. (2009). Continuity of aggression from childhood to early adulthood as a predictor of life outcomes: Implications for the adolescent-limited and life-course-persistent models. *Aggressive Behaviour, 35*, 136-149.
- Hullemann, C.S., Durik, A.M., Schweigert, S.A., & Harackiewicz, J.M. (2008). Task values, achievement goals, and interest: An integrative analysis. *Journal of Educational Psychology, 100*(2), 398-416. doi:10.1037/0022-0663.100.2.398
- Häyrynen, Y.-P., & Hautamäki, J. (1977). *Människans bildbarhet och utbildningspolitiken*. Stockholm: Wahlström & Widstrand
- Inhelder, B. & Piaget, J. (1958). *The early growth of logic in the child*. London: Routledge & Kegan Paul.

- Jahnukainen, M. (2011). Different Strategies, Different Outcomes? The History and Trends of the Inclusive and Special Education in Alberta (Canada) and in Finland. *Scandinavian Journal of Educational Research* 55, 489–502.
- Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, 74(4), 443-471. doi:10.3102/00346543074004443.
- Kenney-Benson, G.A., Pomerantz, E.M., Ryan, A.M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology*, 42(1), 11-26. doi:10.1037/0012-1649.42.1.11
- Kimball, M.M. (1989). A new perspective on women's math achievement. *Psychological Bulletin*, 105, 198-214.
- Kindermann, T.A. (2007). Effects of naturally existing peer groups on changes in academic engagement in a cohort of sixth graders. *Child Development*, 78(4), 1186-1203.
- Kintsch, W. & van Dijk, T.A. (1978). Towards a model of text comprehension and production. *Psychological Review*, 8 (5), 363-394.
- Kirsch, I., de Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). *Reading for change. Performance and engagement across countries. Results from PISA2000*. Paris:OECD.
- Klauer, K.J. (1988). Teaching for learning-to-learn: A critical appraisal with some proposals. *Instructional Science*, 17, 351-367.
- Kline, R.B. (2005). *Principles and Practice of Structural Equation Modeling*. Second edition. New York: The Guildford Press.
- Knotek, S.E. (2005). Sustaining RTI through consultee-centered consultation. *The California School Psychologist*, 10, 93–104.
- Kosunen, S (2014). Reputation and parental logics of action in local school choice space in Finland. *Journal of Education Policy*, 29 (4), 443-466. doi:10.1080/02680939.2013.844859
- Kupari, P., Välijärvi, J., Andersson, L, Arffman, I., Nissinen, K., Puhakka, E., & Vettenranta, J. (2013). *PISA 12 Ensituloksia [PISA 2012: The first results]*. Opetus- ja kulttuuriministeriön julkaisuja 2013:20.
- Kupiainen, S., Hautamäki, J., & Rantanen, P. (2008). EU pre-pilot on learning to learn: Report on the compiled data. 2008-1190/001-001 TRA-TRINDC.

- Kupiainen, S., Marjanen, J., Vainikainen, M.P., & Hautamäki, J. (2011). *Oppimaan oppiminen Vantaan peruskouluissa: Kolmas-, kuudes- ja yhdeksäsluokkalaiset oppijoina keväällä 2010. [Learning to learn in comprehensive schools in Vantaa: Third, sixth and ninth graders as learners in spring 2010]*. City of Vantaa.
- Kupiainen, S., Vainikainen, M.-P., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in a low stakes assessment of cross-curricular skills. *Journal of Educational Psychology, 106* (3), 627-638. doi 10.1037/a0035507
- Kuusela, J. (2010). Oppilaiden sosioekonomisen taustan yhteys koulumenestykseen koulutasolla [School-level relationships of socio-economic background of the pupils and their school achievement]. In M. Rimpelä & V. Bernelius (Eds.). *Peruskoulujen oppimistulokset ja oppilaiden hyvinvointi eriytyvällä Helsingin seudulla [Learning results and welfare of basic education pupils in the differentiating Helsinki area]*. University of Helsinki: Publications of the Department of Geosciences and Geography B1. Helsinki: Yliopistopaino.
- Kuusela, J. (2000). *Tieteellisen paradigman mukaisen ajattelun kehittyminen peruskoulussa [The development of scientific thinking in basic education]*. University of Helsinki, Department of Teacher Education, Research Reports.
- Krkovic, K., Greiff, S., Kupiainen, S., Vainikainen, M.P., & Hautamäki, J. (2014). Gender bias in education: Does the interaction of student and teacher gender cause a biased evaluation of students' abilities? *Educational Research, 56* (2), 244-257. doi:10.1080/00131881.2014.898909
- Kärnä, P., Hakonen, R., & Kuusela, J. (2012). *Luonnontieteellinen osaaminen perusopetuksen 9. luokalla 2011 [Ninth graders' performance in science in 2011]*. Koulutuksen seurantaraportit 2012:2. Helsinki: National Board of Education.
- Ladd, G.W., Kochenderfer-Ladd, B., Visconti, K.J., & Ettekal, I. (2012). Classroom peer relations and children's social and scholastic development: Risk factors and resources. In A.M Ryan & G.W. Ladd (eds.) *Peer Relationships and Adjustment at School*, pp. 11-50. Charlotte, NC: Information Age Publishing, Inc.
- Lappalainen, H.-P. (2011). *Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010 [You will find it ahead. Ninth graders' performance in mother tongue and literature in 2010]*. Koulutuksen seurantaraportit 2011:2. Helsinki: National Board of Education.
- Leadbeater, B.J., Kuperminc, G.P., Blatt, S.J., & Hertzog, C. (1999). A multivariate model of gender differences in adolescents' internalizing and externalizing problems. *Developmental Psychology, 36*(5), 1268-1282.

- Lehto, J.E., Scheinin, P., Kupiainen, S., & Hautamäki, J. (2001). National survey of reading comprehension in Finland. *Journal of Research in Reading, 24* (1), 99-110.
- Lepola, J. (2004). The role of gender and reading competence in the development of motivational orientations from kindergarten to grade 1. *Early Education & Development, 15*(2), 215-240.  
doi:10.1207/s15566935eed1502\_5
- Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J.-E. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. *Learning and Instruction, 18*, 548-564. doi:10.1016/j.learninstruc.2007.11.004
- Lintuvuori, M. (2010). *Erityisopetus muutoksen kynnyksellä. Tilastollinen kuvaus erityisopetusjärjestelmästä ja sen määrällisestä kehityksestä 1970-luvun lopulta vuoteen 2008. [Finnish special education in change – A statistical description of the Finnish special education system from late 1970s to present]* Master's Thesis. Department of Teacher Education. University of Helsinki. <http://urn.fi/URN:NBN:fi-fe201006162060>
- Lintuvuori (in press). Oppimisen ja koulunkäynnin tuen järjestäminen virallisen tilastotiedon ja empiirisen tutkimusaineiston kuvaamina. In M. Jahnukainen, E. Kontu, H. Thuneberg & M.-P. Vainikainen (Eds.) *Erityisopetuksesta kolmiportaiseen tukeen*. The Finnish Educational Research Association.
- Little, T.D., Lopez, D.F., Oettingen, G. & Baltes, P.B. (2001). A comparative-longitudinal study of action-control beliefs and school performance. *International Journal of Behavioural Development, 25* (3), 237-245.
- Logie, R.H., & Pearson, D.G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology, 9*, 241-257.
- Lyytinen, S., & Lehto, J. E. (1998). Hierarchy rating as a measure of text macroprocessing: Relationship with working memory and school achievement. *Educational Psychology, 18* (2), 157-169.
- Lönnqvist, J.-E., Vainikainen, M.-P., & Verkasalo, M. (2012). Teacher and parent ratings of seven-year-old children's personality and psychometrically assessed cognitive ability. *European Journal of Personality, 26*, 504-514.
- Lönnqvist, J.-E., Verkasalo, M., Vainikainen, M.-P. (2011). Parent-teacher agreement on 7-year-old children's personality. *European Journal of Personality, 25*, 306–316.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1* (4), 173–181.



- Mangels, J.A., Butterfield, B., Lamb, J., Good, C., & Dweck, C.S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social cognitive and affective neuroscience*, 1(2), 75-86. doi:10.1093/scan/ns1013
- Marjanen, J., Vainikainen, M.P., Kupiainen, S., Hotulainen, R., & Hautamäki, J. (2014). *Oppimaan oppiminen Vantaan peruskouluissa: Kolmas-, kuudes- ja yhdeksäsluokkalaiset oppijoina vuosina 2013 ja 2010. [Learning to learn in comprehensive schools in Vantaa: Third, sixth and ninth graders as learners in 2013 and 2010]*. City of Vantaa.
- Matsunaga, M. (2008). Item parceling in Structural Equation Modeling: A primer. *Communication Methods and Measures*, 2 (4), 260-293.
- Midgley, M. & Urdan, T. (2001). Academic self-handicapping and achievement goals: A Further examination. *Contemporary Educational Psychology* 26, 61-75. doi:10.1006/ceps.2000.1041,
- Ministry of Education (2012). *Koulutuksen arviointisuunnitelma vuosille 2012-2015 [The plan for educational assessment for years 2012 to 2015]*. Opetus- ja kulttuuriministeriön julkaisuja 2012:14.
- Ministry of Education. (2007). *Erityisopetuksen strategia. [Special education strategy.]* Reports of the Ministry of Education 2007:47, Finland.
- Molnár, Gy., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9(8), 35-45.
- Moreno, A., & Martín, E. (2007). The development of learning to learn in Spain. *The Curriculum Journal*, 18(2), 175-193.
- Mähönen, T.A., Jasinskaja-Lahti, I., & Liebkind, K. (2011). The impact of perceived social norms, gender and intergroup anxiety on the relationships between intergroup contact and ethnic attitudes of adolescents. *Journal of Applied Social Psychology*, 41(8), 1877-1899. DOI: 10.1111/j.1559-1816.2011.00793.x
- National Board of Education (1999). *A Framework for Evaluating Educational Outcomes in Finland*. National Board of Education, Evaluation 8/1999.
- National Board of Education (2004). *National Core Curriculum for Basic Education 2004*. [http://www.oph.fi/english/publications/2009/national\\_core\\_curricula\\_for\\_basic\\_education](http://www.oph.fi/english/publications/2009/national_core_curricula_for_basic_education)
- National Board of Education (2011). *Opetussuunnitelman perusteiden muutokset ja täydennykset 2010 [Changes and additions to the National Core Curriculum for Basic Education 2010]*. Opetushallitus: Määräykset ja ohjeet 2011:20. [http://www.oph.fi/download/132882\\_Perusopetuksen\\_opetussuunnitelman\\_perusteiden\\_muutokset\\_ja\\_taydennykset2010.pdf](http://www.oph.fi/download/132882_Perusopetuksen_opetussuunnitelman_perusteiden_muutokset_ja_taydennykset2010.pdf)

- Niemivirta, M. (2002). Individual differences and developmental trends in motivation: Integrating person-centred and variable-centred methods. In P.R. Pintrich and M.L. Maehr (Eds.) *New directions in measures and methods. Advances in motivation and achievement*, Vol. 12, pp. 241-275. Amsterdam: JAI Press.
- OECD (2004). *Learning for Tomorrow's World. First Results from PISA 2003*. Paris: OECD.
- OECD (2007). *PISA 2006. Science Competencies for Tomorrow's World: Volume 1: Analysis*. Paris: OECD. DOI:10.1787/9789264040014-en
- OECD (2010). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*. OECD Publishing. DOI:10.1787/9789264091450-en
- OECD (2013a). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science (Volume I)*. OECD Publishing. <http://dx.doi.org/10.1787/9789264201118-en>
- OECD (2013b). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. <http://dx.doi.org/10.1787/9789264190511-en>
- OECD (2013c). *PISA 2015: Draft collaborative problem-solving framework*. Retrieved from <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>
- OECD (2013d). *Education at a Glance 2013. OECD indicators*. OECD Publishing. <http://dx.doi.org/10.1787/eag-2013-en>
- Olweus, D. (1979). Stability of aggressive reaction patterns in males: A review. *Psychological Bulletin*, 86(4), 852-875.
- Panula, A.-M. (2013). *Lukemisvaikeudet ja osa-aikainen erityisopetus. Seurantatutkimus esikoulusta yhdeksännen luokan loppuun [Reading difficulties and part time special education – Follow up from preschool to the end of 9<sup>th</sup> grade]*. University of Helsinki, Department of Teacher Education, Research Report 344.
- Patrick, H., Ryan, A.M., & Pintrich, P.R. (1999). The differential impact of extrinsic and mastery goal orientations on males' and females' self-regulated learning. *Learning and Individual Differences*, 11(2), 153-171.
- Piaget, J., & Inhelder, B. (1956). *The Child's Conception of Space*. (Translated from French by F.J. Langdon and J. L. Lunzer). London: Routledge & Kegan Paul.
- Petrill, S.A. (1997). Molarity versus modularity of cognitive functioning? A behavioural genetic perspective. *Current Directions in Psychological Science*, 6, 96-99.

- Powers, K., Hagans, K. & Busse, R.T. (2008). School psychologists as instructional consultants in a Response-to-Intervention model. *The California School Psychologist*, 13, 41–53.
- Райгородский, Д.Я. (ed.) (2008). *Энциклопедия психодиагностики. Психодиагностика детей*, 177-179.
- Rautopuro, J. (Ed.) (2013). *Huödyllinen pakkolasku. Matematiikan oppimistulokset peruskoulun päättövaiheessa 2012 [Ninth graders' performance in mathematics in 2012]*. Koulutuksen seurantaraportit 2013:3. Helsinki: National Board of Education.
- Recommendation 2006/962/EC of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. Official Journal L 394 of 30.12.2006.
- Renninger, K.A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist*, 46(3), 168-184. doi:10.1080/00461520.2011.587723
- Rose, A.J., & Rudolph, K.D. (2006). A review of sex differences in peer relationship processes: Potential trade-offs for the emotional and behavioural development of girls and boys. *Psychological Bulletin*, 132(1), 98-131. doi:10.1037/0033-2909.132.1.98
- Ross, J.D. & Ross, C.M. (1979). *Ross test of Higher Cognitive Processes*. Novato, California: Academic Therapy Publications.
- Ryan, A.M. (2012). Peer relationships and adjustments at school: An introduction. In A.M Ryan & G.W. Ladd (eds.) *Peer Relationships and Adjustment at School*, pp. 1-10. Charlotte, NC: Information Age Publishing, Inc.
- Ryan, A. M. (2000). Peer groups as a context for the socialization of adolescents' motivation, engagement, and achievement in school. *Educational Psychologist*, 35(2), 101-111.
- Ryan, A.M., Jamison, R.S., Shin, H., Thompson, G.N. (2012). Social achievement goals and adjustment at school during early adolescence. In A.M Ryan & G.W. Ladd (eds.) *Peer Relationships and Adjustment at School*, pp. 165-186. Charlotte, NC: Information Age Publishing, Inc.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54–67. doi:10.1006/ceps.1999.1020
- Sabel, C., Saxenian, A, Miettinen, R., Kristensen, P.H., & Hautamäki, J. (2011). *Individualized Service Provision in the New Welfare State: Lessons from Special Education in Finland*. Helsinki: Sitra Studies 62.

- Salo, K. (2010). *Hallinnontasojen välinen dialogi ja erityisopetuksen muutos byrokraattisena instituutiona- Kymppikuntien loppuraportin ja erityisopetuksen strategian sisällönanalyttinen vertailu [Dialogue between governmental administration levels and the change of a special education system as a bureaucratic institution- A comparative content analysis of the ten largest lea's final report and special education strategy.]* Department of Teacher Education. University of Helsinki. <http://urn.fi/URN:NBN:fi-fe201012073066>
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4-27.
- Shayer, M. (1979). Has Piaget's construct of formal operational thinking any utility? *British Journal of Educational Psychology*, *49*, 265-276. doi: 10.1111/j.2044-8279.1979.tb02425.x
- Shaywitz, B.A., Holford, T.R., Holahan, J.M., Fletcher, J.M., Stuebing, K.K., Francis, D.J., & Shaywitz, S.E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly*, *30*(4), 894-906.
- Sheridan, S. & Gutkin, T.B. (2000). The ecology of school psychology: Examining and changing our paradigm for the 21st century. *School Psychology Review*, *29*(4), 485-502.
- Siegler, R.S. (2005). Children's learning. *American Psychologist*, *60*, 769-778.
- Snow, R.E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law*, *2*(3/4), 536-560.
- Song, H.-D., & Grabowski, B. (2006). Stimulating intrinsic motivation for problem solving using goal-oriented context and peer group composition. *Educational Technology Research and Development*, *54* (5), 445-466.
- Steenkamp, J.-B.E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-107.
- Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences*, *19*, 80-90. doi:10.1016/j.lindif.2008.05.004
- Sternberg, R., Castejon, J.L., Prieto, M.D., Hautamäki, J., & Grigorenko, E. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples. *European Journal of Psychological Assessment*, *17*, 1-16.
- Tapola, A., Veermans, M., & Niemivirta, M. (2013). Predictors and outcomes of situational interest during a science learning task. *Instructional Science*, *41* (6), 1047-1064.

- Terry, R., & Coie, J.D. (1991). A comparison of methods for defining sociometric status among children. *Developmental Psychology*, 27(5), 867-880.
- The Association of Educational Assessment – Europe (2012). *European Framework of Standards for Educational Assessment 1.0*.  
[http://www.aea-europe.net/images/downloads/SW\\_Framework\\_of\\_European\\_Standards.pdf](http://www.aea-europe.net/images/downloads/SW_Framework_of_European_Standards.pdf).
- Thuneberg, H. (2007). *Is a majority enough? Psychological well-being and its relation to academic and prosocial motivation, self-regulation and achievement at school*. University of Helsinki, Department of Teacher Education, Research Report 281. Helsinki: Yliopistopaino.
- Thuneberg, H., Hautamäki, J., Ahtiainen, R., Lintuvuori, M, Vainikainen, M.-P., & Hiltavuori, T. (2014). Conceptual change in adopting the nationwide special education strategy in Finland. *Journal of Educational Change*, 15(1), 37-56.
- Thuneberg, H., Vainikainen, M.-P., Ahtiainen, R., Lintuvuori, M., Salo, K, & Hautamäki, J. (2013). Education is special for all – The Finnish support model. *Gemeinsam leben*, 21(2), 67-78.
- Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2012). Achievement goal orientations and academic well-being across the transition to upper secondary education. *Learning and Individual Differences*, 22 (3), 290-305.
- Urdu, T., & Midgley, M. (2001). Academic self-handicapping: What we know and what more there is to learn. *Educational Psychology Review*, 13 (2), 115-138.
- Vainikainen, M.P., Marjanen, J., Kupiainen, S., Gustavson, N., & Hautamäki, J. (2011). *Oppimaan oppiminen Vantaan peruskouluissa: Ensiluokkalaisten oppimisvalmiudet syksyllä 2010. [Learning to learn in comprehensive schools in Vantaa: Learning preparedness of the first graders in autumn 2010.]* City of Vantaa.
- Vainikainen, M.P., Thuneberg, H., Greiff, S., & Hautamäki, J. (submitted). Multiprofessional collaboration in Finnish schools.
- Vainikainen, Thuneberg, & Mäkelä, (in press). Moniammatillinen yhteistyö lähikouluperiaatteen toteuttamisen tukena. In M. Jahnukainen, E. Kontu, H. Thuneberg & M.-P. Vainikainen (Eds.) *Erityisopetuksesta kolmiportaiseen tukeen*. The Finnish Educational Research Association.
- Vainikainen, M.P., Wüstenberg, S., Kupiainen, S., Hotulainen, R., & Hautamäki, J. (submitted). Development of learning to learn in primary school. *International Journal of Lifelong Education*.

- Van Yperen, N. (2003). Task interest and actual performance: The moderating effects of assigned and adopted purpose goals. *Journal of Personality and Social Psychology, 85*(6), 1006-1015. doi:10.1037/0022-3514.85.6.1006
- Varjo, J., & Kalalahti, M. (2011). Koulumarkkinoiden institutionaalisen tilan rakentuminen. *Yhdyskuntasuunnittelu, 49*(4), 8-25.
- Varjo, J., Kalalahti, M., & Silvennoinen, H. (2014). Families, school choice, and democratic iterations on the right to education and freedom of education in Finnish municipalities. *Journal of School Choice: International Research and Reform, 8*(1), 20-48. doi:10.1080/15582159.2014.875408
- Wang, M.-T., & Eccles, J.S. (2013). School context, achievement motivation, and academic engagement: A longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12-23.
- Wang, M.-T., & Eccles, J.S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development, 83*(3), 877-895.
- Wechsler, D. (1981). *WAIS-R: Manual: Wechsler Adult Intelligence Scale – revised*. Harcourt Brace Jovanovich for Psychological Corp.
- Wentzel, K.R. (1988). Gender differences in math and English achievement. *Sex Roles, 18*, 11-12.
- Wentzel, K.R. & Caldwell, K. (1997). Friendships, peer acceptance, and group membership: Relations to academic achievement in middle school. *Child Development, 68*(6), 1198-1209.
- Wentzel, K.R., Donlan, A. & Morrison, D. (2012). Peer relationships and social motivational processes. In A.M Ryan & G.W. Ladd (eds.) *Peer Relationships and Adjustment at School*, pp. 79-107. Charlotte, NC: Information Age Publishing, Inc.
- Wentzel, K., Filisetti, L., & Looney, L. (2007). Adolescent prosocial behaviour: The role of self-processes and contextual cues. *Child Development, 78*(3), 895-910.
- Willms, J.D. (2010). School composition and contextual effects on student outcomes. *Teachers College Record, 112*(4), 1007-1037.
- Wilson, J.T.L., Scott, J.H., & Power, K.G. (1987). Developmental differences in the span of visual memory for pattern. *British Journal of Developmental Psychology, 5*, 249-255.
- Wise, S.L. (2006). An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education, 19*(2), 95-114.

- Wise, S.L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163-183.
- Yang Hansen, K., Gustafsson, J.-E. , & Rosén, M. (2014). School performance difference and policy variations in Finland, Norway and Sweden. In K. Yang Hansen, J.-E. Gustafsson, M. Rosén, S. Sulkunen, K. Nissinen, P. Kupari, R.F. Ólafsson, J.K. Björnsson, L.S. Grønmo, L. Rønberg, J. Mejding, I.C. Borge, & A. Hole. *Northern Lights on TIMMS and PIRLS 2011*. TemaNord 2014:528, 25-48.
- Yang Hansen, K., Rosén, M., & Gustafsson, J.-E. (2011). Changes in the multi-level effects of socio-economic status on reading achievement in Sweden in 1991 and 2001. *Scandinavian Journal of Educational Research*, 55(2), 197-211.
- Zhao, X., Lynch, J. G. & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37, 197–206.
- Zimmer, R.W., & Toma, E.F. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, 19(1), 75-92.





**Appendix 1A.**

Bivariate correlations of the variables used in the models of Substudy 1 for the Helsinki data

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Reading skills first grade	1	,171**	,132**	,159**	,045	,059	,107*	,361**	,251**	,191**	-,041	-,218**	,240**
2 Analogical reasoning first grade	,171**	1	,216**	,272**	,005	-,004	,033	,288**	,193**	,387**	-,039	-,161**	,199**
3 Visuo-spatial memory first grade	,132**	,216**	1	,288**	-,025	,010	,045	,327**	,132**	,267**	,042	-,165**	,123
4 Following instructions first grade	,159**	,272**	,288**	1	-,016	,001	-,028	,278**	,157**	,371**	,017	-,217**	,119
5 Learning orientation third/fourth grade	,045	,005	-,025	-,016	1	,701**	,680**	,060	,000	,036	-,094*	-,081	,029
6 Achievement orientation third/fourth grade	,059	-,004	,010	,001	,701**	1	,606**	,108*	,046	,065	-,149**	-,152**	,052
7 Agency: Effort third/fourth grade	,107*	,033	,045	-,028	,680**	,606**	1	,120**	,042	,064	-,159**	-,113*	,048
8 Mathematical thinking third/fourth grade	,361**	,288**	,327**	,278**	,060	,108*	,120**	1	,282**	,418**	,138**	-,298**	,240**
9 Reading comprehension third/fourth grade	,251**	,193**	,132**	,157**	,000	,046	,042	,282**	1	,297**	-,081*	-,206**	,202**
10 Reasoning skills third/fourth grade	,191**	,387**	,267**	,371**	,036	,065	,064	,418**	,297**	1	-,005	-,139**	,287**
11 Gender (0=Boy, 1= Girl)	-,041	-,039	,042	,017	-,094*	-,149**	-,159**	,138**	-,081*	-,005	1	,096*	-,021
12 Support needs (0= No support, 1= Support)	-,218**	-,161**	-,165**	-,217**	-,081	-,152**	-,113*	-,298**	-,206**	-,139**	,096*	1	-,111
13 Mothers' education	,240**	,199**	,123	,119	,029	,052	,048	,240**	,202**	,287**	-,021	-,111	1

\*\* p<.01. \* p<.05

**Appendix 1B.**

Bivariate correlations of the variables used in the models of Substudy 1 for the Vantaa data

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Reading skills first grade	1	,298**	,136*	,292**	,033	,097	,057	,334**	,209**	,250**	-,099	-,309**	,194**
2 Analogical reasoning first grade	,298**	1	,331**	,369**	,020	,081	,098	,377**	,228**	,375**	-,063	-,265**	,151**
3 Visuo-spatial memory first grade	,136*	,331**	1	,278**	-,005	-,013	-,008	,288**	,136*	,283**	,074	-,154**	,077
4 Following instructions first grade	,292**	,369**	,278**	1	,004	,013	,022	,386**	,104	,346**	-,052	-,215**	,148*
5 Learning orientation third/fourth grade	,033	,020	-,005	,004	1	,690**	,601**	,110*	,035	,048	-,115*	-,055	,017
6 Achievement orientation third/fourth grade	,097	,081	-,013	,013	,690**	1	,598**	,170**	,010	,021	-,195**	-,112*	,073
7 Agency: Effort third/fourth grade	,057	,098	-,008	,022	,601**	,598**	1	,168**	-,032	,046	-,120*	-,208**	-,039
8 Mathematical thinking third/fourth grade	,334**	,377**	,288**	,386**	,110*	,170**	,168**	1	,262**	,470**	,028	-,295**	,141*
9 Reading comprehension third/fourth grade	,209**	,228**	,136*	,104	,035	,010	-,032	,262**	1	,209**	-,086	-,134*	,149*
10 Reasoning skills third/fourth grade	,250**	,375**	,283**	,346**	,048	,021	,046	,470**	,209**	1	-,003	-,160**	,171**
11 Gender (0=Boy, 1= Girl)	-,099	-,063	,074	-,052	-,115*	-,195**	-,120*	,028	-,086	-,003	1	,095	-,006
12 Support needs (0= No support, 1= Support)	-,309**	-,265**	-,154**	-,215**	-,055	-,112*	-,208**	-,295**	-,134*	-,160**	,095	1	-,209**
13 Mothers' education	,194**	,151**	,077	,148*	,017	,073	-,039	,141*	,149*	,171**	-,006	-,209**	1

\*\* p<.01. \* p<.05

## Appendix 2.

Bivariate correlations of the variables used in the models of Substudy 2

Variable	1	2	3	4	5	6	7	8	9	10	11
1 Learning to learn test score fourth grade	1	,573**	,457**	,054	,005	-,010	-,045	,015	-,087*	,083*	-,232**
2 Learning to learn test score sixth grade	,573**	1	,477**	,083*	,036	,064	,047	,134**	-,066	-,002	-,332**
3 Analogical reasoning fourth grade	,457**	,477**	1	,030	,054	,053	,034	,025	-,016	-,051	-,198**
4 Learning orientation: Change	,054	,083*	,030	1	,625**	,616**	,228**	,181**	,162**	,086*	,023
5 Achievement orientation: Change	,005	,036	,054	,625**	1	,507**	,180**	,138**	,139**	,046	,043
6 Agency: Effort: Change	-,010	,064	,053	,616**	,507**	1	,192**	,153**	,128**	,041	-,008
7 Task interest: Arithmetical Operations: Change	-,045	,047	,034	,228**	,180**	,192**	1	,328**	,223**	,029	,017
8 Task interest: Mental Arithmetics	,015	,134**	,025	,181**	,138**	,153**	,328**	1	,221**	,031	-,053
9 Task interest: Reasoning task: Change	-,087*	-,066	-,016	,162**	,139**	,128**	,223**	,221**	1	,027	-,007
10 Gender (0=Boy, 1= Girl)	,083*	-,002	-,051	,086*	,046	,041	,029	,031	,027	1	,091**
11 Support needs (0= No support, 1= Support)	-,232**	-,332**	-,198**	,023	,043	-,008	,017	-,053	-,007	,091**	1

\*\* p<.01. \* p<.05

**Appendix 3.**

Bivariate correlations of the variables used in the models of Substudy 3

Variable	1	2	3	4	5	6
1 LTL Test score	1	,210**	-,364**	,644**	,494**	,510**
2 Mastery attitudes	,210**	1	-,063*	,230**	,076**	,317**
3 Detrimental attitudes	-,364**	-,063*	1	-,368**	-,226**	-,131**
4 Grade point average (GPA)	,644**	,230**	-,368**	1	,418**	,279**
5 Analogical reasoning	,494**	,076**	-,226**	,418**	1	,195**
6 Time on task: All cognitive tasks	,510**	,317**	-,131**	,279**	,195**	1

\*\* p<.01. \* p<.05