Nordic Council of Ministers

Northern Lights on TIMSS and PISA 2018

Northern Lights on TIMSS and PISA 2018

TemaNord 2018:524

Northern Lights on TIMSS and PISA 2018

ISBN 978-92-893-5565-0 (PRINT)
ISBN 978-92-893-5566-7 (PDF)
ISBN 978-92-893-5567-4 (EPUB)
http://dx.doi.org/10.6027/TN2018-524

TemaNord 2018:524
ISSN 0908-6692

Standard: PDF/UA-1
ISO 14289-1

Cover photo: Unsplash.com

Print: Rosendahls
Printed in Denmark

5041 0457
Ecolabelled Printed Matter
Rosendahls

Nordic Council of Ministers/Publication Unit
Ved Stranden 18
DK-1061 Copenhagen K
Denmark
Phone +45 3396 0200
pub@norden.org

Nordic co-operation
Nordic co-operation is one of the world's most extensive forms of regional collaboration, involving Denmark, Finland, Iceland, Norway, Sweden, and the Faroe Islands, Greenland and Åland.

Nordic co-operation has firm traditions in politics, economics and culture and plays an important role in European and international forums. The Nordic community strives for a strong Nordic Region in a strong Europe.

Nordic co-operation promotes regional interests and values in a global world. The values shared by the Nordic countries help make the region one of the most innovative and competitive in the world.

The Nordic Council of Ministers
Nordens Hus
Ved Stranden 18
DK-1061 Copenhagen K, Denmark
Tel.: +45 3396 0200 www.norden.org

Download Nordic publications at **www.norden.org/nordpub**

Contents

Foreword

The results from PISA 2015 and TIMSS 2015 were published in November and December 2016. All of the Nordic countries participated in PISA. Denmark, Finland, Norway, and Sweden participated in TIMSS grade 4, and Norway and Sweden participated in TIMSS grade 8.

The Nordic countries have similarities, but also differences, which makes it interesting and valuable to carry out analyses in a Nordic perspective. In this report, researchers from all of the Nordic countries have performed in-depth analyses on different policy-relevant themes based on the results presented in 2016. The purpose of this report has been to present policy-relevant analyses of TIMSS and PISA in a way that is accessible for policy makers on different levels in the Nordic countries, with the aim to contribute to further development in the education area.

The introductory chapter is an overview of international studies and their significance for the Nordic countries. This chapter is written by Anne-Berit Kavli at the Norwegian Directorate for Education and Training, in cooperation with the Nordic Evaluation Network group. The second chapter deals with social inequality in student performance, and it is a comparison of methodological approaches. The chapter is written by David Reimer, Simon Skovgaard Jensen, and Christian Christrup Kjeldsen. The third chapter about the importance of teachers and their instruction for students' motivation is written by Trude Nilsen, Sigrid Blömeke, and Ronny Scherer. The fourth chapter is written by Magnus Oskarsson, Hanna Eklöf, Marit Kjaernsli, and Helene Sørensen and is a Nordic view on students' interest in science. The fifth chapter analyzes the possible effects of the digitalization of the PISA reading test and is written by Maria Rasmusson and Ulf Fredriksson. The sixth chapter by Bent Sortkaer deals with students' perception of feedback. The final chapter asks the question "Urban advantage in education?" and explains the achievement differences in science between metropolitan and other areas in Finland and Iceland in PISA. This chapter is written by Kari Nissinen, Jouni Vettenranta, Juhani Rautopuro, Ragnar F. Ólafsson, and Almar M. Halldórsson.

The Nordic Evaluation Network group has been acting as the editorial group, led by Anita Wester at the Swedish National Agency for Education. Every paper has also, on two occasions, been reviewed by a panel consisting of Jouni Välijärvi, Finland, Júlíus K. Björnsson, Norway, and Allyson Macdonald, Iceland.

The editorial group wants to thank all of the contributors to this report. Like the previous editions in the Northern Lights series, this publication has received financial support from the Nordic Council of Ministers.

Stockholm in May 2018

Anita Wester
Senior Advisor at the Swedish National Agency for Education
Editor

1. TIMSS and PISA in the Nordic countries

Anne-Berit Kavli, Norwegian Directorate for Education and Training

The Nordic countries are active participants in international large-scale assessments. These studies represent a large and important knowledge base, and they have influenced education policy development globally. The Nordic countries represent a unique "laboratory" for in-depth analyses of the outcomes of these studies because of the many cultural similarities combined with clear national characteristics with respect to results and policy development. The biannual Northern Lights publications aim to present highly policy-relevant analyses in a Nordic context in order to enhance the use and understanding of the data from large-scale assessments, and to stimulate Nordic cooperation.

1.1 Background

Today, TIMSS (Trends in Mathematics and Science Study) and PISA (Programme for International Student Assessment) are the two largest and most widespread international large-scale assessments of learning outcomes.

1.1.1 TIMSS

TIMSS is an IEA [1] study and has been conducted every fourth year since 1995. Like all IEA studies, TIMSS is grade based and curriculum based, and it is designed to assess trends in student achievement in mathematics and science at the primary (grade 4) and

[1] IEA is the International Association for the Evaluation of Educational Achievement.

lower secondary (grade 8) level. Countries can choose to participate at both grade levels or only at grade 4 or grade 8. TIMSS assesses both content knowledge and the students' ability to apply their knowledge, along with questionnaires for students, teachers, parents, and school principals on social background, learning environment, and conditions for learning. TIMSS Advanced is an additional option that assesses final-year upper secondary students' achievement in advanced mathematics and physics.

1.1.2 PISA

PISA is an OECD[2] study designed as a triennial study on relevant skills and competencies acquired by 15 year olds. The OECD conducted PISA for the first time in 2000. The core domains of PISA are literacy in reading, mathematics, and science. PISA is not curriculum based, but is designed to assess fundamental skills that are relevant for work and lifelong learning. These skills are described in frameworks for each domain, and the focus is on how students are able to apply their skills and competencies in real-life situations. Each cycle of PISA also contains an assessment of a new innovative domain, and in 2015 the innovative domain was collaborative problem solving. In 2015, PISA was for the first time conducted as a computer based assessments. PISA is accompanied by background questionnaires for students and principals, and there are optional questionnaires for parents and teachers.

1.1.3 Trend studies

Both PISA and TIMSS are trend studies that are designed to measure the development of learning outcomes over time. This is a very important aspect because most national assessment systems are not designed to measure change over time. Among the Nordic countries, so far only Norway has developed national assessments that can follow change over time, but this system has only recently been introduced.

TIMSS assesses both mathematics and science in each cycle, so countries can calculate trends from the first year they participated in the study. Both Norway and Sweden have participated in TIMSS since 1995 and now have 20 years of trend data from the study.

---

[2] OECD is the Organisation for Economic Cooperation and Development.

In PISA, the domains of reading literacy, mathematics, and science are assessed in each cycle, but only one subject is assessed as a major domain (see Table 1). Trends are calculated from the first time a domain has been assessed as major, which means that countries have trends in reading literacy from 2000, in mathematics from 2003, and in science from 2006.

Table 1: Overview of major domains and innovative domains in PISA

Year	Major domain	Innovative domain
2000	Reading	Students' self-assessment of learning strategies
2003	Mathematics	Problem solving
2006	Science	Assessment of student attitudes towards science
2009	Reading	Electronic reading assessment
2012	Mathematics	Computer-based problem solving
2015	Science	Computer-based collaborative problem solving

In addition to trends in the cognitive domains, both PISA and TIMSS provide analyses on how students' learning environment and conditions for learning have developed over time.

1.1.4 Nordic participation in TIMSS and PISA

All of the Nordic countries have participated in PISA since the beginning in 2000. In total, 72 countries and economies participated in PISA 2015, and among them were all the 35 OECD countries.

Sixty countries and benchmarking regions worldwide participated in TIMSS 2015. All the Nordic countries except Iceland participated in TIMSS grade 4 (grade 5 in Norway), while only Norway and Sweden took part in TIMSS grade 8 (grade 9 in Norway). Norway and Sweden were also among the nine countries that participated in TIMSS Advanced.

In 2015, Norway changed the main target populations in TIMSS to grades 5 and 9, but in order to maintain trends Norway also participated with grade 4 and grade 8 as benchmarking entities. The main reason for this change was for Nordic comparisons. Norwegian pupils start school the year they turn 6, while in Sweden, Finland, and Denmark most children start preschool class the year they turn 6 and then start school the year they turn 7. This means that Norwegian pupils in grade 5 are the same age and have the same total amount of schooling as pupils in grade 4 in the other Nordic countries.

Table 2: Nordic participation in TIMSS

Country	TIMSS Grade 4	TIMSS Grade 8	TIMSS Advanced
Denmark	2007, 2011, 2015	1995	
Finland	2011, 2015	1999,[3] 2011	
Iceland	1995	1995	
Norway	1995, 2003, 2007, 2011, 2015	1995, 2003, 2007, 2011, 2015	1995, 2008, 2015
Sweden	2007, 2011, 2015	1995, 2003, 2007, 2011, 2015	1995, 2008, 2015

Note: In 1995 the sampling design was different from later cycles in TIMSS, and in primary school countries participated with grades 2 and 3 and in lower secondary school with grades 6 and 7.

1.2 Trends in the Nordic countries

Except for Iceland, all of the Nordic countries can now measure development over time both in primary and lower secondary education using data from PISA and TIMSS. Iceland currently only takes part in PISA and does not have international results or trends for primary education, with the exception of the IEA-PIRLS reading literacy study in 2001 and 2006.

TIMSS and PISA have different frameworks and cannot be directly compared, but still the studies complement each other and show quite similar trends. For example, changes observed in TIMSS at primary level from 2007 until 2011 were continued at lower secondary level both in PISA and TIMSS in 2015 (see Figures 1–6).

1.2.1 TIMSS results and trends

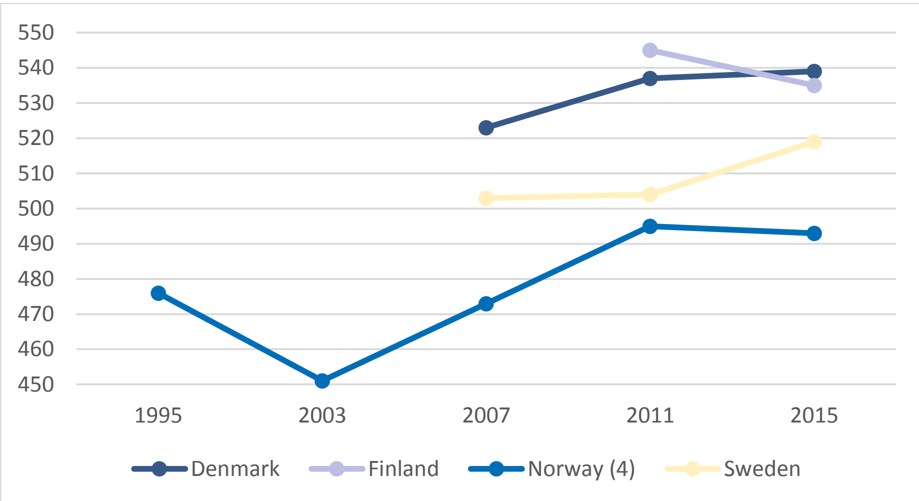
Of the Nordic countries, only Norway has trend data for both populations (primary and lower secondary) for the whole period since 1995. Both in mathematics and science, Norway experienced a significant decline in results in the period from 1995 until 2003. From 2003 until 2015 there has been a significant positive development in mathematics at both grade levels. In science, there has been a positive development at grade 4 since 2003, while at grade 8 the negative trend continued until 2007. After that there have only

[3] Did not meet international requirements for data.

been small changes in science. Norway does not have trend data for grades 5 and 9, but the results for these grades in 2015 were very positive, particularly in mathematics.

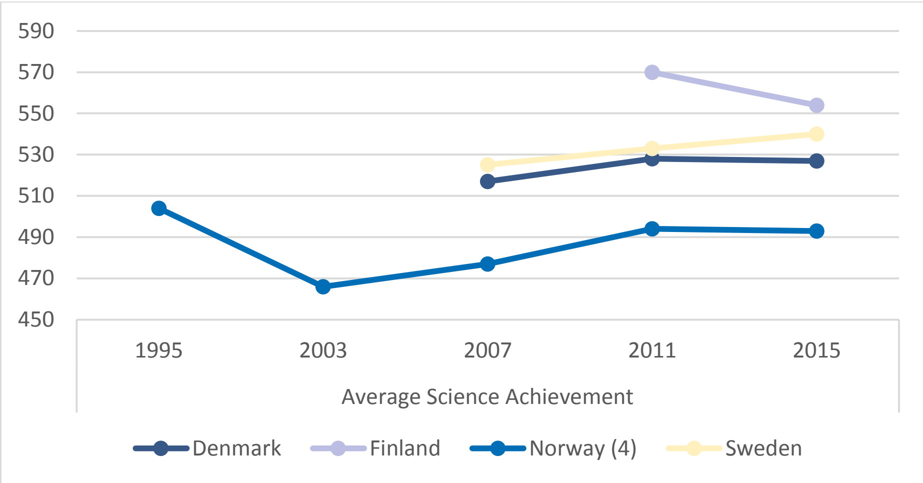
In Sweden, there was a continuous decline in both mathematics and science in grade 8 from 1995 until 2011, while there was a significant improvement from 2011 to 2015. At grade 4, there has been a small improvement in both mathematics and science in the period from 2007 until 2015. Like Sweden, Denmark has seen small improvements in both subjects since 2007 in grade 4, while Finland experienced a decline in both subjects from 2011 to 2015.

Figure 1: Nordic trends in Mathematics – TIMSS grade 4 (Mullis et al., 2016)



Note: Figure 1 shows that both Finnish and Danish pupils in grade 4 are high achievers in mathematics, even if we see a decline in the Finnish results. The Swedish main scores are about 20 points lower than Denmark and Finland, while Norwegian results are the lowest. However, it is important to bear in mind that Norwegian 4th graders are 1 year younger than the others. The Norwegian pupils in grade 5, which was the main population in 2015, had an average score of 549, which was higher than both Finland and Denmark.

Figure 2: Nordic trends in Science – TIMSS grade 4 (Martin et al., 2016)



Note: Figure 2 shows that Finnish pupils in grade 4 are very high achievers in science, while Danish and Swedish pupils have rather similar scores. Again the Norwegian scores are significantly lower, mainly due to their lower age. With the change of main population to grade 5, the Norwegian score in Science was 538, which was slightly below the Swedish score of 540.

Figure 3: Norwegian and Swedish trends in Mathematics – TIMSS grade 8 (Mullis et al., 2016)

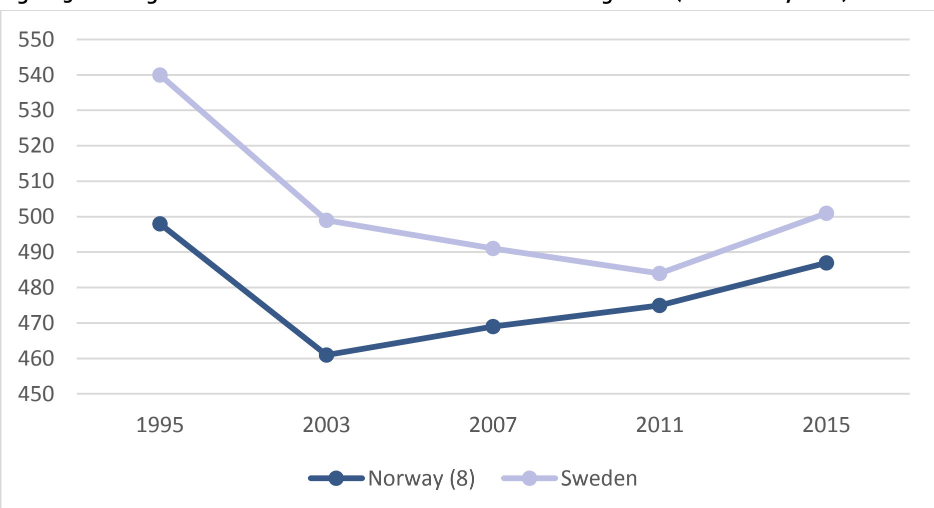
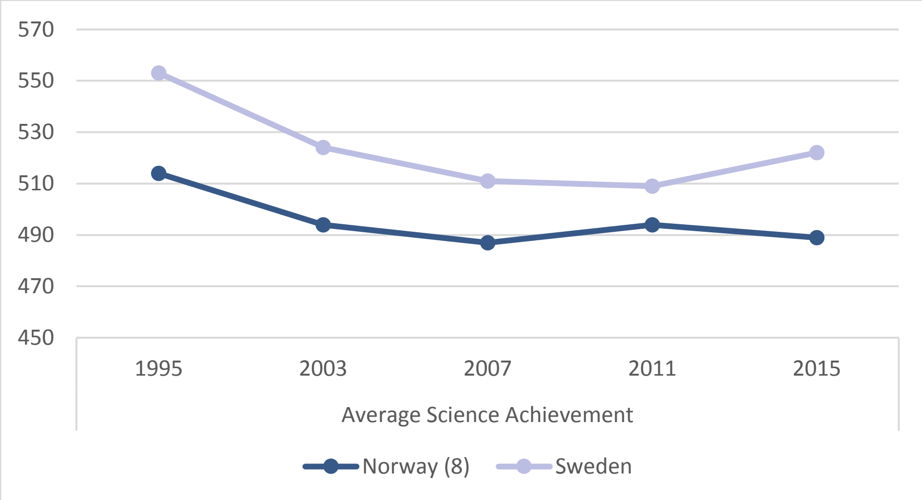


Figure 4: Norwegian and Swedish trends in Science – TIMSS grade 8 (Martin et al., 2016)



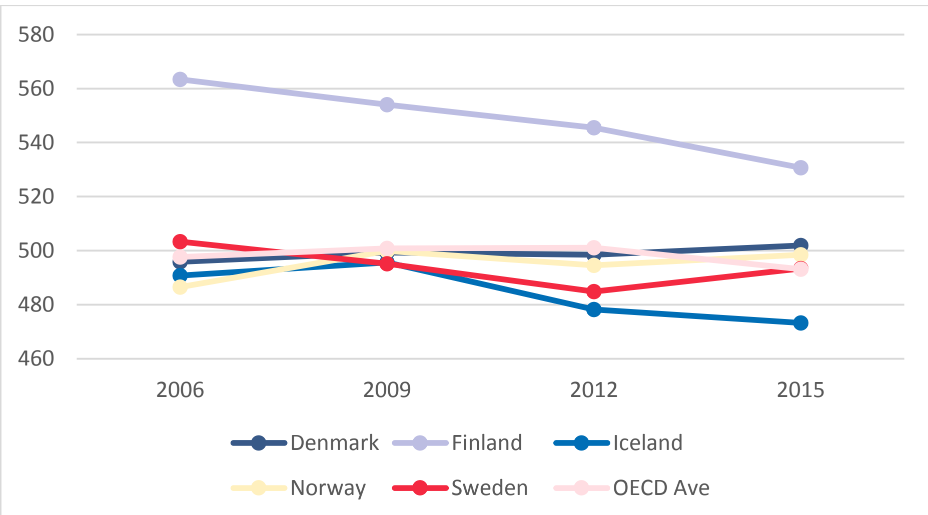
Note: As shown in Figures 3 and 4, the Swedish grade 8 pupils have scored higher than the Norwegian pupils through the whole period, but we also see that the gap has been considerably reduced in mathematics. In 2015, the Norwegian students in grade 9 had an average score of 512 in mathematics and an average score of 509 in science. These students are the same age as the Swedish grade 8 students.

1.2.2 *PISA results and trends*

All of the Nordic countries have taken part in PISA since the beginning in 2000 and now have 15 years of trends. As shown in Figures 5–7, all the Nordic countries except Finland have had results with rather small variations around the OECD average during the whole period.

Finland started with very high results and has been among the top-performing countries in the world during the whole period. Despite these high results, Finland has seen a steady and significant decline in all three domains since 2006. In Sweden, the trend was continuously declining until 2012, but in 2015 Sweden had a significant improvement in all three domains. In Denmark, there have been only small changes. There was a decline in mathematics from 2003 to 2012, but in 2015 the results improved again and Denmark scored at the same level as Finland. Norway experienced a decline in all three domains from 2000 to 2006, but after that there has been a small but significant increase in all domains. The increase has been highest in reading, where Norway scored significantly above the OECD average in 2015. In Iceland, there has been a declining trend, and their results in 2015 were significantly below the OECD average in both science and reading. The OECD average has also declined over this period.

Figure 5: Science trends in PISA – Nordic Countries (OECD, 2016)



Note: In Science, Finland's scores have continuously been very high, and they are still almost 40 points above the OECD average. Denmark, Norway, and Sweden have had scores around the OECD average, while the Icelandic results have declined and are now significantly below the OECD average.

Figure 6: Mathematics trends in PISA – Nordic Countries (OECD, 2016)

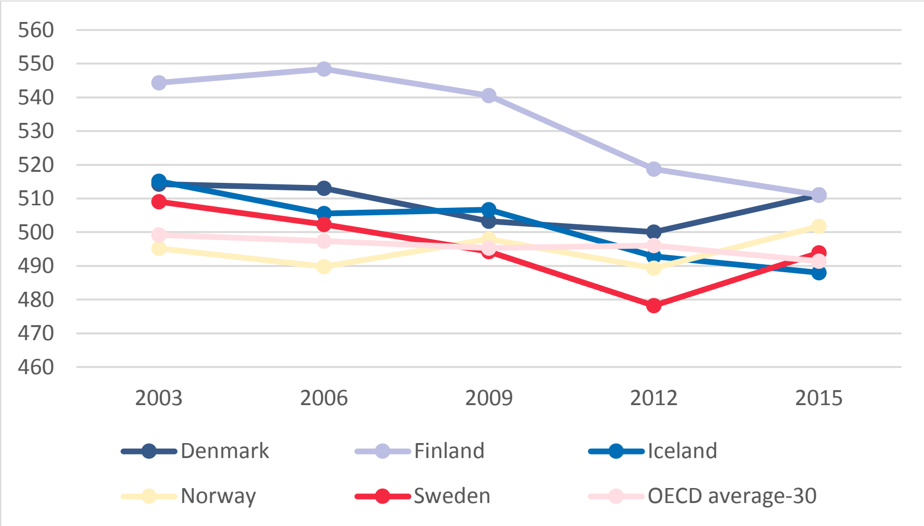
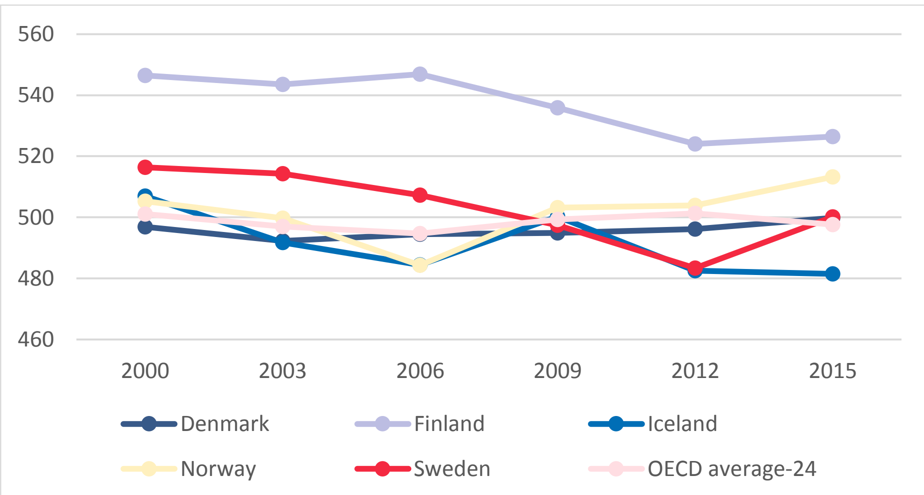


Figure 7: Reading trends in PISA – Nordic Countries (OECD 2016)



Note: Figures 5 and 7 show the development in mathematics and reading, and again the Finnish results have been significantly above the other Nordic countries, even with declining results. The exception is mathematics in 2015, where Denmark and Finland had similar scores.

1.3 Use and impact of international studies in the Nordic countries

In all of the Nordic countries, results from international studies play an important role as part of the evidence base for educational policy development. The main reasons for joining the studies are the wish to analyse strengths and weaknesses of the national education systems in an international and comparative perspective, to follow trends and developments over time, to build international networks, and to learn from other countries. However, it is often emphasised that results from international studies cannot stand alone but have to be analysed in a national context, where national exams, tests, and evaluations play an important role.

For PISA as an OECD study, the decision to join the study is a political decision taken by the Education Ministry. Because the IEA is a non-governmental membership association, the decision process for TIMSS varies across countries depending on how the membership is organised and how the studies are financed. In Norway and Finland, the decision to join the study is made by the Education Ministry. In Sweden, the decision is made by the Swedish National Agency for Education (Skolverket) in consultation with the Education Ministry, while in Denmark participation is decided by Aarhus University in dialogue with the Education Ministry.

In all of the Nordic countries except Finland, the respective national education agencies are responsible for the follow-up of their countries' participation in the international large-scale studies like TIMSS and PISA, while the national research coordinators or project managers in some of the countries are contracted from universities or research institutes.

1.3.1 Attention and impact

In a study on the impact of PISA, Breakspear found that the PISA results have contributed to setting the agenda for policy discussions among policy-makers and experts in many countries and that PISA results are used as evidence to argue for the need for national improvement based on medium or poor performance (Breakspear, 2012). In the Nordic countries, the results of PISA have received much attention and have been used as a basis for educational policy analyses, e.g. in national public reports and in white papers on education.

In the participating countries, TIMSS has provided important data and feedback in mathematics and science, particularly at the primary level. TIMSS is designed as a grade and class-based study with strong links to the curriculum and has led to more in-depth analyses both on curricular content and the relation between teaching characteristics and learning achievements (see, for example, Nilsen & Gustafsson, 2016, and the national websites for PISA and TIMSS reports).[4, 5, 6, 7, 8, 9]

International large-scale assessments are sample-based studies and are designed for system-level analyses. The studies are not intended to give results at the individual level or the school level, and the target groups for reporting have mainly been policy makers, researchers, and other stakeholders at the national level. Still, most of the countries strive to make the results and analyses known, understood, and used by practitioners and leaders at the school level and local school administrations. This is done through conferences and seminars in addition to shorter and more targeted publications. The publications from the Swedish National Agency for Education (Skolverket), "Med fokus på ..."are an example of these.[10]

1.3.2 Educational debates

Results from the international large-scale assessments receive a lot of media attention. In particular, this is the case for PISA, but TIMSS has also received increasing attention. In cases where results have declined or been poorer than expected, this has raised national debates on the quality of education and been an incentive for educational changes and reforms. The form and content of the national debates initiated by PISA and TIMSS have varied among countries and across cycles depending on the achievements in each cycle and the trends over time.

Finland has been among the top achievers during the whole period, even if its results have been somewhat declining both in PISA and TIMSS. This has resulted in quite extensive "educational tourism", where educational policy makers and

---

[4] TIMSS and PISA in Norway: http://www.uv.uio.no/ils/forskning/prosjekt-sider/

[5] TIMSS and PISA in Denmark: http://edu.au.dk/forskning/internationaleundersoegelser/

[6] TIMSS and PISA in Finland: https://ktl.jyu.fi/pirls-timss, https://ktl.jyu.fi/pisa/en

[7] PISA in Iceland: https://www.mms.is/pisa

[8] TIMSS in Sweden: https://www.skolverket.se/statistik-och-utvardering/internationella-studier/timss

[9] PISA in Sweden: https://www.skolverket.se/statistik-och-utvardering/internationella-studier/pisa

[10] https://www.skolverket.se/sok/get?q=Med+fokus+p%C3%A5&search=S%C3%B6k

researchers from all over the world have visited Finland to study and learn from their education system. In Finland, the main concern has not been the academic level, but rather discussions about the learning environment and school culture, pupils' engagement and initiative, and a focus on more generic and overarching skills.

Sweden performed well above the international average in the first rounds of TIMSS and PISA, but showed a significant and declining trend for many years until 2012. This resulted in an intense debate and official reports by the Swedish government ("offentlig utredning") on the quality of Swedish education (SOU 2017:35 and SOU 2017:38). In addition to declining results, the differences between schools have been larger in Sweden than in the other Nordic countries and have increased during the period (Rapport 467, 2018). This has lead to increased attention and debate about equity, especially after PISA 2012.

In Norway, the weak and declining results during the first cycles of PISA and TIMSS caused an intense debate both in the media and among policymakers and stakeholders. This debate started earlier than in Sweden and resulted in several actions and reforms, which seem to have led to positive developments after 2006.

Iceland performed around the international average until the decline in results in 2012. Before then there was little debate about PISA in the country. After 2012, the decline in results caused concern about the educational quality in Iceland, and also led to discussions about the relevance and validity of PISA in the Icelandic context and how much it should influence the country's education policy.

Denmark's performance has been above the international average through the whole period, and there have been only small variations in the results across the different cycles. However, the results have repeatedly been lower than national ambitions and expectations and have led to a series of changes and reforms. The strong focus on weak results from media and policymakers has also led to considerable negative criticism from teacher unions and academia. Parallel to this criticism, there has been a strong methodological debate in Denmark about PISA and the way they calculate scores and rankings (Kreiner & Christensen, 2014). This kind of criticism has also been raised in the other Nordic countries, but not to the same extent as in Denmark.

As the attention to and impact of PISA and other large-scale assessments have increased, there has been an increasing critique in many countries on the strong focus on PISA, and more broadly of an increasing "global testing culture". This critique has been particularly strong from parts of academia and from teacher unions. The critical voices raise questions about the validity and reliability of PISA across countries with very different social, cultural, and economic backgrounds. The critics argue that the focus on rankings and test results leads to a situation where educational quality is only based on rather narrow test results that do not necessarily cover the overarching goals of education. Following this, there is a discussion about the OECD's role and influence in education policy development in relation to what is often perceived as a global standardisation of education policy (Benavot & Meyer, 2013; Smith, 2016).

1.4 Educational policy development in the Nordic countries

Perhaps the most important effect of large-scale assessments has been the fact that education and educational quality have been placed high on the political agenda, both globally and at the national level. The evidence derived from large-scale assessments gives the opportunity to compare the outcomes of different educational systems, to monitor equity and inclusion, and to discuss which factors are important for efficient teaching and learning and thereby provide an evidence base for further policy development.

Despite the criticisms of large-scale assessments, there is strong global agreement that monitoring of educational quality and access and the outcomes of education are both important and necessary in order to improve education, to achieve important educational goals, and to secure equity and inclusion for all (see, for example, UNESCO 2017). However, assessments are only the start of a much longer process. For improvement of education to take place, assessments must be accompanied by contextual analyses, results need to be disseminated and discussed by all stakeholders, and policy changes need to be accompanied by concrete and targeted actions.

In all of the Nordic countries, the strengthened focus on the quality of education during the last decade has led to important changes in education policy. This is not only due to studies like PISA and TIMSS, but is more broadly related to an increased focus on educational governance, efficiency, standard setting, and accountability (see, for example, Burns et al., 2016).

1.4.1 Changes and reforms

In all of the Nordic countries, there have been major educational changes and reforms during the period from 2000 to 2015. The changes include curricular reforms, increased focus on accountability and quality assessment, teacher education and professional development, and a variety of national strategies and support systems to improve the learning environment and to strengthen learning in basic skills. In general, this has led to a strengthening of basic skills like reading, mathematics, and science. Typical for the curricular reforms is the focus on clear achievement goals in all subjects.

In Denmark, the curricular reforms started in 2001 with the reform "Clear Goals". This reform was further developed and simplified as "Common Goals", first in 2009 and then in 2015. This was accompanied by changes in exams and assessments and the introduction of national tests in 2005.

In August 2014, a reform of the Danish Folkeskole was introduced based on political agreement on the need to strengthen academic competences. The background for this was several reports that pointed to weak performance and large variations due to pupils' gender and social and ethnic background. The focus areas in the reform included longer and more varied school days and an enhancement of the teachers' and school leaders' professional competences along with the establishment of national goals and the simplification of rules.

Finland introduced a major curricular reform in 2014. This reform covered pre-primary, primary, and secondary education and was implemented in 2016. The reform emphasizes a common perspective on pedagogy, a culture for learning, and cross-curricular competences, and the focus on pupils' involvement and engagement has been strengthened.

In Norway, the weak results from PISA 2000 initiated the development of a national system for quality assessment. As part of this, national testing in numeracy and reading literacy in Norwegian and English started in 2005. In 2006, the curricular reform "Knowledge Promotion" was introduced both in primary and secondary education. The most important changes in the Knowledge Promotion reform were the strengthened focus on basic skills from the first grade, clearer learning goals in all subjects, and local freedom with respect to school organisation, methods, and learning material. These changes have been accompanied by national strategies to strengthen reading, mathematics, and science and by increased focus and support for professional development for teachers. In addition, the national testing program has been improved and redesigned to measure trends from 2014.

In Sweden, there was a curricular revision that included pre-primary, primary, and secondary education as well as adult education. Strengthening of goals for knowledge and skills also led to changes in pupils' assessment. In order to improve learning in basic skills like reading, mathematics, science, and technical skills, Sweden introduced and conducted large, national strategies for professional development in these domains (Matematiklyftet, Läslyftet, and NT-satsningen). These strategies have been continued as part of the new and broader School Development Program.

In Iceland, a new National Curriculum Framework was published in 2011, and in 2013 new subject curricula with greater emphasis on reading and science were introduced. The new curriculum framework puts weight on competencies with reference to the EU's key competencies and its qualification framework. This framework has also led to changes in the assessment system, with a new grading scale and clearer definitions of competences. A national initiative with a focus on reading was initiated in 2014 with a government white paper on education reform. The government has put significant financial resources into this initiative, and these have been used to develop new reading tests and for counselling and support for municipalities and schools and for awareness raising.

1.5 Introduction to the articles

The aim of this report is to provide more in-depth analyses of the data from TIMSS and PISA in a Nordic context. The Nordic countries share cultural similarities and joint values regarding democracy, equity, trust, and openness, and our educational systems are strong and are based on the same core values. At the same time, our schools face many of the same challenges in a rapidly changing society. We all have concerns regarding vulnerable groups and increasing differences between those who succeed and those who fall out of the system. And we all meet new demands on our education systems, where some of the key words are communication and cooperation, digitalisation, in-depth learning, and problem solving.

Large-scale studies like TIMSS and PISA aim to provide countries with a relevant and updated knowledge base for educational policy development. Analyses of these data in a Nordic context can give us a better understanding of the similarities and differences we are facing and how we can understand the results, learn from each other, and inform the educational policy debate and development in our respective countries.

In this report, the articles analyse and discuss the following important issues in the educational policy debate:

- the importance of interest, motivation, and feedback to students;
- how teachers can make a difference;
- the measurement and impact of social inequality; and
- the transformation to computer-based testing.

1.5.1 Motivation and feedback

There are two articles in the report discussing pupils' motivation and experience of feedback.

Previous PISA results have revealed a comparatively low interest in science among students in the Nordic countries. Eklöf et al. discuss Nordic students' interest, motivation, and self-beliefs in science based on PISA results from 2006 and 2015. Their findings show an increased interest in science in most Nordic countries, but at the same time they observe increased gender differences and greater variation in enjoyment and self-efficacy among students. While enjoyment of science and science self-efficacy are positively related to performance, instrumental motivation and enjoyment of science are associated with an increased likelihood that the student expects a science-related career.

Bent Sortkær's article, "Feedback for everybody? – Variations in students' perceptions of feedback", analyses how teacher feedback is perceived by individual students in Nordic science classrooms. More specifically, the article discusses whether there is a relationship between the amount of feedback perceived by the students and their gender and their social and ethnic background. The analyses indicate significant differences in perceived feedback related to both gender and ethnic background. In all of the Nordic countries, boys perceive much more feedback than girls do. In Finland, Norway, and Sweden, students born in a different country report more perceived feedback than native students. The report also shows a close relationship between science performance and the amount of feedback.

1.5.2 How teachers can make a difference

In the article "How important are teachers and their instruction for student motivation and achievement in science?", Nilsen et al. discuss the relationship between instructional quality and learning outcomes. They also analyse which aspects of teacher quality are directly related to the quality of instruction and how this in turn is related to students' learning. The analyses show positive relations between teachers' pedagogical competence and student outcomes both in primary and lower secondary school, while teachers' formal education seems to be more important in the higher grades. The article also discusses how the findings can be related to teachers' professional development. Another important finding is how teachers' self efficacy and motivation are important for students' learning.

1.5.3 Impact of social inequality

It is well documented that students' socio-economic background is a strong predictor for learning outcomes. However, there is more uncertainty about how these background factors should be measured and how different types of measures and analytical methods influence the results. This is discussed in the article "Social inequality in student performance in the Nordic countries: A comparison of methodological approaches" by Reimer et al. Their analyses show that the correlations between student background and learning outcomes are quite complex and need to be more nuanced. For example, the association between parental background and test achievement seems systematically higher for girls than for boys, and there is a tendency to overestimate the effect for low-performing students, while the association is underestimated for the high performers. Also, how parental background is measured has implications for the results, which shows that researchers and political advisors need to very thoughtful about which indicators to choose and how to use them.

Another discussion related to social background is how regional differences and differences between urban and rural schools can be explained. This is discussed in the article from Nissinen et al. where differences between capital and rural regions in Iceland and Finland are analysed. In both countries, students from the capital regions of Helsinki and Reykjavik outperform students from the rural areas. The article shows that these differences to a large extent can be explained by students' families' socio-economic status and cultural capital, as well as students' own ambitions and expectations.

1.5.4 *Implications of computer-based testing*

In 2015 PISA was transformed from paper based to computer-based testing, and in 2019 the same will happen in TIMSS. Also, at a national level the Nordic countries are in the process of transforming their national assessments to computers. There are many advantages to computer-based assessments. For example, computer-based assessment opens up for more diverse and varied types of test items that are more in line with real-life situations. Thus, it gives the possibility to assess skills that cannot otherwise be tested. Computer-based assessments also open up for more individually adapted testing and have shown to be more motivating for students because they use the tools and environments that young people today are more and more used to and which they utilize in both their schoolwork and in their leisure time.

At the same time, there is concern that a change of test mode can influence the results and by that reduce the reliability of trend data. In their article, Rasmusson and Fredriksson discuss how the change to computer-based assessment might influence students' results in reading. They conclude that among the Nordic countries there have not been any dramatic changes in the results on reading literacy in PISA 2015. Only in Sweden, in reading literacy, can a major change in results be observed, which might be related to the amount of time students spend on the Internet. Even if it is not possible to clearly show whether the change of test mode in PISA 2015 has influenced the results, the authors warn that there are reasons to be careful when comparisons are made of PISA results from 2015 with results from earlier PISA studies. The comparative link between the PISA studies from different years might be weaker in 2015 than earlier. Still, the article concludes that computer-based testing moves the test practice closer to the everyday practice of many students, and also has many advantages for test administration.

1.6 Why Northern Lights?

PISA and TIMSS represent the two largest international comparative studies on learning outcomes, and they have both had great influence on educational policy development world wide. The Northern Lights publications aim to provide in-depth analyses in a Nordic context in order to stimulate Nordic cooperation, make better and more informed use of the data, and encourage a debate on the future use and development of large-scale international assessments in a Nordic context.

1.7 References

Benavot, A. & H.-D. Meyer (2013). PISA, power, and policy: the emergence of global educational governance. Oxford, Symposium Books.

Breakspear, S. (2012). The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance, Paris: OECD Publishing.

Burns, T., et al. (2016). Education Governance in Action: Lessons from Case Studies. OECD Publishing "för kunskap och likvärdighet".

Kreiner, S. & Christensen, K.B. Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. Psychometrica, April 2014, Vol. 79(2), pp.201–31.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Science. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/

Nilsen, T. and J.-E. Gustafsson (2016). Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time, Springer International Publishing: Imprint: Springer. Vol. 2.

OECD (2016), PISA 2015 Results (Volume 1): Excellence and Equity in Education, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264266490-en

Rapport 467 (2018) Analyser av familjebakgrundens betydelse för skolresultaten och skillnader mellan skolor. En kvantitativ studie av utvecklingen över tid i slutet av grundskolan. Skolverket.

Smith, W. C. (2016). The Global testing culture: shaping education policy, perceptions, and practice. Oxford, Symposium Books.

SOU 2017:35 Samling för skolan – Nationell strategi för kunskap och likvärdighet.

SOU 2017:38 Samling för skolan – Nationella målsättningar och utvecklingsområden.

UNESCO (2017) Accountability in Education: Meeting our Commitments. Global Education Monitoring Report 2017/8.

2. Social inequality in student performance in the Nordic countries: A comparison of methodological approaches

David Reimer,[1] Simon Skovgaard Jensen,[1,2] and Christian Kjeldsen[1]

[1] Danish School of Education, Aarhus University.
[2] Via University College, Holstebro.

2.1 Introduction

Knowing the extent to which parental background affects the educational performance of students is highly relevant for both policy makers and researchers. High levels of educational inequality in a country can be a sign of insufficient support structures and negative learning environments for students from disadvantaged backgrounds, and the relationship between social origin and educational attainment has received significant attention in the Nordic countries. In both Denmark and Norway, recent school reforms have been introduced with the goal of increasing educational performance and at the same time reducing inequality in educational achievement (Olsen, Hopfenbeck, and Lillejord 2013; Rasmussen, Holm, and Rasch-Christensen 2015).[11] Nevertheless, the Nordic welfare states' efforts to "equalize education" by reducing disadvantages for less privileged students have a far longer tradition (Erikson and Jonsson 1996).

[11] To provide one example: One explicit goal of the Danish compulsory school reform, which was implemented in 2014, was to decrease the impact of social background in respect to students' academic achievement: "Folkeskolen skal mindske betydningen af social baggrund i forhold til faglige resultater" (see Kommunernes Landsforening 2013, p. 1).

Notwithstanding these efforts, a body of literature based on recent studies of the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) has documented substantial inequality in student performance according to students' social background in all OECD countries, including the Nordic welfare states (Marks 2006; OECD 2010, 2012).

Overall, the existence of pronounced levels of inequality according to students' social background is not contested. However, there is less agreement among researchers regarding the question of how background-related inequality should be measured. On the one hand, there is diversity in the type of parental background information that researchers use to assess inequality. Should one, for example, examine the relationship between the parents' level of education and their children's school performance, or is parental occupation the more relevant or objective indicator? Apart from the issue of identifying the "best" parental social background measure, there is also substantial methodological diversity in the way the strength of the association between social background and student performance is calculated. Some studies report relatively simple measures, such as mean performance differentials between students from varying backgrounds, while the PISA consortium uses the amount of explained variance (R^2) from regression analyses to quantify the extent of educational inequality in a country.

It follows that in this chapter we address two research questions. Based on data from the latest PISA (2015) study for all Nordic countries, we ask first *to what extent the usage of different parental background indicators (such as parental education, occupation, or PISA's own index) changes the conclusions about the degree of educational inequality in the different countries*. Second, we explore *whether the extent of inequality varies when we use a different statistical technique, quantile regression*, which provides a more nuanced understanding of the relationship between social background and educational achievement across the entire distribution of achievement in the PISA tests.

The remainder of the chapter is structured as follows. In the first part, we briefly discuss the main concerns related to the measurement of parental social background. Next, we explain the most common statistical method used by the OECD to calculate inequality. Subsequently, we present our analyses of inequality in educational achievement using a number of selected social background indicators. In the second part of the chapter, we report the results of our quantile regression. Finally, in the third section of this chapter we summarize our results and discuss the possible implications they have for policy and educational research.

2.2 Measuring parental social background

White (1982) as well as Sirin (2005) conclude in their international reviews of the literature measuring the association between socioeconomic status (SES) and academic achievement that there is considerable diversity in the operationalization of socioeconomic status. A variety of indicators such as parental class, education, and socio-economic status are frequently used interchangeably without any theoretical rationale, which makes comparisons of results across studies difficult (Sirin 2005; White 1982; White et al. 1993).

Overall, the measurement of parental background can be related to two more broad discussions in the social sciences. First, there is the discussion about whether parental background can be measured with one *continuous* indicator, such as a status-index, or whether a *categorical* approach, such as parental class or parental level of education, should be preferred (Ganzeboom, De Graaf, and Treiman 1992; Jonsson et al. 2009). When applying a continuous indicator, such as any index for the measurement of socioeconomic status, it is assumed that differences between parental background groups can be measured in one dimension only. Social scientists applying categorical approaches assume that members of society belong to clearly distinguishable groups or classes. Furthermore it is assumed that group members are similar to each other (internal homogeneity) but differ markedly from members of other groups (external heterogeneity) (Ganzeboom et al. 1992:3–4). Applying this logic to, for example, a social class classification would imply that all members of the "working class" are very similar to each other but are very different from members of all other classes. The theoretical rationale for grouping different individuals into different classes varies according to the respective categorical scheme. However, categorical approaches have in common the assumption of multidimensionality, meaning that group membership is determined by more than a single factor – such as skill level and personal responsibility in the well-known Erikson–Goldthorpe–Portocarero (EGP) class schema (Erikson, Goldthorpe, and Portocarero 1979).

Second, the question of *which background dimension* should be used to measure parental background is another issue that is debated in the scientific literature. Social scientists often refer to the work of sociologist Pierre Bourdieu who differentiated between parental cultural, economic, and social capital (Bourdieu 1986), and many empirical studies have tried to measure these different kinds of capital. Bourdieu's distinction is reflected in the measurement of parental background in reports published by the OECD's PISA consortium, which most often use the so-called ESCS index, a composite index that measures parental *e*conomic, *s*ocial, and *c*ultural *s*tatus (OECD 2016b:339). Table 1 gives an overview of the three components that make up the ESCS index. Parental economic status is measured with the HOMEPOS index, which is based on various questions regarding material possessions in the home. Social status is measured based on the HISEI index that measures the occupational status of the parent with the highest status, and cultural status is measured based on the education of the parent with the highest education, as measured in years.

Table 1: PISA Index of Economic, Social and Cultural Status (ESCS)

	Economic Status	Social Status	Cultural Status
Components, description	*HOMEPOS*, Index of all household possessions	*HISEI*: Highest International Socio-Economic Index	*PARED*: (Highest) Parental Education
Explanation	Index of all household possessions (OECD 2017b). The index consists of a wide range of questions such as: Do you have a room of your own? Do you have a computer you can use for school work? How many cars are there in your home? How many tablets are there in your home? How many books are there in your home?	Index that measures the highest occupational status of the parents. The ranking of the parents' occupational status is based on the International Socio-Economic Index (ISEI), which ranks occupation based on the relationship with income and education (Ganzeboom and Treiman 1996; OECD 2017b)	This is the highest level of parental education as measured in years. It is based on the International Standard Classification of Education (ISCED, see UNESCO Institute for Statistics 2012), which is recoded into the educational level of the parents in years of education (OECD 2017a)

The ESCS index has been criticized because it somewhat arbitrarily gives equal weight to all three dimensions it is comprised of (Carnoy and Rothstein 2013; Lefebvre 2016). While it is not an easy task to decide how important each background dimension should be, giving "possessions at home" the same relevance as parental occupation would need further clarification, which the PISA consortium does not provide. Moreover, the logic behind the weighting and scaling procedure used by OECD statisticians to produce the index scores for each individual student in the respective countries is complex and difficult to follow (Carnoy and Rothstein 2013:41; Rutkowski and Rutkowski 2015:263). Likewise, the reliability and validity of the index that measures economic status, the HOMEPOS index (see Table 1), has been disputed. It is also worth mentioning that Bukodi and Goldthorpe (2013) show that based on a trend analysis of different British cohort studies, the three components of parental background, parental class, and parental education all have independent effects on educational attainment that change in different ways across cohorts.

Another relevant issue related to the measurement of parental background relates to the fact that they are based on students' self-reports (in PISA, these are 15-year-old students). The different student-reported measures have proven to be of disparate quality in a cross-national comparison – with "books at home" being less reliable than children's report of parental occupation (Jerrim and Micklewright 2014). Keeping in mind the limitations of self-reports, we argue that these analyses can still be very valuable in order to show how robust the conclusions are across different self-reported measures of parental background. Furthermore, we have no reason to believe that a potential bias in students' self-reports, such as less reliable reporting of parental education or occupation among less privileged students, would be drastically different between the Nordic countries, i.e. this bias should not affect conclusions regarding between-country differences.

2.3 Parental background measurement in PISA studies

In the following section, we provide a brief overview of previous measurements of parental social background used in the context of reports based on the PISA. Ehmke and Siegel (2005) provide similar overviews of different measurements of socioeconomic background used in TIMSS, PIRLS,[12] and PISA studies conducted from 2000 to 2005 in Germany. They report that since the second cycle of the PISA study (2003), the PISA consortium has exclusively used the ESCS index, while publications based on TIMSS, as well as previous PISA reports, used a variety of categorical measures for parental occupational position, including both education and a continuous measurement of occupation status. The shift towards the use of the ESCS index can be seen in national PISA reports from Nordic countries starting in the year 2003 (see, for example, reports from Norway, Denmark and Sweden: Kjærnsli et al. 2007; Mejding 2004; Skolverket 2007). Interestingly, the use of the ESCS in favor of other established social background measures has been problematized in a number of Nordic reports over the last decade. The authors of the Norwegian national report on

[12] Progress in International Reading Literacy Study.

PISA 2000, for example, mention the OECD index, but they use another index instead. They state that *"you get a different picture of the situation in Norway as regards the relationship between SES and achievements. Interestingly, this is largely a consequence of the use of different measures for SES!"* (Lie et al. 2001:230 [own translation]). Subsequently in 2003, Norway used another index for SES: *"Since PISA no longer contains 'clean' measures for economic or social capital, the focus of the analyses in this chapter will be on the total SES"* (Kjærnsli et al. 2004:202 [own translation]). In 2006, the authors of the national PISA report in Norway made use of the ESCS, but also presented recalculated measures for the years 2000 and 2003 (Kjærnsli et al. 2007).

There is also considerably heterogeneity in the measurement of parental background reports based on the other major international assessment study, the TIMSS, which is also conducted in the Nordic countries. In many national reports, the number of books at home is used as the indicator to measure parental background (for example, Allerup 2008, 2012). Other national reports also draw on parental education and/or occupation. More recently, the "Home Resources for Learning Scale" that was developed by the International Association for the Evaluation of Educational Achievement (IEA) has been used in many TIMSS reports (Bergem, Kaarstein, and Nilsen 2016; Skolverket 2016; Vettenranta et al. 2016).

2.4 Choice of parental background indicators

For the purpose of the current chapter, we utilized PISA's ESCS index as a baseline measurement to estimate the strength of the association between parental social background and academic achievement. Thus we present estimates for each of the three *continuous* parental background that comprise the ESCS (see OECD 2016b): the international socio-economic index of occupational status of the father or mother, whichever is higher (*HISEI*) (Ganzeboom and Treiman 1996), the level of education of the father or mother, whichever is higher, converted into years of schooling (*PARED*), and the index of household possessions (*HOMEPOS*).[13] While the first part of our analysis can be considered a partial replication of Siegle and Ehmke's (2005) study, we go beyond their work by also comparing estimates for one established *categorical*

[13] A detailed overview over the individual items of the HOMEPOS index will be given in the next section.

background variable, the European Socioeconomic Classification (*ESeC*), which is a revised and further developed version of the established EGP class schema (Rose and Harrison 2012). Occupations are grouped into ESeC classes according to the average skill levels of different occupations, employment status (employer, self-employed, or employee), and the degree to which work effort can be monitored (the type of employment contract, see Bihagen, Nermo and Erikson (2012)). The ESeC is constructed based on a classification of occupations (the ISCO08-03 classification) using the coding routine provided by Harrison (2017).[14] We used a *seven-class* version of the schema in order to have a sufficient number of observations across parental class categories in each of the Nordic countries (see Table 7).[15] Overall, the background indicators we chose to focus on for this chapter are all frequently used in both educational research and in other social sciences (Sirin 2005).

Table 2: Analyzed parental background indicators

Name (official acronym)	Level of measurement
PISA index of economic, social, and cultural status (*ESCS*)	Continuous
The international socio-economic index of *occupational status* of the father or mother, whichever is higher (*HISEI*)	Continuous
The level of education of the father or mother, whichever is higher (*PARED*)	Continuous
PISA index of all household possessions (*HOMEPOS*)	Continuous
European Socioeconomic Classification (*ESeC*)	Categorical

2.5 Calculation of the degree of inequality

Apart from the issue of how to measure parental social background, there is also considerable methodological diversity in the way the strength of the association between social background and student test performance is calculated. Publications from the PISA consortium typically use the amount of explained variance (R^2) in an ordinary least squares (OLS) regression model as an indicator for the social inequality gradient; the more variance in student test performance is explained by the measure of

[14] See: http://ekharrison.weebly.com/uploads/2/3/9/9/23996844/eseco83digit.sps, accessed 23.09.2017

[15] For a more detailed description of the ESeC class schema, see: https://www.iser.essex.ac.uk/archives/esec/user-guide/the-european-socio-economic-classification (accessed 28.08.2017).

parental social background, the higher the level of inequality (see for example OECD 2010, p. 55).

- R^2: Amount of explained variance (the coefficient of determination) in an ordinary least squares regression model.
- The value ranges between 0 and 1.
- 0 = the independent variable (social background) does not explain any variance in student learning (i.e., there is no inequality).
- 1 = the independent variable (social background) completely explains all variance in student learning (i.e., there is complete inequality).
- For further reading, see Allison (1999).

Using the amount of explained variance as a single indicator for the measurement of inequality reduces complexity and provides a very intuitive summary statistic.[16] However, this measure also has some drawbacks given that important nuances in the relationship between social background and inequality across the entire distribution of test performance might be missed. More concretely, this has the implication that one assumes that the association between social background and academic performance is the same no matter how well the students perform in the (PISA) test. This implies that social background is just as important among low-performing, average-performing, and high-performing students. However, to obtain a more nuanced understanding of the generation of inequality, it would be helpful to determine whether social background indicators relate in different ways depending on the distribution of student test performance (Costanzo and Desimoni 2017:3).

[16] One should note that the PISA reports with special focus on the relationship between students' social background and educational achievement provide a number of alternative measures, including the proportion of students (boys or girls) performing below a certain threshold (level 2), the proportion of resilient students, and the slope of the socio-economic gradient (OECD 2010, 2012).

2.6 Data and methods

To determine whether the strength of the association between parental background and student performance varies according to the respective background variable that is used, we made use of data from the most recent PISA test in 2015 and focused on the Nordic countries. Because the core domain tested in 2015 was science, the analyses were mostly based on this domain. PISA uses an item-response model to summarize the test performance of 15-year-old students, and it provides multiple plausible values for each tested domain. We applied the OECD's suggested procedure (OECD 2009a) to take into account all plausible values by using the SAS code provided by IEA's IDB Analyzer (Version 4.0.13). We also used the IEA IDB Analyzer in order to implement the OECD's recommendations for using weights to account for PISA's sampling design as well and to correct for student nonresponse (OECD, 2009, p. 57-58). Missing data imputation for the different parental background variables in our analyses was performed using the statistical program R (R Core Team 2017) and applying the statistical package MICE (Buuren & Groothuis-Oudshoorn 2011).[17]

2.7 Different background dimensions and educational inequality across the Nordic countries

2.7.1 Univariate distributions of different parental background indicators

To begin with, we present the distribution of our five different social background variables across the Nordic countries (Tables 4–7). To obtain a more nuanced picture of the distributions of the respective variables, we report the values across seven percentiles in addition to the variables' means and standard deviations.

[17] A more detailed description of our missing data imputation is provided in the appendix.

- Percentiles: Students are sorted in rank order from lowest to highest ESCS score, and the values of the student ESCS scores are then divided into 100 equally sized groups from lowest to highest.
- For example: The 10th percentile for the ESCS indicates that 10% of the ESCS values lie at or below this value and 90% lie above it.
- Quantiles: These are essentially the same as percentiles – only that they are indexed by sample fractions.
- The most common quantiles have special names: We can divide the sample into four equally large groups (quartiles) or five equally large groups (quintiles). The 20th percentile is equivalent to the first quintile.

The OECD standardizes the ESCS index across all participating OECD countries with a mean of 0 and a standard deviation of 1 (OECD 2017b:339–40). In Table 3, we see that all Nordic countries have an average ESCS value greater than zero. This indicates that the mean ESCS value of students from the Nordic countries is substantially higher than the average of students across all participating OECD countries. The mean ESCS scores are lower in Finland and Sweden than in the other Nordic countries.

Table 3: Mean, standard deviation (Std), and percentiles (P) of ESCS across the Nordic countries

Country	N	Mean	Std	P5	P10	P25	P50	P75	P90	P95
DK	7,161	0.58	0.86	−1.02	−0.70	0.00	0.76	1.28	1.53	1.65
FI	5,882	0.25	0.75	−0.96	−0.74	−0.32	0.29	0.87	1.16	1.31
IS	3,371	0.72	0.74	−0.60	−0.28	0.23	0.84	1.28	1.53	1.71
NO	5,456	0.47	0.74	−0.82	−0.56	−0.01	0.58	1.02	1.30	1.46
SV	5,458	0.32	0.82	−1.08	−0.80	−0.25	0.43	0.96	1.27	1.43
All	27,328	0.40	0.81	−0.98	−0.71	−0.18	0.51	1.03	1.35	1.51

Note: Based on PISA 2015; imputed data and weighted estimates.

Furthermore, a comparison of ESCS standard deviations (the average spread of the ESCS values around the mean value) reveals that the dispersion of ESCS seems to be quite similar across the Nordic countries, even if the standard deviation is slightly higher in Denmark than in the rest of the Nordics. Finally, it is interesting to note that the ESCS score for Iceland at both the 5th and 10th ESCS percentile is substantially higher than in the other countries, which suggests that there are fewer very disadvantaged students in Iceland.

Next, we inspected the first of the three components of the ESCS separately, starting with the index measuring highest parental occupation status (HISEI, Table 4). We see that the mean HISEI, standard deviations, and scores across the different percentiles do not vary much across the Nordic countries. However, the mean HISEI score for Finland is a little lower than for the other countries, which is in line with the comparatively lower average value for ESCS (Table 3).

Table 4: Mean, standard deviation (Std) and percentiles (P) of HISEI across the Nordic countries

Country	N	Mean	Std	P5	P10	P25	P50	P75	P90	P95
DK	7,161	53.93	21.89	22.00	25.00	31.00	57.00	75.00	81.00	82.00
FI	5,882	52.03	21.29	24.00	26.00	31.00	55.00	72.00	81.00	82.00
IS	3,371	58.76	20.03	25.00	27.00	43.00	64.00	76.00	81.00	85.00
NO	5,456	59.81	20.41	25.00	27.00	43.00	65.00	77.00	82.00	82.00
SV	5,458	55.41	20.97	25.00	27.00	32.00	59.00	75.00	81.00	84.00
All	27,328	55.36	21.28	24.00	26.00	33.00	59.00	75.00	81.00	82.00

Note: Based on PISA 2015; imputed data and weighted estimates. Based on the parent with the higher ISEI score.

Next we compared the average years of education for the parent with the highest level of education (PARED). In Table 5, we see that the average educational level among parents in the Nordic countries varies to some degree with Denmark and Iceland having higher levels of average parental education than the other three Nordic countries. Given that the years of education stem from conversions of educational degrees into years of education, a comparison of the original variable for educational attainment (highest educational degree of parents) would certainly be more informative in a Nordic context because a number of different degrees are assigned the same number of years of education (see OECD 2017a).

Table 5: Mean, standard deviation (Std), and percentiles (P) of PARED across the Nordic countries

Country	N	Mean	Std	P5	P10	P25	P50	P75	P90	P95
DK	7,161	16.15	2.84	10.00	10.00	16.00	18.00	18.00	18.00	18.00
FI	5,882	15.15	2.01	12.00	12.00	14.50	16.50	16.50	16.50	16.50
IS	3,371	16.40	2.43	10.00	13.00	14.00	18.00	18.00	18.00	18.00
NO	5,456	14.23	1.86	12.00	12.00	12.00	14.00	16.00	16.00	16.00
SV	5,458	14.30	2.31	10.00	11.50	12.00	16.00	16.00	16.00	16.00
All	27,328	14.91	2.43	10.00	12.00	14.00	16.00	16.50	18.00	18.00

Note: Based on PISA 2015; imputed data and weighted estimates. Based on the parent with the greater number of years of education.

Interestingly, results for the PISA index of all household possessions (HOMEPOS) (Table 6) were quite similar to the results for the overall ESCS index – even if Finland seemed to be even more of an outlier given the low mean value on the HOMEPOS index.

Table 6: Mean, standard deviation (Std), and percentiles (P) of HOMEPOS across the Nordic countries

Country	N	Mean	Std	P5	P10	P25	P50	P75	P90	P95
DK	7,161	0.48	0.64	−0.53	−0.26	0.12	0.48	0.87	1.22	1.43
FI	5,882	0.11	0.73	−1.00	−0.69	−0.29	0.09	0.51	0.91	1.27
IS	3,371	0.49	0.73	−0.52	−0.27	0.08	0.43	0.85	1.35	1.74
NO	5,456	0.60	0.83	−0.68	−0.36	0.16	0.62	1.09	1.52	1.82
SV	5,458	0.40	0.91	−1.03	−0.66	−0.13	0.40	0.93	1.44	1.77
All	27,328	0.40	0.82	−0.86	−0.53	−0.07	0.40	0.88	1.34	1.65

Note: Based on PISA 2015; imputed data and weighted estimates.

Finland had the lowest HOMEPOS index score, while the other countries were closer to each other. However, given that the HOMEPOS index has recently been criticized based on different levels of reliability by country and poor cultural comparability (Rutkowski and Rutkowski 2015), one should be cautious when interpreting its scores across countries (we return to the measurement of HOMEPOS in the next section). Furthermore, based on these descriptive comparisons, it seems likely that Nordic differences in the overall ESCS index seem mostly related to differences in the HOMEPOS index and to a lesser extent to the two other components – highest parental job status (HISEI) and highest education (PARED) – which showed less variation between the Nordic countries. Finally, we measured the distribution of the ESeC classes across the Nordic countries (Table 7).

Table 7: Distribution of highest ESeC across the Nordic countires (in %)

	DK	FI	IS	NO	SW	All
Higher grade professionals	36.66	32.53	42.79	45.28	36.64	37.72
Lower grade professionals	15.42	18.44	23.66	22.43	23.40	20.37
Higher-grade white-collar workers	10.07	8.20	7.33	6.89	6.15	7.63
Lower-grade white-collar workers	0.23	1.11	0.28	0.72	0.60	0.65
Higher-grade blue-collar workers	14.69	24.92	11.89	12.68	17.86	17.44
Skilled workers	7.60	8.04	6.30	3.30	4.19	5.60
Semi- and non-skilled workers	4.70	2.38	1.96	1.93	2.44	2.82
Other status*	8.03	2.33	1.48	4.00	4.34	4.63
No Answer	2.59	2.05	4.31	2.77	4.37	3.14
Number of respondents	7,161	5,882	3,371	5,456	5,458	27,328

Note: Based on PISA 2015; imputed data and weighted estimates. Based on the parent in the highest class category. *Other status includes homemaker, student, social welfare recipient, do not know, and vague status.

As mentioned before, we used a seven-category version of the ESeC classification. We chose to keep the observations in the two categories *Other status* and *No answer* in our analyses in order to be able to compare the same number of observations for the ESeC analyses as for the other background indicators.[18] We found that among 15-year-old students in the Nordic countries in the 2015 assessment, the majority had at least one parent in the two highest-ranked categories of professionals. These values were higher than estimates from population data (see for example Juul 2012 for Denmark). Apart from possible reporting error (Jerrim and Micklewright 2014), this can be attributed to the fact that only the parent with the higher class was considered and that there could have been occupational upgrading among the parents of 15-year-old students in the year 2015. Finland is somewhat of an outlier with the lowest relative proportion of parents in the category of professionals, whereas students from Norway and Iceland reported having the highest proportion of parents from the professional categories. Overall, these univariate distributions demonstrate that the class or socioeconomic position of the student body in the Nordic countries is relatively similar to each other even if Finland seems to have a somewhat lower proportion of parents from the higher grade professionals class and that the opposite can be observed in Norway.

[18] Robustness checks where we excluded these groups from the analyses yielded similar results.

2.7.2 *Different indicators for the bivariate association between parental background and students' academic achievement*

We now present our measurements of the strength of the association between our five different background measures (Table 2) and educational achievement using the PISA score in the science domain. In the first step, we calculated the parameter that is most frequently used in the context of the PISA studies, which is the percentage of variance in (science) performance that is explained by the different measures for parental background (R^2, Table 8).

Table 8: Percentage of variance in science performance explained by ESCS, HISEI, PARED, and ESeC

		ESCS	HISEI	PARED	Homepos	ESeC
Country	N	R^2 (adj)[19]	R^2 (adj)	R^2 (adj)	R^2 (adj)	R^2 (adj)
DK	7,161	0.10	0.11	0.04	0.08	0.12
FI	5,882	0.10	0.09	0.04	0.05	0.09
IS	3,371	0.05	0.05	0.03	0.02	0.05
NO	5,456	0.08	0.08	0.01	0.05	0.09
SV	5,458	0.12	0.15	0.05	0.05	0.16
All	27,328	0.09	0.10	0.04	0.04	0.11

Note: Based on PISA 2015; imputed data and weighted estimates.

The second column of Table 8 represents the replication of the estimates provided in the OECD reports (see OECD 2016a, p. 8).[20] We see that the ESCS index explains between 5% (Iceland) and 12% (Sweden) of the variance in science performance across all Nordic countries. Measuring parental background with ESCS thus leads us to conclude that educational inequality is highest in Sweden and lowest in Iceland, while Norway (8%) seems to be in the middle and Denmark and Finland seem to be closer to the Swedish result (10% for both).

In the second step, we inspected the different indicators that ESCS is comprised of in order to determine if the measurement of inequality varies across the different

[19] Because one categorical variable for parental class background, EseC, is entered in the regression model having nine binary variables (with the last one as the reference category), we calculate the adjusted R^2, which corrects for the fact that a greater number of predictors in the model can inflate the ordinary R^2. However, in case of just one predictor variable, the ordinary R^2 and the adjusted R^2 are identical.

[20] Slight, if any, deviations between our and the OECD's report can be attributed to differing sample sizes due to differences in imputation strategy for missing values.

parental background indicators. Starting with highest parental occupational status (HISEI), we can see that between 5% (Iceland) and 15% (Sweden) of the variance in science performance is explained by this variable. The ranking of countries in terms of the relative degree of inequality is the same as the ranking based on ESCS, even if Sweden appears to be slightly more unequal. Furthermore, parental occupational status explains, on average, just as much of the variation in science scores as the ESCS index. There is relatively little variation across the Nordic countries in terms of the explanation of variance in science scores based on the variable for years of parental education (PARED). Only between 1% (Norway) and 5% (Sweden) of the test-score variance is explained by PARED. Finally, the third element of ESCS, the index for household possessions (HOMEPOS) explains between 2% (Iceland) and 8% (Denmark) of the variance in science performance. The low association between parental education and performance is surprising given that parental education is typically one of the main predictors for children's educational success (Bukodi and Goldthorpe 2013). Interestingly, the HOMEPOS index is also the only background indicator where Sweden does not emerge as the most unequal country. Educational inequality is highest in Denmark based on this measure. Finally, the categorical ESeC variable, which measures parental (highest) social class, can explain almost identical proportions of variance in science scores as the continuous HISEI index across all Nordic countries, suggesting that there is indeed internal homogeneity within and external heterogeneity between the different classes, which justifies the use of this classification.

There are at least three takeaways from this first analysis. First, the relative "inequality-ranking" of the Nordic countries is relatively stable across the different parental background indicators with Sweden emerging as the most unequal and Iceland as the most equal country. Second, the index measuring material possessions in the home (HOMEPOS) seems to be an outlier given that Denmark and not Sweden seems to be the country with the most inequality. Third, the indicator variables that are based on parental occupation (the HISEI index and the ESeC classification) can explain just as much variance in test scores as the much more complicated PISA index ESCS.

Given the increased focus on gender differences in educational performance in the Nordic countries (Jóhannesson, Lingard, and Mills 2009; Sortkaer and Reimer 2018), we repeated the analyses above separately for boys and girls (Table 9).

Table 9: Percentage of variance in science performance explained by ESCS, HISEI, PARED, and ESeC for boys and girls separately

			ESCS	HISEI	PARED	HOMEPOS	ESeC
Country		N	Adjusted R^2	Adjusted R^2	Adjusted R^2	Adjusted R^2	Adjusted R^2
DK	Girls	3602	0.11	0.11	0.11	0.11	0.11
	Boys	3559	0.10	0.10	0.10	0.06	0.12
FI	Girls	2863	0.12	0.09	0.09	0.07	0.08
	Boys	3019	0.09	0.09	0.09	0.03	0.09
IS	Girls	1741	0.07	0.05	0.05	0.03	0.05
	Boys	1630	0.03	0.04	0.04	0.00	0.06
NO	Girls	2706	0.08	0.07	0.07	0.06	0.07
	Boys	2750	0.08	0.10	0.10	0.05	0.11
SV	Girls	2731	0.14	0.14	0.14	0.07	0.15
	Boys	2727	0.11	0.15	0.15	0.04	0.17
All	Girls	13643	0.10	0.09	0.09	0.05	0.09
	Boys	13685	0.09	0.10	0.10	0.03	0.12

Note: Based on PISA 2015; imputed data and weighted estimates.

The results of this analysis show that for both the PISA ESCS index and the HOMEPOS index the association between student background and science performance seems to be stronger for girls than for boys. For the background measures that are based on parental occupation, HISEI and especially ESeC, the opposite pattern emerges. There are almost no gender differences for the parental education measure, PARED. In quite a few cases, within-country gender differences seem to be considerably larger than cross-country differences in the strength of the association between social background and science performance. The gender differences in explained variance (R^2) remain largely the same if we inspect test results for other PISA domains (e.g. reading and mathematics, results available on request).

Next, following Bukodi and Goldthorpe (2013), we explored whether the three ESCS dimensions contribute independently of the other two components of the ESCS to the explanation of science scores in a combined multivariate model (Table 10). This analysis can tell us whether the use of the ESCS summary index conceals important differences in the way parental status, education, and wealth (household possessions) affect student performance. Together, the three items explain slightly more variance than the ESCS index alone, and Denmark and Sweden emerge as the most unequal countries. Otherwise, the ranking of Nordic countries in terms of the level of inequality does not change substantially in this model specification. With the exception of the years of education variable (PARED), which does not reach statistical significance in either Denmark or Norway, all three dimensions of the ESCS index contribute independently to the explanation of the variance in science performance across the Nordic countries. As mentioned above, the results for the PARED variable might be related to the fact that educational background in the context of the Nordic (and other European) educational systems with a strong vocational education sector is an inherently categorical variable and should not be used as a continuous (years of education) predictor in the regression model.

Table 10: Individual coefficient estimates for HISEI, PARED, and HOMEPOS and R2 in a combined multivariate model

	Variable	Estimate	R^2 adj.
DK	HISEI	1.02[3]	0.14
	PARED	0.68	
	HOMEPOS	27.56[2]	
FI	HISEI	1.01[2]	0.11
	PARED	4.36[1]	
	HOMEPOS	16.26[2]	
IS	HISEI	0.75[2]	0.06
	PARED	4.00[1]	
	HOMEPOS	8.05[1]	
NO	HISEI	1.15[3]	0.11
	PARED	-0.33	
	HOMEPOS	18.25[2]	
SV	HISEI	1.60[3]	0.16
	PARED	2.50[1]	
	HOMEPOS	11.39[2]	
All	HISEI	1.16[3]	0.11
	PARED	3.09[2]	
	HOMEPOS	12.10[2]	

Note: [1] $p < 0.05$.
[2] $p < 0.01$.
[3] $p < 0.001$. Note: Based on PISA 2015; imputed data and weighted estimates.

In line with the R^2 values from the bivariate models (Table 8), HISEI seems to have the strongest association with science performance in Sweden compared to the other Nordic countries, while the coefficient for HOMEPOS is largest in Denmark. Furthermore, from additional stepwise models (not reported), we can conclude that the ISEI index for occupational prestige seems to make a stronger contribution to the explanation of the variance in science performance than the other two components of ESCS across all countries.

2.7.3 The index for household possessions (HOMEPOS) in the Nordic countries

Given the irregular pattern of results for HOMEPOS in Table 8 and Table 9 and the fact that the reliability and cross-country comparability of this index has been the subject of criticism (Rutkowski and Rutkowski 2015), we took a closer look at the index for the Nordic countries. HOMEPOS is a summary index that includes a number of measures for family wealth possessions (WEALTH), cultural possessions (CULTPOSS), home educational resources (HEDRES), ICT[21] resources (ICTRES), and some additional items (OECD 2016b). On the subscale for wealth, each country adds up to three country-specific items. For PISA 2015, the Nordic countries added the following items, respectively (Table 11):

Table 11: The three country-specific questions for the PISA HOMEPOS index for each Nordic country (PISA 2015)

Country	Item
Denmark 1	Musical instrument (e.g. piano, guitar, violin)
Denmark 2	Smart TV
Finland 1	Laptop
Finland 2	Home alarm system
Iceland 1	Security guard or home security system
Iceland 2	Hot tub
Iceland 3	Home help (not used for the Nordic index)
Norway 1	Tablets, e.g. iPad
Norway 2	iPhone
Sweden 1	Piano
Sweden 2	Whirlpool
Sweden 3	Espresso machine (not used for the new Nordic index)

There are some redundancies between the country-specific items and the core items that are answered by all students. For example "iPhone" in Norway or "musical instrument" (piano) in Denmark and Sweden are country-specific questions, while all students are asked about the presence of cell phones with Internet access and musical instruments in the core questions. Wealth in homes with musical instruments or smartphones will thus be exaggerated because they are counted twice. Another potential problem with HOMEPOS is ceiling effects. Ceiling effects indicate that all

[21] Information and Communication Technology.

respondents score near or at the top of the scale so that all students have similar high scores and there is little separation between them.

Furthermore, due to the OECD-based standardization, the variance in HOMEPOS might be relatively small in the Nordics. Thus we computed two alternative (Nordic) HOMEPOS indices. Following the OECD's procedure (OECD 2016b), we computed a HOMEPOS index that is largely identical to the OECD index but is standardized at the level of the Nordic countries. The other alternative Nordic HOMEPOS index excludes the country-specific items and is also standardized at the mean for the Nordic countries.[22] Subsequently, we compared the relative proportion of explained variance with the OECD's original index and the two Nordic HOMEPOS indices (Table 12). Excluding the country-specific items did not reduce the proportion of explained variance in science scores. In Iceland, the Nordic HOMEPOS even explained a little (1%) more than the original OECD index. Furthermore, just standardizing HOMEPOS based on the Nordic and not the OECD average seemed to increase its explanatory power. On this basis, it might be advisable that the Nordic countries in close cooperation develop a common set of country-specific items to prevent possible ceiling effects and to increase comparability among the Nordic countries.

Table 12: Explained variance in science performance with two different versions of HOMEPOS

Country	N	Nordic (std) HOMEPOS with country-specific items	Nordic (std) HOMEPOS without country-specific items	OECD (std) HOMEPOS
DK	7,161	0.08	0.09	0.08
FI	5,882	0.05	0.06	0.05
IS	3,371	0.03	0.04	0.02
NO	5,456	0.06	0.06	0.05
SV	5,458	0.07	0.07	0.05
All	27,328	0.06	0.06	0.04

Note: Based on PISA 2015; imputed data and weighted estimates.

[22] We should mention that standardization, e.g. transforming the scale so that 0 represents the average, does not remove ceiling effects per se. However, standardizing the HOMEPOS so that 0 represents the Nordic and not the OECD average has the advantage that values near the top of the distribution are not clustered together as closely.

2.8 A different approach for measuring inequality: Quantile regression

We now turn to the second part of the chapter in which we explore whether an alternative approach for estimating the strength of the association between social background and student test performance, quantile regression, will lead to different conclusions. In the following, we briefly explain what kind of information quantile regression can provide compared to or in addition to results that are obtained through the conventional OLS regression methods that are used in the OECD's PISA reports. To reduce complexity, all analyses in this section are based only on the PISA ESCS index.

2.8.1 Estimates from OLS models with and without control variables

In the conventional OLS regression model, a coefficient estimate [b] indicates the expected change in the dependent variable if the independent variable increases by one unit on the respective parental background dimension. In Table 13, we show the OLS coefficient estimates for the association between ESCS and science performance. We see that, for example, in Denmark a one unit (standard deviation) increase in ESCS will lead to an estimated increase in science performance of 33.8 points. It is assumed that this association is the same across the entire performance distribution of students.

Table 13: Coefficient estimates for ESCS on science score, including confidence intervals

Country	N	Coefficient estimate	R^2 adj.
DK	7,161	33.84 (30.5; 37.2)	0.10
FI	5,882	40.41 (35.9; 44.9)	0.10
IS	3,371	26.94 (22.5; 31.4)	0.05
NO	5,456	37.34 (33.0; 41.7)	0.08
SV	5,458	43.91 (39.7; 48.1)	0.12
All	27,328	36.84 (34.7; 39.0)	0.09

Note: Based on PISA 2015; imputed data and weighted estimates.

2.8.2 Estimates from quantile regression

By applying OLS regression, we assume that the association between parental background (ESCS, Table 13) and science scores is the same for students who performed poorly, average, or well on the science test. However, it is not unlikely that the strength of the association between parental background and sciences scores differs depending on how well the students did on the test. It is possible that the association will be stronger among high performers than among average or low performers (or the other way around). If this is true, the OLS regression will over- and underestimate the association between ESCS and PISA score for some students depending on how well they did on the test. By applying quantile regression, we can check whether the OLS coefficients, which refer to the mean, are an appropriate summary of the relationship between parental background across the entire distribution of science performance.[23] Simply put, using quantile regression has the same benefits as examining the median and/or the 25th and 75th quantile of a distribution compared to only looking at the mean of a distribution – which is the case in OLS regression analysis. Just as the inspection of different quantiles gives a more complete picture of a distribution compared to only looking at the mean, estimates from quantile regression give a more complete picture of the association between two variables than the single OLS coefficient (Koenker and Hallock 2001).

We report our results in Figure 1 using the so-called "quantile process plot" for each of the Nordic countries. In every country-specific plot, the x-axis indicates the quantile of the students' science scores, while the y-axis shows the coefficient estimates for ESCS. For each plot, every 5th quantile is estimated from the 5th to the 95th quantile.[24] The standard OLS coefficient for ESCS and PISA score within the given country is also plotted as a straight horizontal line and serves as a point of reference to evaluate to what extent the coefficients for the different quintiles deviate from the OLS

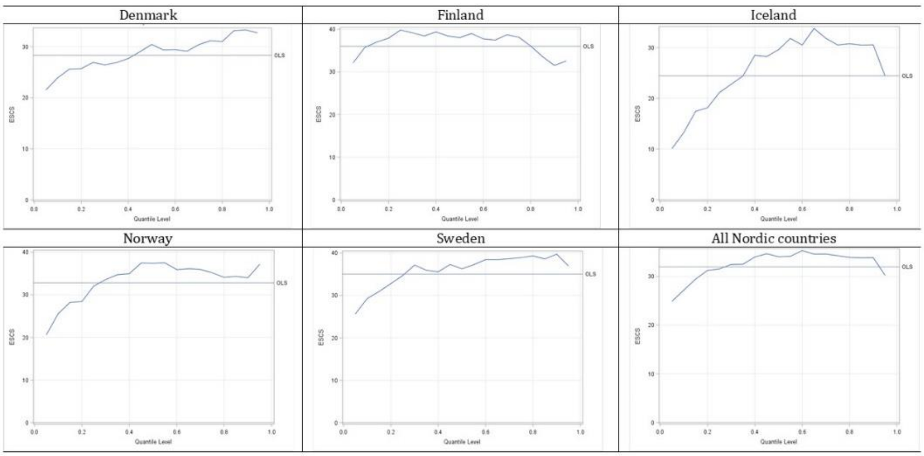
[23] Lefebvre (2016) as well as Constanzo and Desimoni (2017) have performed a similar analyses for Canada and for Italy, respectively.

[24] The estimation of quantile regression does not allow for the implementation of the OECD's suggested technique to work with multiple plausible values. As a result, we only use one plausible value (PV1SCIE). As a robustness check, all analyses were conducted for the other plausible values (2–10). The shape of the association (available on request) looked almost identical for each plausible value. Furthermore, based on simulations conducted by the OECD, one can assume that for samples with more than 6,000 cases the results based on one plausible value should not deviate substantially from analyses that properly take all plausible values into account (OECD 2009b:44).

coefficient.[25] All analyses were conducted with the control variables of student gender, age, grade, and immigration status.[26]

We start our discussion with the quantile regression results for *Denmark*. The blue line for the results of the quantile regressions shows that the coefficient for ESCS varies from 22 to 33 points. The ESCS estimate slightly increases between the 15th quantile up to the 90th where the effect estimate peaks at 33 points. From the 15th quantile the effect decreases more rapidly. The lowest ESCS coefficient estimate can be observed at the 5th quantile (22 ESCS points).

Figure 1: Quantile process plots for ESCS



Note: Based on one plausible value (PV1SCIE), weighted and imputed data.

The quantile regression results for *Finland* show lower coefficient estimates at both the high and low end of the test distribution, but a higher constant coefficient from about the 25th quantile up to the 65th quantile. The sizes of the estimates are larger than in the Danish case ranging from 32 to 40. Again, we observe that the association between

---

[25] To reduce complexity, we do not discuss whether quantile regression results are statistically different from the OLS results (but these results are available on request).

[26] We perform so-called conditional quantile regression because we control for additional covariates and interpret coefficients just as ordinary OLS coefficients. While this might not be unproblematic (Borah and Basu 2013; Killewald and Bearak 2014), we prefer this more intuitive interpretation. Furthermore, estimates from quantile regression models without additional covariates were essentially the same (results available on request).

ESCS and the PISA score is dependent on the relative placement of students in the test distribution. In the Finnish case, the OLS overestimates the association between ESCS and PISA score for both the lowest and highest-scoring students while slightly underestimating the association in the middle of the distribution. Next, the quantile regression results for *Iceland* show a steep increase in the size of the ESCS coefficient from the lower quantiles up to the 55th quantile. The coefficient then stabilizes before decreasing slightly from the 70th percentile up to the 90th percentile. In the Icelandic case, the estimates have a relatively large range from 10 at the lowest to 34 at the highest. For Iceland, the quantile regression results suggest that the average OLS coefficient is not a precise summary of the relationship between ESCS and science performance. Given that Iceland emerged as the most equal country in the previous analyses, this finding should be kept in mind. The results for *Norway* and *Sweden* are very similar with a smaller coefficient estimate at the lower end of the performance distribution. In Norway the ESCS coefficient increases up until the 45th quantile, and then stabilizes. In Sweden it stabilizes around the 30th quantile. Finally, looking at the combined plot for *all the Nordic countries*, we see that the estimates look similar to those of Sweden, Norway, Iceland, and Denmark. The ESCS coefficient starts decreasing slowly from about the 40th quantile, with a relatively steep drop at the 20th quantile. From the 40th quantile up to the 85th quantile, the estimate is somewhat constant, with a small decline from the 85th to the 95th quantile. The plot also indicates that the association between the ESCS coefficient and science score might both be under- and overestimated for different students depending on their science score.

To summarize, with the exception of Finland, the pattern of an inverted u-shape can be observed in the Nordic countries. We see that the association between ESCS and PISA score is weaker for the students who are at the lowest end of the PISA score within their respective countries. When we move up the test distribution, the association seems to stabilize. Comparing the quantile regression with the OLS estimates, we see that the latter overestimates the association between ESCS and PISA score for the students who performed poorly on the test but underestimates the association for students who are not in the lowest end of the quantiles. Interestingly, this pattern of results was also observed in another European country, Italy (Costanzo and Desimoni 2017), but not in Canada (Lefebvre 2016).

The lower effect of ESCS at the bottom quantiles found in most Nordic countries suggests that a few of the students from middle to high-ESCS homes do not perform as well as one would expect given their social position. These results might indicate that when looking at the association between parental social background and school performance, the estimates from the OLS regression might not be suitable. The association between parental social background and student performance is potentially significantly weaker at lower quantiles. Possibly, parental resources at home might not be as much of an advantage for students who fall below a certain threshold in terms of their academic abilities. Given that policy efforts to increase learning are often directed at lower-performing students, these findings should be kept in mind.

2.9 Conclusions and recommendations for policy and research

This chapter has provided a detailed analysis of two important methodological aspects related to the estimation of the strength of the association between indicators of parental social background and student performance in international large-scale assessment studies in the Nordic countries. Based on data from the PISA 2015 study, we first explored whether the degree of inequality is substantially different across the Nordic countries if different dimensions of parental background are considered. We used PISA's ESCS index as the baseline measurement and compared the association between ESCS, the three dimensions comprising ESCS (highest parental occupational status (HISEI), highest parental years of education (PARED), and PISA's index of household possessions (HOMEPOS)) as well as a categorical variable measuring parental class (ESeC). The comparisons of models measuring the variance in student test performance that is explained showed that the relative ranking of the Nordic countries remained relatively stable across the different indicators. Based on the 2015 PISA data, Sweden emerged as the most unequal and Iceland as the most equal country across the different indicators. However, results based on the index used by the OECD to measure wealth or a family's economic resources, the HOMEPOS index, did not quite fit this pattern. Another notable finding from the first part of this chapter was that the association between parental background and test achievement seems to be somewhat higher for girls than for boys.

The second part of this chapter was devoted to showing whether average (mean-based) estimates from the OLS regression model are an appropriate representation of the association between student social background and student achievement. To this end, we estimated quantile regression models and compared the quantile regression with the average standard OLS regression estimates. Overall, we observed a similar pattern across all Nordic countries except Finland, where the OLS coefficients overestimated the effect/association between ESCS and PISA score for the students who performed poorly on the PISA test and underestimated the association for students in the higher quantiles of the achievement distribution. We interpret this finding to mean that low-performing students are potentially confronted with a number of other barriers that are not related to the socioeconomic status of their parents. Some of these students might, for example, have a diagnosis or other stressors such as parental divorce or sickness that might potentially play a role in their achievement on the test.

What can be learned from these findings? First, the results show that researchers and political advisors need to be very thoughtful when choosing the indicator variables for measuring parental background because the conclusions can in fact change if different indicators are used. The index used in the PISA studies, the ESCS, is a relative complex construction, and the reliability of one key component of ESCS, the index for household possessions (HOMEPOS), across nations has been questioned (Rutkowski and Rutkowski 2015). From a practical perspective, it is also not easy to identify "low-ESCS groups" in reality because multiple combinations of the indicator variables comprising the index could lead to a low ESCS status. Furthermore, inequality measurement for the different components of ESCS might vary across countries – which is another problem of a composite index. In this respect, the simpler index for highest parental occupational status (HISEI) or the social class grouping (ESeC) might be a viable straightforward alternative. Our analyses showed that these two indexes explain almost or just as much variation in science performance as the ESCS index. Additionally, the Nordic countries should consider regularly taking advantage of the availability of parental background information that can be extracted from the administrative registers and should compare the results of register-based information with self-reported parental background information to assess the size and direction of potential biases arising from the students' answers (see for example Engzell and Jonsson 2015). Another policy-relevant finding of our work refers to the OECD's indicator for household possessions, HOMEPOS, which is one of the key components for measuring ESCS. This index might to some degree underestimate the association between material wealth and student performance due to the standardization at the OECD mean as well redundant country-specific items. The latter might be helped by a concerted effort to harmonize country-specific wealth items in the Nordic countries.

In the second part of our analysis, we introduced quantile regression analyses as an alternative to the established OLS regression method to calculate the degree of inequality in test performance. Our findings demonstrated that the relationship between parental background and student achievement changes across the distribution of test performance. Targeted efforts directed at low-performing students should be aware of this limitation of (average-based) statistical analyses. Students who performed relatively poorly might face a number of challenges such as mental health problems, sickness, or other problems in the family that cannot be easily remedied even if parents have sufficient economic, social, and cultural resources.

2.10 References

Allerup, Peter. 2008. *Danske 4. Klasseelever I TIMSS 2007*. Fjerritslev: Forlag1.dk.

Allerup, Peter. 2012. *Danske 4. Klasseelever I TIMSS 2011*. Fjerritslev: Forlag1.dk.

Allison, Paul D. 1999. *Multiple Regression: A Primer*. Newbury Park: Pine Forge Press.

Bergem, Ole Kristian, Hege Kaarstein, and Trude Nilsen. 2016. *Vi Kan Lykkes I Realfag: Resultater Og Analyser Fra TIMSS 2015*. Oslo: Universitetsforlaget.

Bihagen, Erik, Magnus Nermo, and Robert Erikson. 2012. "Social Class and Employment Relations: Comparisons between the ESeC and EGP Class Schemas Using European Data." Pp. 89–113 in *Social Class in Europe: An introduction to the European Socio-economic Classification*, edited by D. Rose and D. Harrison. Abingdon: Routledge.

Borah, Bijan J. and Anirban Basu. 2013. "Highlighting Differences between Conditional and Unconditional Quantile Regression Approaches through an Application to Asess Medication Adherence." *Health Economics* 22(9):1052–70.

Bourdieu, Pierre. 1986. "The Forms of Capital." Pp. 241–58 in *Handbook of Theory and Research for the Sociology of Education*, edited by J. Richardson. New York: Greenwood Press.

Bukodi, Erzsébet and John H. Goldthorpe. 2013. "Decomposing 'Social Origins': The Effects of Parents' Class, Status, and Education on the Educational Attainment of Their Children." *European Sociological Review* 29(5):1024–39.

Buuren, Stef van and Karin Groothuis-Oudshoorn. 2011. "*Mice* : Multivariate Imputation by Chained Equations in *R*." *Journal of Statistical Software* 45(3).

Carnoy, Marting and Richard Rothstein. 2013. *What Do International Tests Really Show about U.S. Student Performance*. Washington D.C.: Economic Policy Institute.

Costanzo, Antonella and Marta Desimoni. 2017. "Beyond the Mean Estimate: A Quantile Regression Analysis of Inequalities in Educational Outcomes Using INVALSI Survey Data." *Large-Scale Assessments in Education* 5(1):14.

Ehmke, Timo and Thilo Siegle. 2005. "ISEI, ISCED, HOMEPOS, ESCS." *Zeitschrift Für Erziehungswissenschaft* 8(4):521–39.

Enders, Craig C. 2010. *Applied Missing Data Analysis*. New York: Guilford Press.

Engzell, P. and J. O. Jonsson. 2015. "Estimating Social and Ethnic Inequality in School Surveys: Biases from Child Misreporting and Parent Nonresponse." *European Sociological Review* 1–14.

Erikson, Robert, John H. Goldthorpe, and Lucienne Portocarero. 1979. "Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden." *The British Journal of Sociology* 30(4):415–41.

Erikson, Robert and Jan O. Jonsson. 1996. "The Swedish Context: Educational Reform and Long-Term Change in Educational Inequality." Pp. 65–93 in *Can Education Be Equalized? The Swedish Case in Comparative Perspective*, edited by R. Erikson and J. O. Jonsson. Boulder: Westview Press.

Ganzeboom, Harry B. G., Paul M. De Graaf, and Donald J. Treiman. 1992. "A Standard International Socio-Economic Index of Occupational Status." *Social Science Research* 21(1):1–56.

Ganzeboom, Harry B. G. and Donald J. Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research* 25(3):201–39.

Jerrim, J. and J. Micklewright. 2014. "Socio-Economic Gradients in Children's Cognitive Skills: Are Cross-Country Comparisons Robust to Who Reports Family Background?" *European Sociological Review* 30(6):766–81.

Jóhannesson, Ingólfur Ásgeir, Bob Lingard, and Martin Mills. 2009. "Possibilities in the Boy Turn? Comparative Lessons from Australia and Iceland." *Scandinavian Journal of Educational Research* 53(4):309–25.

Jonsson, Jan O., Matthew Di Carlo, Mary C. Brinton, David B. Grusky, and Reinhard Pollak. 2009. "Microclass Mobility: Social Reproduction in Four Countries." *AJS; American Journal of Sociology* 114(4):977–1036.

Juul, Jonas Schytz. 2012. *Inddeling Af Befolkningen I Sociale Klasser*. København: Arbejderbevægelsens Erhversråd.

Killewald, Alexandra and Jonathan Bearak. 2014. "Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression." *American Sociological Review* 79(2):350–57.

Kjærnsli, Marit, Svein Lie, Rolf V. Olsen, Astrid Roe, and Are Turmo. 2004. *Rett Spor Eller Ville Veijer? Norske Elevers Prestasjoner I Matematikk, Naturfag Og Lesing I PISA 2003*. Oslo: Universitetsforlaget.

Kjærnsli, Marit, Svein Lie, Rolf Vegar Olsen, and Astrid Roe. 2007. *Tid for Tunge Løft: Norske Elevers Kompetanse I Naturfag, Lesing Og Matematikk I PISA 2006*. Oslo: Universitetsforlaget.

Koenker, Roger and Kevin F. Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives* 143–56.

Kommunernes Landsforening. 2013. "Fakta Om Folkeskolereformen." (august):1–2.

Lefebvre, Pierre. 2016. *Socioeconomic Gradient Literacy and Numeracy Skills*. Québec: Research Group on Human Capital and Department of Economics School of Management, Université du Québec à Montréal.

Lie, Svein, Marit Kjærnsli, Astrid Roe, and Are Turmo. 2001. *Godt Rustet for Framtida? Norske 15-Åringers Kompetanse I Lesing Og Realfag I et Internasjonalt Perspektiv*. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Marks, G. N. 2006. “Are between- and within-School Differences in Student Performance Largely due to Socio-Economic Background? Evidence from 30 Countries.” *Educational Research* 48(1):21–40.

Mejding, Jan. 2004. *PISA 2003: Danske Unge I En International Sammenligning*. København: Danmarks Pædagogiske Universitets Forlag.

OECD. 2009a. *PISA Data Analysis Manual: SAS, Second Edition*. Organisation for Economic Co-operation and Development (OECD).

OECD. 2009b. *PISA Data Analysis Manual: SPSS*.

OECD. 2010. *PISA 2009 Results : Overcoming Social Background – Equity in Learning Opportunities and Outcomes (Volume II)*. Paris: OECD.

OECD. 2012. *Equity and Quality in Education – Supporting Disadvantaged Students and Schools*. Paris: OECD.

OECD. 2016a. *PISA 2015 Results in Focus*. Paris: OECD.

OECD. 2016b. *PISA Technical Report 2015: Chapter 16 Scaling Procedures and Construct Validation of Context Questionnaire Data*. Paris: OECD.

OECD. 2017a. “Annex D – Mapping of ISCED to Years.” P. 435 in *PISA 2015 Technical Report*. Paris: OECD.

OECD. 2017b. “Chapter 16. Scaling Procedures and Construct Validation of Context Questionnaire Data.” Pp. 290–344 in *PISA 2015 Technical Report*. Paris: OECD.

Olsen, Rolf Vegar, Therese N. Hopfenbeck, and Sølvi Lillejord. 2013. “Elevenes Læringssituasjon Etter Kunnskapsløftet.” *Norsk Pedagogisk Tidsskrift2* 97(6):355–69.

Rasmussen, Jens, Claus Holm, and Andreas Rasch-Christensen. 2015. *Folkeskolen – Efter Reformen*. København: Hans Reitzels Forlag.

Rose, David and Eric Harrison. 2012. *Social Class in Europe : An Introduction to the European Socio-Economic Classification*. London: Routledge.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Rutkowski, David and Leslie Rutkowski. 2015. “Measuring Socioeconomic Background in PISA: One Size Might Not Fit All.” *Research in Comparative and International Education* 8(3):259–78.

Sirin, Selcuk R. 2005. “Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research.” *Review of Educational Research* 75(3):417–53.

Skolverket. 2007. *PISA 2006 – 15-Åringars Förmåga Att Förstå, Tolka Och Reflektera – Naturvetenskap, Matematik Och Läsförståelse*. Stockholm: Skoleverket.

Skolverket. 2016. *TIMSS 2015: Svenska Grundskoleelevers Kunskaper I Matematik Och Naturvetenskap I Ett Internationellt Perspektiv*. Stockholm: Skolverket.

Sortkaer, Bent and David Reimer. 2018. “Classroom Disciplinary Climate of Schools and Gender: – Evidence from the Nordic Countries.” *School Effectiveness and School Improvement*, DOI: 10.1080/09243453.2018.1460382.

UNESCO Institute for Statistics. 2012. *International Standard Classification of Education: ISCED 2011*. Montreal: UNESCO Institute for Statistics.

Vettenranta, Jouni, Jenna Hiltunen, Kari Nissinen, Eija Puhakka, and Juhani Rautopuro. 2016. *Lapsuudesta Eväät Oppimiseen: Neljännen Luokan Oppilaiden Matematiikan Ja Luonnontieteiden Osaaminen: Kansainvälinen TIMSS-Tutkimus Suomessa*. Jyväskyla: Jyväskylän yliopistopaino.

White, Karl R. 1982. "The Relation between Socioeconomic Status and Academic Achievement." *Psychological Bulletin* 91(3):461–81.

White, Sammis B., Peter D. Reynolds, Marc M. Thomas, and Nance J. Gitzlaff. 1993. "Socioeconomic Status and Achievement Revisited." *Urban Education* 28(3):328–43.

2.11 Appendix: Missing imputation

Missing data imputation was performed with the R package mice (Buuren and Groothuis-Oudshoorn 2011). This statistical package is intended for multiple imputations of multivariate data with incomplete data. MICE performs sequential regression imputation, which can handle both categorical and continuous variables. Instead of imputing multiple datasets (Rubin 1987) for the incomplete variables, we used one imputed dataset with 20 iterations in order for the statistical analysis to be coherent with the use of multiple plausible values. Each variable in this iterated conditional model (Enders 2010:275) is imputed variable-by-variable. We apply different types of imputation for each variable. We apply logistic regression combined with predictive mean matching instead of polytomous logistic regression in order to meet the computational limits. For the predictor matrix, the variables in the analysis were used.

3. The relation of science teachers' quality and instruction to student motivation and achievement in the 4th and 8th grade: A Nordic perspective

By Trude Nilsen,[27] Ronny Scherer,[28] and Sigrid Blömeke

3.1 Abstract

Teachers and their instruction are the lifeblood of education and are vital to student outcome. However, little research has been conducted in this field in the Nordic countries, and fewer still has investigated these effects using student motivation in science as the outcome. We address this gap and investigate relations between aspects of teacher quality, the quality of their instruction, and student motivation and achievement in science. Findings from TIMSS 2015 data from the Nordic countries (grades 4 and 8) indicate that teachers' instructional quality had a positive and significant relation to student achievement and motivation in both grades in most countries. Moreover, the type of teacher competences reflecting more general pedagogical aspects (i.e., collaboration, self-efficacy in pedagogical content knowledge, and teacher motivation) had positive and significant relations to student outcomes in both grades, while teachers' formal qualifications seemed to be of more importance in grade 8 than in grade 4. The implications for policy and practice are discussed.

27 Department of Teacher Education and School Research, Faculty of Educational Sciences, University of Oslo, Norway
28 Centre for Educational Measurement (CEMO), Faculty of Educational Sciences, University of Oslo, Norway

3.2 Introduction

Educational policy-makers and researchers have been concerned about students' motivation and recruitment to subjects such as science, technology, engineering, and mathematics (STEM) over the last two decades (OECD, 2006; Osborne & Dillon, 2008). This concern is increasing as societies are becoming more and more digitalized and as challenges related to sustainable development remain largely unsolved. The need for competence in STEM is therefore growing, and it is of utmost importance to motivate students for these subjects in school.

Existing evidence in subjects outside the STEM domains suggests that two pathways may lead to long-term achievement in and recruitment to science—a cognitive pathway and a commitment pathway (Snow, 1994). Indeed, student motivation in science and related subjects determines their long-term achievement in—and later recruitment to—this field (Bøe, 2012; Eccles & Wigfield, 2002; Osborne & Dillon, 2008; Schiepe-Tiska, Roczen, Müller, Prenzel, & Osborne, 2016).

Considering this issue, the key question is: How can student motivation and achievement in science be fostered effectively? In other words, what are the key factors that may determine science motivation and achievement? One possible answer to this question concerns the key drivers of education—teachers. Teachers are the group with the closest proximity to students during their everyday instruction, and the literature generally supports that both teacher quality (e.g., their educational level, specialization, and self-efficacy) and the quality of their instruction are related to cognitive and motivational outcomes in education (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Neumann, Kauertz, & Fischer, 2012; Pintrich, 2003; Scherer & Nilsen, 2016; Seidel & Shavelson, 2007). However, rarely do researchers find *direct relations* between teacher quality and student outcomes—although there is evidence for *indirect relations* via instruction (Blömeke, Olsen, & Suhl, 2016; Goe, 2007). Moreover, most international studies examining these indirect relations between teacher quality and student outcomes focus on cognitive outcomes such as student achievement; student motivation often plays a minor role (Goe, 2007). For the Nordic countries, information about how to support student motivation to learn science through teachers and teaching is particularly scarce and is thus urgently needed.

The present study addresses this need by asking the following questions:

- What are the relations between teachers' instructional quality and their students' achievement and motivation in science in the Nordic countries?
- What aspects of teacher quality are directly related to their instructional quality as well as to their students' achievement and motivation in the Nordic countries?
- What aspects of teacher quality are indirectly (via instructional quality) related to student achievement and motivation?
- Are these relations between teacher quality, instructional quality, and student achievement or motivation the same for younger students (in grade 4) and older students (in grade 8)?

Answering these questions will help us to understand what aspects of teacher quality are important and how to design instruction so that teachers may be able to support student motivation and achievement in science in grade 4 and grade 8.

3.3 Conceptual Framework

This section presents the theoretical foundations of student achievement and motivation as the two major outcomes of education, and this is followed by the theoretical foundations of teacher quality and instructional quality. The outline concludes with a brief review of existing research on the relations between teacher quality, instructional quality, and student outcomes.

3.3.1 Student outcomes

Student motivation

According to social-cognitive theory, motivation can be defined as an internal state that arouses, directs, and sustains goal-oriented behavior (Bandura, 1997; Eccles & Wigfield, 2002). The two core dimensions of motivation are *intrinsic* and *extrinsic motivation.* Intrinsic motivation occurs when students enjoy a subject because they are interested in it for itself, while extrinsic motivation occurs when students engage in a subject to accomplish external objectives, such as earning good grades (Deci & Ryan, 1985;

Eccles & Wigfield, 2002). Because extrinsic motivation was not measured in grade 4 in the Trends in Mathematics and Science Study (TIMSS), we focus only on intrinsic motivation as the crucial motivational counterpart to achievement.

Student achievement

Science achievement covers a broad range of areas, including knowledge in biology, chemistry, earth science, and physics as well as certain related skills such as managing experiments and observations, testing hypotheses, and explaining natural phenomena. Science achievement thus represents a complex construct.

The state of research shows that intrinsic motivation is positively related to student achievement (Pintrich, 2003; Schiepe-Tiska et al., 2016). This finding has been established by a large body of studies in the Nordic countries (Cerasoli, Nicklin, & Ford, 2014; OECD, 2016), supporting once more the relevance of student motivation.

3.3.2 Teacher quality and their instruction

Teacher quality

Teacher quality can be divided into two aspects: teachers' formal qualifications and their competence (e.g., Goe, 2007; Kuger, Klieme, Jude, & Kaplan, 2016). Teachers' formal qualifications typically include the level of their teacher education, their specialization, and their participation in professional development. Teachers' competence includes their professional knowledge and cognitive skills (Blömeke, Gustafsson, & Shavelson, 2015) as well as their beliefs, self-efficacy, and collaboration skills (Goe, 2007; Kuger et al., 2016; Seidel & Shavelson, 2007).

Teachers' professional knowledge comprises many aspects, two of which have gained importance recently—Content Knowledge (e.g., about electricity in physics) and pedagogical content knowledge (e.g., how to use inquiry in science; Abell, 2008; Baumert et al., 2010). It is, however, difficult to measure teacher knowledge. Self-efficacy in content knowledge and pedagogical content knowledge—that is, the extent to which teachers trust their knowledge in these two areas—is more easily accessible but is only an indirect indicator of teachers' professional knowledge (Blömeke, Olsen et al., 2016; Tatto et al., 2012).

Teacher collaboration skills have also been found to affect student achievement according to, for instance, a review of previous research by Goddard, Goddard, and Tschannen-Moran (2007). This review also included an empirical analysis of 4th graders, and their conclusion was that higher levels of teacher collaboration in working for school improvement were associated with higher student achievement in mathematics and reading.

The different aspects of teacher quality vary in their importance for student outcomes across countries (Blömeke, Olsen et al., 2016), and what is important for student outcomes in one country might not be as important in another. Furthermore, teacher quality is known to be indirectly linked to student outcomes via their instructional quality (Baumert et al., 2010; Nilsen & Gustafsson, 2016). The direct effect of teacher quality on student outcomes is often less informative. For instance, if the direct relation between teachers' level of teacher education and student achievement is negative, this might simply reflect that highly educated teachers are more often assigned to low-achieving students than to high-achieving students. Such compensatory approaches have been revealed for several educational systems, for example, in Norway (Gustafsson, Nilsen, & Hansen, 2016).

Instructional quality

Instructional quality comprises instructional practices that research has found to be important for student outcomes (Kuger et al., 2016; Neumann et al., 2012; Nilsen & Gustafsson, 2016). The most common European understandings of instructional quality are inspired by studies that characterize effective practices according to *cognitive activation*, *teacher support*, and *classroom management* (Baumert et al., 2010; Blömeke, Busse, Kaiser, König, & Suhl, 2016; Klieme, Pauli, & Reusser, 2009; Kunter et al., 2013). Although these studies use mathematics as the subject domain, similar aspects of instructional quality have been found in the domain of science (Neumann et al., 2012). The core practices of good instruction are:

- *Cognitive activation* comprises instructional activities in which students are challenged cognitively, for example, through evaluating, integrating, and applying knowledge in the context of problem solving (Baumert et al., 2010; Hiebert & Grouws, 2007). For instance, the teacher might use inquiry practices such as having students interpret data from experiments in science or asking students to engage in challenging tasks (Minner, Levy, & Century, 2010).
- *Teacher support* includes practices such as listening to and respecting students' ideas and questions, showing an interest in every student's learning, providing feedback, and adapting practices to the individual (Blömeke, Olsen et al., 2016).

Teacher support further reflects clear and comprehensive instruction and includes practices of setting clear learning goals, providing a summary at the end of the lesson, and connecting new and old topics (Cohen & Grossman, 2016; Nilsen & Gustafsson, 2016; Raudenbush, 2008).

- *Classroom management* reflects the time spent on task and disciplinary practices such as reducing levels of noise and disorder in the classroom (Kyriakides, Creemers, & Antoniou, 2009; van Tartwijk & Hammerness, 2011).

Although these practices form the core of instructional quality, their conceptualizations may vary across countries and frameworks. The specific conceptualizations outlined here largely overlap with those outlined in studies of educational effectiveness (e.g., Creemers & Kyriakides, 2008; Kane & Staiger, 2012; Pianta, Hamre, & Allen, 2012).

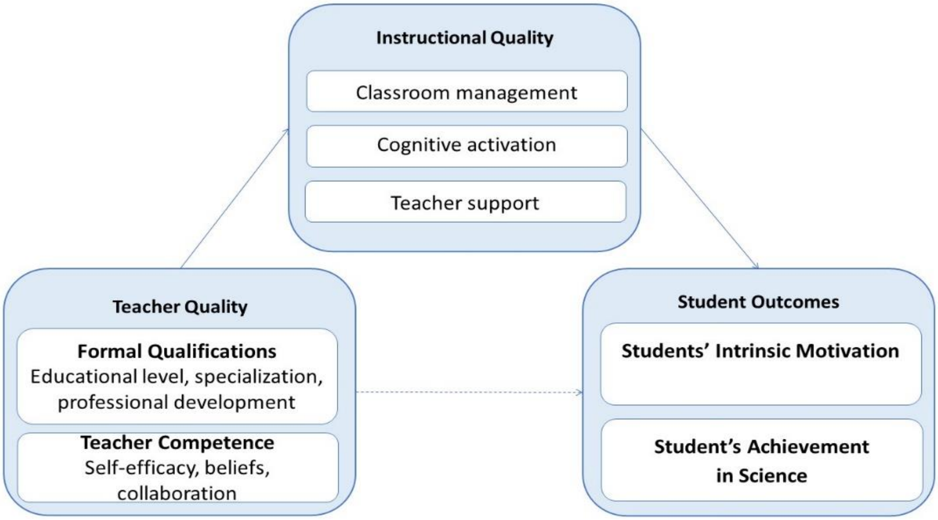
3.3.3 The relation between teacher quality, instructional quality and student outcomes

As previously mentioned, the relations between teacher quality and student outcomes may be indirect via instruction rather than direct (Creemers & Kyriakides, 2008; Goe, 2007). No matter how high a teacher's level of teacher education or specialization is, these characteristics can only influence student achievement or student motivation if they are visible in the teacher's instructional quality. If, for example, teacher's pedagogical content knowledge increases, but does not inform classroom instruction, it is unlikely that student learning will increase. This idea can be labeled as "the *mediating* role of instructional quality", and this represents the core assumption of the present study.

Figure 1 illustrates this role of instructional quality as a potential mediator. Indeed, the conceptual frameworks underlying the most renowned studies in the US (e.g., Darling-Hammond, 2006; Kane & Staiger, 2012; Pianta et al., 2012; Raudenbush, 2008) and Europe (e.g., Blömeke, Busse, et al., 2016; Klieme et al., 2009; Kunter et al., 2013; Neumann et al., 2012; Nilsen & Gustafsson, 2016) and within the educational effectiveness framework (e.g., Kyriakides et al., 2009) assume that the relation between teacher quality and student outcomes is indirect, thus considering instructional quality to be a mediator.

However, only a few studies have data available that provide information on the different teacher, instructional, and student aspects depicted in Figure 1. The present study therefore adds substantial new results to the state of research, with particular relevance to science as a subject in the Nordic countries.

Figure 1: Theoretical framework for the relations between teacher quality, instructional quality, and student outcomes



Although the relations described in Figure 1 are often considered to be generalizable across subjects, countries, and age groups, studies have shown that this is not necessarily the case (Blömeke, Olsen et al., 2016). Given that previous research was mostly conducted in English- or German-speaking countries, using mathematics achievement as the outcome (Goe, 2007; Neumann et al., 2012), the findings associated with these studies might not transfer to the Nordic context or to science. Further, not much is known about how these relations may change with students' age. A synthesis of studies by Goe (2007) showed that formal qualifications and especially specialization and content knowledge of teachers matter more in secondary than in primary school. Our study will therefore add a comparison of relations between grade levels 4 and 8 considering the different aspects of teacher quality, instructional quality, and student outcomes.

Moreover, few studies of the relations among teacher quality, instruction, and educational outcomes have focused on the domain of science (Goe, 2007; Neumann et al., 2012; Seidel & Shavelson, 2007). This is surprising because the relevance of teacher quality and instruction quality for student motivation should be evident (e.g., Fauth et al., 2014; Scherer & Nilsen, 2016). From a conceptual perspective, social-cognitive and motivational theories agree that motivation can be enhanced, and some argue that to be intrinsically motivated to learn, students must participate in instructional activities that are personally meaningful to them and that allow for positive experiences (Pintrich, 2003; Schiepe-Tiska et al., 2016).

3.4 Hypotheses

Our review of the state of research revealed gaps in previous research that point to a great need to examine the relations between teacher quality, instructional quality, and students' cognitive and motivational outcomes in the domain of science in the Nordic countries. With respect to the four research questions outlined above (see section 1), we tested the following hypotheses:

- There are significant positive relations between teachers' instructional quality and their students' achievement and motivation in science in the Nordic countries.
- There are rather few direct relations of teacher quality to their students' achievement and motivation in the Nordic countries.
- There is a substantial number of significant and positive indirect (via instructional quality) relations between teacher quality and student achievement and motivation.
- The relations between teacher quality and student outcome are not necessarily the same for younger students (in grade 4) and older students (in grade 8).

3.5 Method

3.5.1 Sample

The dataset used for the present study is that of the TIMSS. TIMSS is the only international large-scale study that samples entire classes within schools and collects background information from teachers. Only such a design allows for the examination of teacher variables that may explain differences in student outcomes between classrooms. Our sample was taken from the TIMSS 2015 science study.

With respect to the Nordic countries, Norway and Sweden participated in grade 8 and Denmark, Finland, Norway, and Sweden participated in grade 4. In grade 8, Norwegian students were taught science as an integrated subject, while Swedish students were taught physics, biology, and chemistry as separate subjects. Hence, in the student questionnaire, Swedish students reported on their motivation in each of the science subjects whereas Norwegian students reported on their motivation to learn science in an integrated way. A reliable comparison of their results is thus not possible from our perspective, and we therefore excluded motivation as an outcome from the grade 8 study. Table 1 provides a more detailed description of the samples.

Table 1: Summary statistics

		Denmark	Finland	Norway	Sweden
Grade 4	Number of students	3,710	5,015	4,164	4,142
	Number of classrooms	296	290	296	280
	Average number of students per classroom	19.3	14.9	19.1	19.5
Grade 8	Number of students			4,795	4,090
	Number of classrooms			225	706
	Average number of students per classroom			22.3	17.5

If more than one teacher was assigned to a group of students, data from all teachers were included by treating these as separate "classrooms" in order to assess the different teachers' influences on students. In grade 8, Sweden reported on 706 classrooms, which was more than any of the other countries. This is because there are several science teachers per classroom, one for each science subject (e.g. biology).

3.5.2 Data and variables used

The different aspects of teacher quality were measured by teachers' self-reports and included both formal qualifications and teacher competence. We provide information on the variables asked in the grade 8 teacher questionnaire in science. The questions were similar in grade 4. Teachers' formal qualifications included the following:

- Educational level ("What is the highest level of formal education you have completed?" with seven choices such as "Bachelor's or equivalent level—ISCED Level 6").
- Specialization in science or science education ("During your post-secondary education, what was your major or main area(s) of study?" with nine choices such

as "Physics"). This was split up for our purpose into teachers with a specialization in science on the one hand and those with a specialization in science education on the other.

- Professional development in certain science content areas ("In the past two years, have you participated in professional development in any of the following?" with seven types such as "Science pedagogy/instruction" to be marked as "Yes" or "No") and in terms of time ("In the past two years, how many hours in total have you spent in formal <in-service/professional development> [e.g., workshops, seminars, etc.] for science?" with five choices such as "Less than 6 hours")

Measures of teacher competence included the extent to which teachers collaborate with other science teachers ("How often do you have the following types of interactions with other teachers?" with seven types such as "Discuss how to teach a particular topic" to be rated from "Very often" to "Almost never"), how motivated they are for their work ("How often do you feel the following way about being a teacher?" with seven statements such as "I am proud of the work I do" to be rated from "Very often" to "Almost never"), their self-efficacy in pedagogical content knowledge ("In teaching science to this class, how would you characterize your confidence in doing the following?" with ten statements such as "Making science relevant for students" to be rated from "Very high" to "Low"), and their self-efficacy in content knowledge ("How well prepared do you feel you are to teach the following science topics?" with 22 topics covering the range of all science topics in the TIMSS framework to be rated from "Very well prepared" to "Not well prepared").

Instructional quality was measured by teachers' self-reports of practices that pertain to cognitive activation and support ("How often do you do the following in teaching this class?" with seven activities such as "Ask students to complete challenging exercises that require them to go beyond the instruction" to be rated from "Every lesson" to "Never").

Students' intrinsic motivation in science was reported by the students ("How much do you agree with these statements about learning science?" with nine statements such as "I enjoy learning science" to be rated from "Agree a lot" to "Disagree a lot").

Student achievement was assessed with standardized tests, including 168 items at grade 4 and 215 items at grade 8 reflecting knowledge in the different domains of science (life science, earth science, and physics at grade 4; biology, chemistry, earth science, and physics at grade 8). Students' responses to these items were calculated as so-called “plausible values”—values that resulted from statistical models that included information about the correctness of responses and available background variables. In the present study, all plausible values were included.

3.5.3 Analysis

TIMSS has a hierarchical design, with students nested in classrooms/teachers—a design that calls for multilevel analysis. We therefore took a two-level approach, with students on the first level and classrooms on the second level. We further used confirmatory factor analysis to estimate the degree to which each item measured the intended characteristic and the degree to which the analyses fit the data (Hox & Roberts, 2011). To compare relations among teacher, instructional, and student characteristics across countries, a certain degree of comparability is needed. In other words, it needed to be ensured that students and teachers had the same understanding of a question across countries. To test this, a procedure was applied called “measurement invariance testing”. Overall, sufficient levels of measurement invariance were supported by the data (Millsap, 2011).

Separate models for each aspect of teacher quality were estimated because these aspects are often highly correlated with each other, and this can lead to serious problems with the precision of the results. To investigate the relation between teacher quality, instructional quality, and student outcomes, two sets of models were estimated with the data from grade 4, one in which achievement was the outcome and one with students’ intrinsic motivation as the outcome. In grade 8, only achievement was used as the outcome variable such that one type of model was sufficient. To test whether the hypothesized mediation processes between teacher quality, instructional quality, and student outcomes were significant, we used a procedure (test) in the M*plus* software package. Considering the number of classrooms for each sample and the number of variables in all models, the level of statistical significance was set to $p < .10$.

Because the aim of the present study was to explain differences between teachers/classrooms, not between individual students, all results are reported at the classroom level. To enable comparisons across countries, standardized results are presented. These represent the change in student outcomes if the teacher or instructional quality changes by one unit. The fit of all models tested ranged from acceptable to very good.

All analyses were conducted with the statistical software package M*plus* version 8 (Muthén & Muthén, 1998-2017).

3.6 Results and Interpretation

A summary of the overall patterns of results is provided here, while more detailed information is provided in the appendix.

3.6.1 Results for grade 4

Relation between instructional quality and student outcomes

The findings concerning our first research question are summarized in Table 2.

For Denmark, science teachers' self-reported instructional quality did not have a significant relation to student achievement in science as assessed with the TIMSS test, but it did have a significant and positive relation to student's intrinsic motivation (+). In Finland, the opposite was the case, and instructional quality had a significant and positive relation to science achievement in grade 4 (+) but not to intrinsic motivation. In Sweden and Norway, instructional quality had a significant and positive relation to both outcomes (+). The details are provided in Table A1 (first row) in the appendix.

Table 2: Relations between instructional quality and student achievement and students' intrinsic motivation in grade 4

Grade 4	Denmark		Finland		Sweden		Norway	
	Student Achievement	Student Motivation	Student Achievement	Student Motivation	Student Achievement	Student Motivation	Student Achievement	Student Motivation
Instructional quality	NS	+	+	NS	+	+	+	+

Note: NS = relation not significant, + = relation significant and positive.

These results show that in all four Nordic countries instructional quality has an important influence on student outcomes in grade 4. This applies particularly to Norway and Sweden where instructional quality is significant and positively related to both student achievement and student motivation. This applies also to Denmark and

Finland where one of the two student outcomes is positively related to instructional quality. This means overall that higher instructional quality in grade 4 in the Nordic countries is associated with stronger intrinsic motivation to learn science and/or greater student achievement in science.

The insignificant findings for the other respective outcomes in Denmark and Finland may have the following explanations: Either the data set does not capture the specific type of practices that are important to student outcomes in these countries, or the analysis is not sensitive enough to capture significant relations because of the relatively small number of teachers included. It might, for instance, be particularly important for grade 4 science teachers in Denmark or Finland to teach students how to be able to switch between different representations of the same phenomenon or to be able to handle inquiry approaches (Treagust, Duit, & Fischer, 2017). However, these types of practices are not captured in the construct.

Direct and indirect effects of teacher quality on student outcomes

The second research question asked about what aspects of teacher quality are directly related to instructional quality and student outcomes, while the third question asked whether instructional quality might mediate any of these aspects so that we can identify indirect effects of teacher quality on student outcomes. Figure 2 illustrates these relations:

- A refers to the direct effect of teacher quality on instructional quality.
- C refers to the direct effect of teacher quality on student outcomes.
- Teacher quality can, in addition, have an indirect effect on student outcomes via instructional quality. This indirect effect is called "mediation" (MED) and is the product of A and B.

Figure 2: The mediation model between teacher quality, instructional quality, and student outcomes

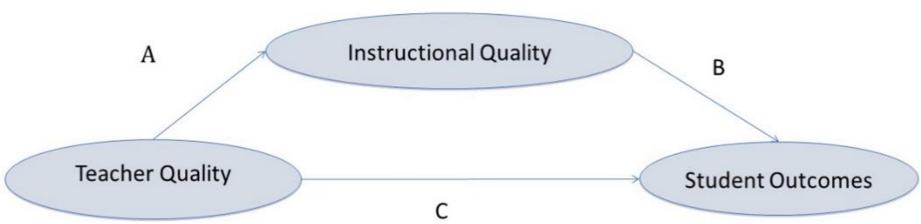


Table 3 summarizes the results regarding these effects (see Table A1 in the appendix with details for the models with achievement as the outcome and Table A2 with details for the models with student motivation as the outcome). The effect of instructional quality on student outcomes (estimate B in Figure 2) has already been reported in Table 2 and will not be repeated in Table 3 for the sake of clarity.

Table 3: Direct and indirect effects of teacher quality on instructional quality and on student achievement or students' intrinsic motivation

Grade 4	Denmark		Finland		Sweden		Norway	
	Achieve-ment	Motiva-tion	Achieve-ment	Motiva-tion	Achieve-ment	Motiva-tion	Achieve-ment	Motiva-tion
Educational level	NS	C –	NS	C –	NS	NS	A + C – MED +	A + MED +
Specialization in science	NS	C –	NS	NS	NS	A + MED +	NS	A + MED +
Content of professional development	NS	NS	NS	C –	A +	A + MED +	NS	NS
Hours of professional development	NS	C –	NS	NS	NS	A + MED +	NS	A +
Teacher motivation	A +	NS	A +	NS	A + C +	NS	A + C + MED +	A + MED +
Self-efficacy content knowledge	NS	NS	NS	C +	NS	NS	NS	NS
Self-efficacy pedagogical content knowledge	A +	A +	A +	C +	A + MED +	A + MED +	A +	A + MED +
Collaboration	A +	A+	A + MED +	A +	A + MED +	A + MED +	A + MED +	A + MED +

Note: A = direct effect of teacher quality on instructional quality, C = direct effect of teacher quality on student outcomes, MED = indirect effect of teacher quality on student outcomes via instructional quality. NS = not significant, + = relation significant and positive, – = relation significant and negative. For the sake of clarity, relations between instructional quality and student outcomes are not displayed again (see Table 2 for a summary of these).

The overall pattern. The overall pattern in the results shows that, across all countries, science teachers' characteristics have a significant relation to student outcomes in grade 4, although the relevance varies by characteristic, outcome, and country. The relations are more significant between science teachers' competence (i.e., their motivation, self-efficacy regarding content knowledge, self-efficacy regarding pedagogical content knowledge, and collaboration activities) and student outcomes than between science teachers' formal qualifications (i.e., their educational level, content and hours of professional development, and specialization) and student outcomes. These results point to the particular relevance of teachers' competence.

Furthermore, more aspects of science teachers' quality are significantly related to students' intrinsic motivation to learn science than to students' achievement in science. These results show how important it is to take motivational outcomes into account when discussing teacher effects instead of just focusing on student achievement in grade 4. Finally, whereas all significant relations between teachers' competence and student outcomes are positive, some significant relations between teachers' formal qualifications and students' intrinsic motivation are negative. These results might point to a potential ambiguity when looking only at formal qualifications. The following sections discuss these findings in greater detail.

The different aspects of teacher quality

Educational level. The more detailed picture shows that teachers' educational level had a significant and positive relation to instructional quality (A +) and student outcomes, in terms of both achievement and motivation, mediated by instructional quality (MED +) in Norway. In contrast, higher levels of education were significant and negatively associated with student motivation in Denmark and Finland (C –). Furthermore, also in Norway a negative direct relationship of teachers' educational level to achievement was seen. These unexpected negative relations may indicate that highly educated teachers teach students with lower motivation, which could be a result of compensatory teaching approaches in grade 4. At the same time, it could also mean that teachers with a higher educational degree are less skilled at motivating students to learn science.

Specialization in science. Surprisingly, teachers' specialization in science did not have a significant and positive relation to student achievement in any of the four Nordic countries. One reason for this could be the way specialization was measured, where teachers answered whether or not they had specialized in, for instance, science during their education. This is a very rough and inaccurate measure, and given the small sample of teachers, the method might not have been sensitive enough to capture any effects. However, in Sweden and in Norway teachers' specialization was positively associated with their instructional quality (A +) and to students' intrinsic motivation (MED +). One question that arises is why there were significant findings only in Norway and Sweden. One possible explanation is that far more teachers have specialization in science in Norway (26%) and Sweden (66%) than in Denmark (14%) and Finland (13%). Another question is why there were significant relations to motivation and not to achievement. Further research is needed to answer this, but a teacher who has great content knowledge may be able to inspire their students more, perhaps due to their own interest in the subject, and perhaps because they are able to make the content more engaging than teachers who do not have the same deep content knowledge. This still does not answer why there were no significant relations between specialization and achievement. However, according to Goe (2007), teacher content knowledge is more important for older students, while teacher support and pedagogical content knowledge may play a greater role in primary school.

In contrast to Sweden and Norway, specialization in science in Denmark had a negative relation to motivation (C –). As mentioned with respect to teachers' educational level, this unexpected result may either point to a compensatory approach where specialized teachers are purposely assigned to classes with lower motivation to learn science, or to a lower ability of such teachers to motivate students. There were no significant findings for Finland.

Content and hours of professional development. Both indicators of professional development were of strong relevance in Sweden. Science teachers who reported broader participation in different types of professional development courses (e.g., science curricula, assessments) and those science teachers who had taken part in more hours of professional development reported higher instructional quality (A +) and had students with stronger intrinsic motivation (MED +). Sweden was, therefore, the country with the most significant findings in grade 4. In fact, Sweden has spent considerable resources on professional development lately (Gustafsson & Nilsen, 2017; Mullis, Martin, Goh, & Cotter, 2016), and this is reflected in the data as almost 60% of Swedish teachers reported that they had participated in such courses, while only around 20% of the Norwegian teachers, 20% of Finnish teachers, and 34% of Danish teachers had done so. The Swedish results suggest that this huge effort might be paying off.

In the other three countries, the content or the hours of professional development were rarely significant and positively related to student outcomes. In Denmark and

Finland, enrollment in professional development even had a direct and negative relation to intrinsic motivation (C –). These negative direct relations may indicate that it was teachers struggling with motivating their students in grade 4 in particular who took part in professional development activities.

Teacher motivation. In all four Nordic countries there were significant findings for grade 4 teachers' motivation for their work in the models where achievement served as the outcome variable. In these cases, teacher motivation had a positive and significant relation to instructional quality (A +). In Sweden and Norway, teacher motivation also had a direct (C +) and indirect effect (MED +) on student achievement in science. In Norway, the relevance of teacher motivation was even greater because it also had an indirect effect on student motivation. In all Nordic countries, these findings show how important it might be that science teachers have positive feelings regarding their job, for example, being proud to be a science teacher.

Teachers' self-efficacy regarding their content knowledge and their pedagogical content knowledge. In contrast to our expectations, how well-prepared science teachers felt for teaching the range of topics included in the TIMSS framework was rarely relevant for student outcomes. This applied to all four Nordic countries and to potential direct as well as indirect effects. This suggests that content knowledge alone might not be sufficient to feel prepared for science teaching in grade 4. Only in Finland was there a direct relation between teachers' self-efficacy in content knowledge and their students' intrinsic motivation.

In contrast, teachers' self-efficacy regarding their pedagogical content knowledge was highly relevant either to instructional quality or to student outcomes in all Nordic countries in grade 4. This type of teacher self-efficacy better reflects their competence beliefs concerning pedagogical practices (e.g., inquiry approaches or providing challenging tasks) than self-efficacy in content knowledge does. For the models with achievement as the student outcome, teachers' self-efficacy in pedagogical content knowledge had a positive and significant relation to instructional quality in all four Nordic countries (A +). In Sweden, instructional quality also mediated the effect of teachers' self-efficacy on student achievement. For the models with students' intrinsic motivation as the outcome, teachers' self-efficacy in pedagogical content knowledge had a positive and significant relation to instructional quality in all Nordic countries (A +) except Finland. In Finland, there was a positive and significant direct relation to intrinsic motivation (C +). In Sweden and Norway, there was also a significant positive relation between self-efficacy in pedagogical content knowledge and intrinsic motivation as mediated by instructional quality (MED +).

Teacher collaboration. This characteristic of grade 4 science teachers was most often associated with instructional quality and student outcomes in all four Nordic countries. All relations were significant and positive, indicating that science teachers reporting to collaborate more often with other teachers provided higher levels of instructional quality (according to their self-reports) and promoted better student achievement and greater student motivation to learn science. In Denmark, teacher collaboration had a positive and significant relation to instructional quality in both the achievement and the motivation model (A +). In Finland, teacher collaboration also had a positive and significant relation to instructional quality in both models (A +). In addition, a significant effect of teacher collaboration was found on student achievement mediated by instructional quality (MED +). In both Norway and Sweden, the additional mediation effect was found for both types of student outcomes.

3.6.2 Results for grade 8

Our final research question asked about the extent to which the results for grades 4 and 8 are consistent. Notice that, in contrast to the analyses of the grade 4 data, achievement was the only outcome variable that could be used in grade 8. Furthermore, participation of the Nordic countries in TIMSS was limited to Sweden and Norway.

As was the case with grade 4 data, instructional quality had a significant positive relation to achievement in grade 8 in both Sweden and Norway (see Table 4). This consistency strengthens the relevance of instructional quality for student outcomes in the Nordic countries.

Table 4: Relations between instructional quality and student achievement in grade 8

Grade 8	Sweden	Norway
Instructional quality	Student Achievement +	Student Achievement +

Note: Relations between instructional quality and student achievement in grade 8

Table 5 shows the results in greater detail. It includes both grade 4 and grade 8 to allow comparisons between the two grade levels (see Table A3 in the appendix for detailed results for grade 8). The effect of instructional quality on student outcomes has already been reported in Table 4 and is not repeated in Table 5 for the sake of clarity.

Table 5: Direct and indirect effects of teacher quality on instructional quality and student achievement in Sweden and Norway

	Sweden		Norway	
	G4	G8	G4	G8
Educational level	NS	NS	A + Med +	NS
Specialization in science	NS	A + C+ Med +	NS	A + Med +
Content of professional development	A +	C +	NS	A +
Hours of professional development	NS	A + Med +	NS	A +
Teacher motivation	A + C +	A + Med +	A + Med +	A + Med +
Self-efficacy content knowledge	NS	A + Med +	NS	A +
Self-efficacy pedagogical content knowledge	A + Med +	A + Med +	A +	A + C +
Teacher collaboration	A + Med +	A + C + Med +	A + Med +	A + C + Med +

Note: A = direct effect of teacher quality on instructional quality, C = direct effect of teacher quality on student achievement, MED = indirect effect of teacher quality on student achievement via instructional quality, NS = not significant, + = relation significant and positive, – = relation significant and negative. For the sake of clarity, relations between instructional quality and student achievement are not shown again (see Table 4 for a summary of these).

Educational level. A comparison of the results between grades 4 and 8 for teacher quality showed that for teachers' educational level there were no significant relations to student achievement in science, neither for Sweden nor for Norway, in grade 8.

However, a significant positive effect on instructional quality (A +) and a mediation effect (MED +) were found for Norway in grade 4.

Specialization in science. For specialization in science, there were no significant relations to student achievement in science in either country in grade 4. However, in both countries a significant positive association between such a specialization and instructional quality existed (A +) as well as a direct effect on student achievement in Norway (C +) and an indirect effect on student achievement mediated by instructional quality in grade 8 in Sweden (MED +). These findings may indicate that subject-specific content knowledge plays a more significant role in grade 8 than in grade 4.

Content and hours of professional development. How broadly science teachers had taken professional development courses had a significant and positive relation to instructional quality in grade 4 in Sweden (A +) and a positive direct relation to student achievement in science in grade 8 (C +). In Norway, there were no significant findings for professional development courses in grade 4, but these had a positive and significant relation to instructional quality in grade 8 (A +).

Concerning the hours spent on professional development, there were no significant relations in grade 4 in either of the two countries. However, in grade 8 the length of professional development was significantly and positively related to instructional quality in both countries (A +). Furthermore, a significant positive relation between hours of professional development and science achievement mediated by instructional quality was seen in Sweden in grade 8 (MED +). These findings may indicate that professional development plays a larger role in grade 8 than in grade 4.

Teacher motivation. The relevance of teacher motivation for student achievement in science was found to be significant in grade 4, and this significance was confirmed in grade 8 in both Norway and Sweden. In both countries, teacher motivation had a significant and positive relation to instructional quality in grade 8 (A +). In addition, in both countries there was a significant indirect effect of teacher motivation on student achievement mediated by instructional quality (MED +).

Self-efficacy regarding content knowledge and pedagogical content knowledge. For teachers' self-efficacy regarding content knowledge, there were no significant relations in either country in grade 4. However, in grade 8 there was a significant positive effect of self-efficacy on student achievement in science in both countries (A +). Furthermore, a mediation effect was found in Sweden, which means that self-efficacy regarding content knowledge also had an indirect effect on science achievement (MED +). These results might again indicate that content knowledge plays a more important role in grade 8 than in grade 4.

For self-efficacy regarding pedagogical content knowledge, the grade 4 results had revealed a high relevance of this indicator of teacher quality for student achievement in science, and this relevance was confirmed for grade 8. In both countries, self-efficacy regarding pedagogical content knowledge had a significant positive effect on student achievement in science in grade 8 (A +). In Sweden, an additional mediation effect was found (MED +), and in in Norway an additional direct effect was found in grade 8 (C +).

Teacher collaboration. For teacher collaboration, there was consistency between the results for grade 4 and grade 8. The high relevance found for grade 4 with a significant positive effect of teacher collaboration on instructional quality (A +) was also found in grade 8 in both countries along with an additional significant indirect effect on student achievement mediated by instructional quality (MED +). In addition, in both countries a direct effect of teacher collaboration on student achievement was seen for grade 8. However, the direction was different and was positive in Sweden (C +) but negative in Norway (C –).

To summarize the comparisons between grades 4 and 8, it seems as if the indicators of teacher quality matter at least equally as much in grade 8 as in grade 4, and even more strongly in all content-related cases. The latter applies to teachers' formal qualifications in terms of specialization in science and content as well as hours of professional development and to teacher competence in term of self-efficacy regarding content knowledge. Teacher collaboration, self-efficacy regarding pedagogical content knowledge, and teacher motivation for their work seem to be of equal importance in both grades. These latter constructs represent more pedagogical aspects.

3.7 Summary and Discussion

Considering previous research on the relations between instructional quality, teacher quality, and student outcomes, there are, to our knowledge, no studies examining these relations for representative samples in the Nordic countries.

This study showed that science teacher's instructional quality had a positive and significant relation to student achievement in science in grades 4 and 8 in all Nordic countries except Denmark and to student motivation to learn science in grade 4 in all

Nordic countries except Finland. In Finland, instructional quality had a positive and significant relation to achievement but not to motivation, while the opposite was true for Denmark. These results are the first indicators for the strong relevance of teachers for student outcomes on the one hand and the importance of taking student motivation into account in addition to student achievement on the other.

The relevance of teachers for student outcomes could be confirmed with respect to several specific indicators of teacher quality. However, there were rarely *direct* teacher effects. The effects were almost always *indirect*, which means they were mediated by instructional quality. The type of teacher competences reflecting pedagogical aspects (i.e., teacher collaboration, teachers' self-efficacy regarding their pedagogical content knowledge, and teachers' motivation for their job) in particular had many positive and significant indirect effects on student outcomes in both grades.

Teacher self-efficacy in certain domains has been the focus of studies for some time (Zee & Koomen, 2016). Despite the variety of findings, one observation that is consistent across these studies is that teacher self-efficacy determines teaching practices and other relevant teacher characteristics (Klassen & Tze, 2014; OECD, 2014). The present study adds to this body of research by providing evidence for (a) the significant and positive relations between self-efficacy measures and student outcomes and (b) the mediating role of instructional quality for this relation.

How important it is to distinguish between achievement and motivation could also be confirmed with regards to the respective indicators of teacher quality. In grade 4, teacher competence was more often related to student motivation than to achievement. Our findings are largely in line with a study using TIMSS 2011 grade 4 data in the domain of mathematics and achievement as an outcome (Blömeke, Olsen, et al., 2016), thus extending these studies to the domain of science and to student motivation as an outcome.

Teachers' formal qualifications seemed, in contrast, to be of more importance for student achievement in grade 8 than in grade 4. This applied particularly to those characteristics related to content such as teachers' self-efficacy regarding their content knowledge. If one disregards the context of subject and culture and compares the present findings to the systematic review by Seidel and Shavelson (2007), the present results are largely in line with their findings. Seidel and Shavelson included mathematics, science, and reading achievement as outcome variables and found that domain-specific components had the largest effect size ($d = 0.4$) on student outcome (cognitive and affective). Similarly, in a research synthesis of relations between teacher quality, instructional quality, and student outcomes, Goe (2007) states that formal qualifications, and especially domain-specific specialization and content knowledge, matter more in secondary than in primary school. This seems to be a very robust finding. We have shown here that it was important to distinguish in some cases between the four Nordic countries. This applied particularly to teachers' formal qualifications, which were generally more important in Sweden and Norway than in Denmark and Finland. In Norway, educational level and the specialization in science were particularly relevant. This might be a result of a certain degree of backlog compared to the other Nordic countries such that teachers with a higher degree (i.e., a Master's degree) or a specialization in science represent a small but particularly well-trained group of teachers.

Sweden stood out with respect to professional development. There were far more significant relations for Sweden in both grades and stronger relations to outcomes than in the other countries. This result might on the one hand indicate differences in the allocation of resources as Sweden has put considerable effort into professional development. On the other hand, quality differences might underlie these differences in effects (Gustafsson & Nilsen, 2017). Previous research found that professional development only has an effect if the activities meet certain quality criteria (Timperley, Wilson, Barrar, & Fung, 2007), and Desimone (2009) classified these quality features into a focus on content, active learning, and a certain minimum length of the professional development course in order to be sustainable (Blömeke, Olsen, et al., 2016).

3.8 Limitations

All international large-scale surveys follow cross-sectional designs at the student, classroom, and school level; hence, there are issues pertaining to omitted variables and reversed causality. The data must therefore be interpreted with caution even though the data are of high quality, the samples are representative, and the methods of analysis are robust.

Another limitation refers to the restricted availability of data – of all the Nordic countries, only Sweden and Norway participated in TIMSS 2015 in grade 8. Moreover, in Sweden, science in grade 8 is not taught in an integrative way, that is, physics, chemistry, biology, and earth science form separate school subjects. Because Swedish students reported on their motivation for each of these subjects, comparisons of motivation and its relations to other constructs between Norway and Sweden are hardly possible.

Finally, items measuring instructional quality behave differently across countries, that is, while some of them are positively related to student outcomes in one country, they might be negatively related to student outcomes in another country. Perhaps a broader conceptualization of instructional quality will resolve these issues. To this end, future research should examine what aspects of instructional quality are important for student outcomes in the Nordic countries.

3.9 Conclusions for policy-makers, teacher educators, and researchers

It is important to note that causal inferences based on international large-scale data cannot be drawn. Still, when carefully examining the present findings in light of previous research, some implications and generalizations may be drawn. We would especially like to highlight the five following implications and contributions to policy and educational stakeholders, teacher education, educational research, and practice:

- The most important lesson learned may be the relevance of mediation or indirect effects in teaching quality. The teacher characteristics examined here rarely had direct effects on student outcomes but needed to become visible in terms of instructional quality. Without the latter, the former turned out to be almost meaningless. This result has clear implications for policy making and support mechanisms for the teaching profession. Note that teachers' competence and formal qualifications are important, but it is the interaction between these characteristics and what happens in the classroom that is most important.
- Another important lesson learned might be the relevance of taking student motivation into account when examining school effectiveness. As pointed out in the beginning of this article, motivation plays a major role in educational and job-related choices, for example, for further recruitment to the STEM field (Schiepe-Tiska et al., 2016), and it supports achievement. Our data show that several teacher characteristics are of particular relevance when it comes to strengthening student motivation. Thus, the dominating focus on achievement might be short-

sighted. Promoting students' motivation in primary school requires competent and qualified teachers who provide high-quality instruction.

- Strengthening teachers' and student teachers' self-efficacy and competence within pedagogical content knowledge, for instance, to teach inquiry, to assess students' understanding, to provide challenging tasks, and to link their teaching to students' daily life, can improve instructional quality. For teacher education and professional development in lower secondary school, it might be especially important to provide teachers and teacher students with high formal qualifications (e.g. educational level), including subject-specific specialization (e.g. in physics). This seems to be more important in grade 8 than grade 4.
- Professional development relates positively to the quality of science teachers' instruction in Sweden, which, in turn, has a positive effect on student outcomes in science. It seems that professional development could enhance student outcomes via their instructional quality if a certain length of time, breadth of content, and quality is ensured as part of the development program. We suggest looking more closely into the Swedish model in this respect given the many and large effects seen in this study.
- Teachers' motivation for their work and their collaboration is strongly related to student outcomes. It might therefore be important to support teachers not only in their self-efficacy to teach a specific subject, but also to help them consider their profession in a meaningful way. In addition, the benefits of teacher collaboration could, for example, be exploited by allocating time for collaboration during regular school work.

Given the limitations associated with the design of international large-scale assessments such as TIMSS, there is a clear need to substantiate the above-described findings and implications in longitudinal studies, preferably with equally representative samples. Moreover, further research is needed to disentangle how these relations operate for different groups of students (e.g., across gender or proficiency levels) because it is possible that certain practices promote learning in one group of students while other types of practice might promote learning in a different group. It must also be emphasized that one cannot look at these effects in isolation; they exist and work simultaneously and must therefore be examined together as this study does.

Altogether, the findings presented here address the pertinent aim of enhancing student motivation and learning outcomes in science education, both of which are important for future competence in and recruitment to the STEM fields. Our findings further point to the importance of instructional quality and certain aspects of teacher quality for student outcomes, and these findings may inform policy decisions, for example, through teacher education or professional development. In addition, the results show that specific features are relevant in the four Nordic countries, but not to the same degree. These similarities and differences necessitate further inquiries and show that "one-size-fits-all" might not apply. In other words, the results call for care with respect to inferences on their generalizability across Nordic countries, as they might not transfer readily from one country to another.

3.10 References

Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International journal of science education, 30*(10), 1405-1416.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: WH Freeman and Co.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal, 47*(1), 133-180.

Blömeke, S., Busse, A., Kaiser, G., König, J., & Suhl, U. (2016). The relation between content-specific and general teacher knowledge and skills. *Teaching and Teacher Education, 56*, 35-46. doi:10.1016/j.tate.2016.02.003

Blömeke, S., Gustafsson, J.-E., & Shavelson, R. J. (2015). Approaches to competence measurement in higher education: Hogrefe Publishing.

Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of Student Achievement to the Quality of Their Teachers and Instructional Quality. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes* (pp. 21-50): Springer.

Bøe, M. V. (2012). Science choices in Norwegian upper secondary school: What matters? *Science Education, 96*(1), 1-20.

Caprara, G. V., Barbaranelli, C., Steca, P., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology, 44*(6), 473-490. doi:10.1016/j.jsp.2006.09.001

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological bulletin, 140*(4), 980.

Cohen, J., & Grossman, P. (2016). Respecting complexity in measures of teaching: Keeping students and schools in focus. *Teaching and Teacher Education, 55*, 308-317.

Creemers, B., & Kyriakides, L. (2008). The dynamics of educational effectiveness. *A contribution to policy, practice and theory in contemporary schools. Abingdon: Routledge*.

Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational researcher, 35*(7), 13-24.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*: Springer Science & Business Media.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational researcher, 38*(3), 181-199.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*(1), 109-132.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1-9.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. Teachers college record, 109(4), 877-896.

Goe, L. (2007). The Link between Teacher Quality and Student Outcomes: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.

Gustafsson, J.-E., & Nilsen, T. (2017). *Changes in Mathematics Performance from 2011 to 2015 in Norway and Sweden as a Function of Teacher Professional Development*. Paper presented at the ECER 2017, Copenhagen.

Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2016). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation*.

Gustafsson, J. E., & Nilsen, T. (2016). The Impact of School Climate and Teacher Quality on Mathematics Achievement: A Difference-in-Differences Approach. In T. Nilsen & J. E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes* (pp. 81-95): Springer.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. *Second handbook of research on mathematics teaching and learning, 1*, 371-404.

Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774-786. doi:10.1037/a0032198

Hox, J., & Roberts, J. K. (2011). *Handbook of advanced multilevel analysis*: Psychology Press.

Kane, T., & Staiger, D. O. (2012). Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains. Research Paper. MET Project. *Bill & Melinda Gates Foundation*.

Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59-76.

Klieme, E., Pauli, C., & Reusser, K. (2009). The pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. *The power of video studies in investigating teaching and learning in the classroom*, 137-160.

Kuger, S., Klieme, E., Jude, N., & Kaplan, D. (2016). *Assessing contexts of learning: An international perspective*: Springer.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology, 105*(3), 805-820. doi:10.1037/a0032583

Kyriakides, L., Creemers, B., & Antoniou, P. (2009). Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education, 25*(1), 12-23

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*: Routledge.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474-496.

Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (Eds.). (2016). *TIMSS 2015 Encyclopedia: Education Policy and Curriculum in Mathematics and Science*. Boston College: TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user´s guide* (Eight ed.). Los Angeles, CA: Muthén & Muthén.

Neumann, K., Kauertz, A., & Fischer, H. E. (2012). Quality of instruction in science education *Second international handbook of science education* (pp. 247-258): Springer.

Nilsen, T., & Gustafsson, J.-E. (Eds.). (2016). *Teacher quality, instructional quality and student outcome. Relationships across countries, cohorts and time.* (Vol. 2): Springer

OECD. (2006). Evolution of Student Interest in Science and Technology Studies Policy Report. http://www.oecd.org/dataoecd/16/30/36645825.pdf

OECD. (2014). Teacher Self-Efficacy and Job Satisfaction: Why They Matter *TALIS 2013 Results: An International Perspective on Teaching and Learning*. Paris: OECD Publishing.

OECD. (2016). *Pisa 2015 results in focus* (2226-0919). Retrieved from

Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections* (Vol. 13): London: The Nuffield Foundation.

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions *Handbook of research on student engagement* (pp. 365-386): Springer.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667.

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206-230.

Rosén, M., & Gustafsson, J.-E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-scale Assessments in Education, 4*(1), 1-19. doi:10.1186/s40536-016-0020-8

Scherer, R., & Nilsen, T. (2016). The Relations Among School Climate, Instructional Quality, and Achievement Motivation in Mathematics *Teacher Quality, Instructional Quality and Student Outcomes* (pp. 51-80): Springer.

Schiepe-Tiska, A., Roczen, N., Müller, K., Prenzel, M., & Osborne, J. (2016). Science-related outcomes: Attitudes, motivation, value beliefs, strategies *Assessing Contexts of Learning* (pp. 301-329): Springer.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research, 77*(4), 454-499. doi:10.3102/0034654307310317

Snow, R. E. (1994). Abilities in academic tasks. *Mind in context: Interactionist perspectives on human intelligence*, 3-37.

Tatto, M. T., Peck, R., Schwille, J., Bankov, K., Senk, S. L., Rodriguez, M., . . . Rowley, G. (2012). *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-MM)*: ERIC.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). Teacher professional development and learning: Best evidence synthesis iteration (BES): Wellington: Ministry of Education.

Treagust, D. F., Duit, R., & Fischer, H. E. (2017). *Multiple representations in physics education* (Vol. 10): Springer.

Tschannen-Moran, M., & Barr, M. (2004). Fostering student learning: The relationship of collective teacher efficacy and student achievement. *Leadership and Policy in Schools, 3*(3), 189-209.

van Tartwijk, J., & Hammerness, K. (2011). The neglected role of classroom management in teacher education. *Teaching Education, 22*(2), 109-112. doi:10.1080/10476210.2011.567836

Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research, 86*(4), 981-1015.

3.11 Appendices

Table A1: Standardized regression coefficients at the classroom level for relations between teacher quality, instructional quality, and student achievement in science in grade 4 (*p < .10, **p < .05, NS = not significant)

Variable in focus	Direct and indirect effects of this variable	Denmark	Finland	Sweden	Norway
Instructional Quality	Direct effect on achievement	NS	0.18 **	0.16 *	0.33 **
Educational level	Direct effect on instructional quality	NS	NS	NS	0.27 **
	Direct effect on achievement	NS	NS	NS	-0.19 **
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	0.10 **
Specialization in science	Direct effect on instructional quality	NS	NS	NS	NS
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	NS
Content of professional development	Direct effect on instructional quality	NS	NS	0.18**	NS
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	NS
Hours of professional development	Direct effect on instructional quality	NS	NS	NS	NS
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	NS
Teacher motivation	Direct effect on instructional quality	0.17**	0.18**	0.21**	0.18**
	Direct effect on achievement	NS	NS	0.14*	0.25**
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	0.05*
Self-efficacy content knowledge	Direct effect on instructional quality	NS	NS	NS	NS
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS	NS	NS

Variable in focus	Direct and indirect effects of this variable	Denmark	Finland	Sweden	Norway
Self-efficacy pedagogical content knowledge	Direct effect on instructional quality	0.36**	0.19**	0.51**	0.19*
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS	0.09*	NS
Teacher collaboration	Direct effect on instructional quality	0.31**	0.30**	0.35**	0.32**
	Direct effect on achievement	NS	NS	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	0.06*	0.06*	0.09**

Note: For the sake of clarity, the effect of instructional quality on student achievement is only displayed in the first row and is not repeated in the later models.

Table A2: Standardized regression coefficients at the classroom level for relations between teacher quality, instructional quality, and students' intrinsic motivation in science in grade 4. (*p < .10, **p < .05, NS = not significant)

Variable in focus	Direct and indirect effects of this variable	Denmark	Finland	Sweden	Norway
Instructional Quality	Direct effect on student motivation	0.25**	NS	0.31**	0.33**
Educational level	Direct effect on instructional quality	NS	NS	NS	0.23**
	Direct effect on student motivation	−0.14*	−0.15**	NS	−0.14**
	Indirect effect on motivation mediated by instructional quality	NS	NS	NS	0.09**
Specialization in science	Direct effect on instructional quality	NS	NS	0.19**	0.18**
	Direct effect on student motivation	−0.18**	NS	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	0.06*	0.06*
Content of professional development	Direct effect on instructional quality	NS	NS	0.31 **	NS
	Direct effect on student motivation	NS	−0.13*	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	0.10**	NS

Variable in focus	Direct and indirect effects of this variable	Denmark	Finland	Sweden	Norway
Hours of professional development	Direct effect on instructional quality	NS	NS	0.19 **	0.12**
	Direct effect on student motivation	−0.20**	NS	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	0.05*	NS
Teacher motivation	Direct effect on instructional quality	NS	NS	NS	0.22**
	Direct effect on student motivation	NS	NS	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	NS	0.07*
Self-efficacy content knowledge	Direct effect on instructional quality	NS	NS	NS	NS
	Direct effect on student motivation	NS	0.20**	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	NS	NS
Self-efficacy pedagogical content knowledge	Direct effect on instructional quality	0.32**	NS	0.50**	0.27**
	Direct effect on student motivation	NS	0.21**	NS	NS
	Indirect effect on motivation mediated by instructional quality	NS	NS	0.16**	0.09*
Teacher collaboration	Direct effect on instructional quality	0.22**	0.28**	0.30**	0.26**
	Direct effect on student motivation	NS	NS	NS	−0.18*
	Indirect effect on motivation mediated by instructional quality	NS	NS	0.09**	0.11**

Note: For the sake of clarity, the effect of instructional quality on student achievement is only displayed in the first row and is not repeated in the later models.

Table A3: Standardized regression coefficients at the between level for relations between teacher quality, instructional quality, and student achievement in science in grade 8. (*p < .10, **p < .05, NS = not significant)

Variable in focus	Direct and indirect effects of this variable	Sweden	Norway
Instructional Quality	Direct effect on student achievement	0.20**	0.19**
Educational level	Direct effect on instructional quality	NS	NS
	Direct effect on student achievement	NS	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS
Specialization in science	Direct effect on instructional quality	0.19**	0.26**
	Direct effect on student achievement	−0.20**	NS
	Indirect effect on achievement mediated by instructional quality	0.04**	0.05*
Content of professional development	Instructional quality ON	NS	0.19**
	Direct effect on student achievement	0.08*	NS
	Indirect effect on achievement mediated by instructional quality	NS	NS
Hours of professional development	Direct effect on instructional quality	0.11**	0.15*
	Direct effect on student achievement	NS	NS
	Indirect effect on achievement mediated by instructional quality	0.02*	NS
Teacher motivation	Direct effect on instructional quality	0.47**	0.38**
	Direct effect on student achievement	NS	NS
	Indirect effect on achievement mediated by instructional quality	0.09**	0.07*
Self-efficacy content knowledge	Direct effect on instructional quality	0.23**	0.15*
	Direct effect on student achievement	NS	NS
	Indirect effect on achievement mediated by instructional quality	0.04**	NS
Self-efficacy pedagogical content knowledge	Direct effect on instructional quality	0.53**	0.59**
	Direct effect on student achievement	NS	0.20*
	Indirect effect on achievement mediated by instructional quality	0.11**	NS
Teacher collaboration	Direct effect on instructional quality	0.39**	0.36**
	Direct effect on student achievement	0.12**	−0.17*
	Indirect effect on achievement mediated by instructional quality	0.06**	0.09**

4. Nordic students' interest and self-belief in science

Magnus Oskarsson, Departement of Mathematics and Science Education, Mid Sweden University.
Marit Kjærnsli, Department of Teacher Education and School Research, University of Oslo.
Helene Sørensen, Danish School of Education, Aarhus University.
Hanna Eklöf, Department of Applied Educational Science, Umeå University.

Modern society requires people with positive attitudes toward science and who have science-related competences. This article will focus on Nordic students' enjoyment in learning science, their self-efficacy in solving scientific problems, and their instrumental motivation to learn science. Findings from PISA 2006 and PISA 2015 show an increasing interest in science in most Nordic countries, and more students in all Nordic countries except Denmark indicate that they expect to have a science-related occupation in the future. At the same time, we can see increased gender differences and greater variation in enjoyment and self-beliefs among students. Enjoyment of learning science and science self-efficacy correlate positively with performance, while instrumental motivation and enjoyment of learning science are associated with an increased likelihood that the student expects to have a science-related career. These findings are discussed in relation to changes in society over the past decade in terms of the visibility and use of science and technology.

4.1 Background

Students' interest in, motivation to learn, and self-beliefs in relation to science are regarded as important factors for their science competence, their future career choices, and their understanding of the role of science in modern society. Student interest and motivation is thus a core issue in educational settings because achievement motivation is assumed to interact with achievement behaviour in important ways (Pintrich & Schunk, 2002; Wigfield & Eccles, 2002). It is often claimed, and has been empirically shown, that an interested and highly motivated student performs better in achievement situations, has higher educational aspirations, expends more effort in learning new tasks, and uses more efficient self-regulating strategies compared to less motivated students (Pintrich & Schunk, 2002). Enjoyment in learning, self-beliefs, and valuing of a subject or task are often seen as parts of a larger interest/motivation construct. Several studies have demonstrated that enjoyment in learning and positive self-beliefs (self-concept, self-efficacy) are positively related to achievement (Bandura, 2010; Krapp & Prenzel, 2011; Marsh & Craven, 2006; Skaalvik & Valas, 1999; Zimmerman, 1992). In contrast, the value students attribute to different subjects has been shown to have a relatively weak relationship with performance compared with other background variables, but there is evidence that value perceptions do predict future achievement choices such as enrolment in mathematics education (Wigfield & Eccles, 2000).

The importance of interest and motivation variables is also evident in the PISA frameworks. The PISA definition of science literacy recognises that a student's response to a science-related issue requires more than just skills and knowledge; it also depends on how able and "willing" the student is "to engage" with the issue. The student questionnaire in PISA measures students' attitudes towards science in the following three areas: interest in science and technology, environmental awareness, and valuing scientific approaches to enquiry. These three areas were selected for measurement in PISA because a positive attitude towards science, a concern for an environmentally sustainable way of life, and a disposition to value the scientific approach to enquiry were considered characteristic of a scientifically literate individual. Thus, the extent to which individual students are interested in science and recognise its value and implications is considered an important measure of the outcome of compulsory education. If scientific approaches to enquiry are valued, it is possible to understand the way scientific claims are supported by the data and by reasoning and to understand the difference between facts and opinions (OECD, 2016a).

However, even if interest in science and the motivation to learn and work with science are perceived as important parts of science literacy in policy documents, education frameworks, and research, this does not automatically result in interested students. Rather, low interest in science and in pursuing a science career has been perceived as a problem in many countries (Bybee & McCrae, 2011), and it is claimed that school science fails in engaging students and that students think that school science is difficult and not relevant for everyday use or for their future jobs (Schreiner, 2005; Lyons, 2006; Sørensen 2008; Jidesjö, Oscarsson, Karlsson, & Strömdahl, 2009; Oskarsson, 2011).

The OECD is concerned that the proportions of students who choose careers in science are insufficient for the needs of modern society (OECD, 2016a), and several reports describe a decline in enrolment and graduation rates for science-related fields and perceived shortages of science graduates in the labour market (Gago et al., 2004). Studies of students' interest in science and attempts to raise interest in science and increase recruitment to science-related occupations show the importance of personal relevance for the students (Gago et al., 2004; Osborne & Dillon, 2010; Teknikdelegationen, 2010).

Previous cycles of PISA and TIMSS have shown that in the Nordic countries the students' interest in science is low compared to other countries in the EU and OECD (Martin, Mullis, Foy, & Stanco, 2012; OECD, 2007). TIMSS data have further shown that students in grade 4 are more positive towards science than students in grade 8 in most participating countries (Mullis et al., 2016). Oskarsson et al. (2017) showed that the lower interest in science among grade 8 students in Sweden is an important explanation behind Swedish grade 4 students having relative better scores in TIMSS Science than grade 8 students (Oskarsson, Eliasson, & Karlsson, 2017).

This article focuses on students' interest in science in a Nordic perspective and over time using data from PISA 2006 and PISA 2015. The Nordic perspective has been in focus in other reports as well, and the interested reader is referred to Kjærnsli & Jensen, 2016; Sørensen & Dohn 2016a, 2016b. Science was the main subject in PISA in 2006 and PISA 2015, thus data from these two studies provide opportunities to study changes in student perceptions of science over the last decade. Students in these two years were asked about their interest in learning science, their beliefs in their science competence, and the value they placed on learning science. For teachers and policy makers, an informed discussion about the relative importance of different interest variables, about changes in student interests, and about possible explanations for observed changes might provide valuable information about the role of interest and motivation in science education.

Today's society faces global challenges such as environment degradation, sustainability, and climate changes. Science, technology, and modern industry are on one hand part of the problem with emissions of carbon dioxide and the use fossil fuels as just one example. On the other hand, science and technology provide possible solutions to many of these problems with solar power plants, windmills, and other sustainable ways of producing energy.

Recently, major changes have taken place in society related to our use of everyday science and new technology. Today 15-year-old students live in an information-saturated society that is very different from only a decade ago. Technology has brought changes in how young people communicate with peers and interact with others. The Nordic countries are on top among all countries in the number of computers in school, students' use of computer in their leisure time, and their use of social networks (OECD, 2017), and this has likely had an impact on how students come into contact with and how they value science and technology.

The focus in the current paper is on the Nordic countries, which share many cultural and historical characteristics. They also have similar educational systems, and previous studies using PISA data have shown that students in Denmark, Iceland, Norway, and Sweden tend to follow a "Nordic pattern" in their answers to questions about attitudes. Finland also has many similarities with the other Nordic countries, but it also has some important differences (Kjærnsli & Lie, 2011). Although it has been shown, that the Nordic countries have several characteristics in common and often are treated as a single entity, it does not mean that they are in fact the same. One of the purposes of the present study was to explore possible differences between the Nordic countries.

4.1.1 Aims

Students' attitudes and motivation in relation to science are important for their future career choices and for their understanding of the role of science in modern society. It is therefore important to study how different factors such as instrumental motivation, enjoyment of learning science, and self-efficacy are related to each other and whether these variables can predict students' performance and their future career choice. This article reports descriptive findings and relationships between these variables in PISA 2006 and PISA 2015 for the Nordic countries and the OECD as a whole.

4.1.2 *Research questions:*

- What levels of self-reported enjoyment in learning science, science self-efficacy, and instrumental motivation do students in the Nordic countries report in PISA? Has their interest in and motivation to learn science changed over time, and are there differences between the Nordic countries?
- Are there differences between boys and girls in terms of interest in and motivation to learn science and in their expectations of pursuing a science career, and have possible gender differences changed over time in the Nordic countries?
- How are different interest variables related to each other, and how do these variables affect science performance and students' willingness to consider a career in science?

4.2 Data and analyses

We used student questionnaire and performance data from all Nordic countries from PISA 2006 and 2015. The choice arose from our focus on science, and science was the main subject in both years. The sample in the Nordic countries was in total 22,170 15-year-old students in 2006 and 27,331 15-year-old students in 2015 (OECD, 2009, 2016b).

First, we selected relevant data to answer our research questions from the PISA databases and processed them to provide the dataset for our detailed analysis. In order to be able to study changes over time, we chose to work with the three interest/motivation constructs that were assessed in both 2006 and 2015 (*enjoyment of learning science*, *instrumental motivation*, and *science self-efficacy*) along with an open-ended item asking the students about their future career expectations. We sought to identify changes from 2006 to 2015, differences between boys and girls, and changes in response patterns for individual items as well as relationships between the respective interest/motivation constructs and student's performance in science in PISA. We also investigated the interrelationships between the different interest indices, their effect on performance when modelled together rather than separately, and their effect on the likelihood that the student is interested in a science-related occupation. In the analyses, we used the motivation/interest scales, or "indices", already provided in the PISA databases. All index scales in 2015 were set to allow for comparisons with the corresponding index scales in PISA 2006 and are thus suitable for trend analyses (OECD, 2016b).

The index value for the average OECD student was zero the first time it was calculated (in 2006) and had a standard deviation of 1. This means that two thirds of the OECD students' index values would be between the values of −1 and 1. It is important to point out that negative values on the index do not imply that students responded negatively. Instead, a negative value means that the student's response had a value below the average response across the OECD countries. Likewise, students with positive values on the index are those who responded more positively than an average student in the OECD (see more details in Annex 1 in the OECD report).

In all analyses, student weights and all plausible values were used (for more information about the use of weights and plausible values as estimates of student proficiency, see OECD, 2009, 2016b). In the processing and analysis of the data, we used the Excel, SPSS/IEA IDB Analyzer, and Mplus 7.2 (Muthén & Muthén, 1998-2002) software packages. Both the IDB Analyzer and Mplus are analytical tools that are suitable for use with PISA-type data. In particular, Mplus is suitable for so-called multilevel analysis. Because students participating in PISA are clustered in schools, the school level is a source of variation that might be relevant to consider in the analyses. A general rule of thumb is that school-level variation is considered non-trivial if it accounts for more than 5% of the total variation. Two-level analyses performed in Mplus, however, suggested that only a small amount (1%–4%) of the variation in the interest indices was at the school level. Therefore, we decided to continue with and report only single-level analyses, but with standard errors adjusted to account for the clustered structure of the data. The same result was obtained regardless of the software used, and below the SPSS output (obtained by using the IDB Analyzer) is presented.

4.3 Results

PISA distinguishes between two forms of motivation for learning science. Students might learn science because they like it and find it interesting and/or because they think that science can be useful in their lives. This is the background for investigating the students' desire to learn and their instrumental motivation supplemented with science-related career expectations. In the first part of the results section, we will present the results for three of the constructs in PISA, *enjoyment of learning science*, *instrumental motivation to learn science*, and *science self-efficacy*. We will start with an international outlook regarding one of the constructs, *enjoyment of learning science*, and then focus on the Nordic perspective for all three constructs. In the second part of the results section, we will present findings for the Nordic countries for the open-ended question about career expectations. Both the constructs and the open-ended questions were identical in PISA 2006 and PISA 2015. For each construct, we will present the index values (mean and standard deviation) for the Nordic countries and the OECD, respectively, and we will explore gender differences and changes from PISA 2006 to PISA 2015. In addition to analysing these constructs one by one, we also performed a correlation analysis and a multiple regression analysis to explore the relationships between the three constructs and their relative effects on the student's scores in science when modelled together. The results of this analysis are presented at the end of the first part of the results section. Furthermore, we present results for the students' expectations of science-related careers, and in connection with this we provide the results of a logistic regression analysis with the different interest variables as predictors of science career expectations.

4.3.1 *Enjoyment of learning science*

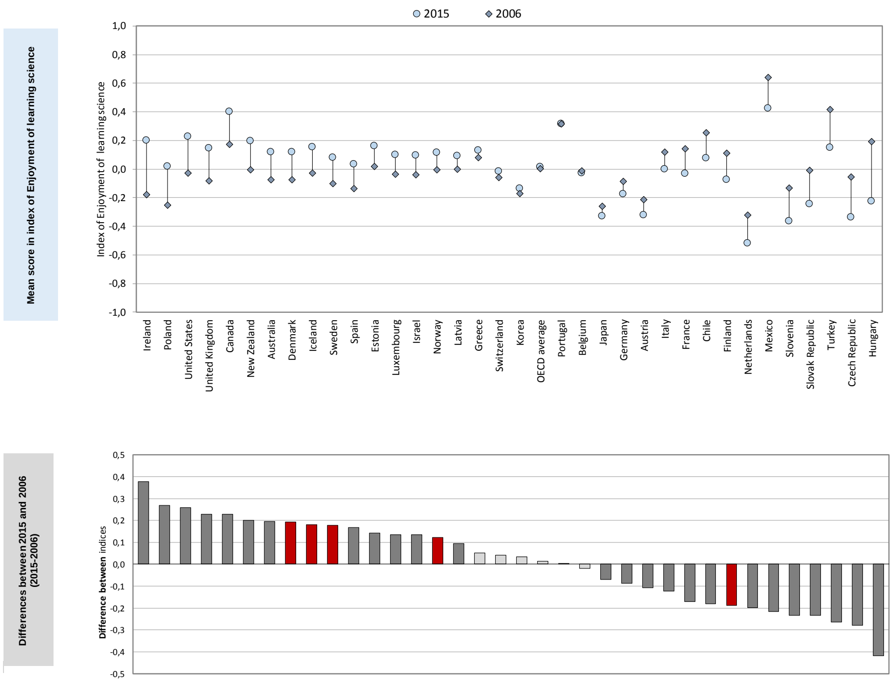
The construct, *enjoyment of learning science* is measured in PISA by five statements. The students were asked how much they disagreed or agreed with each statement, and the response categories were "strongly disagree", "disagree", "agree", and "strongly agree":

- I generally have fun when I am learning science topics.
- I like reading about science.
- I am happy working on science.
- I enjoy acquiring new knowledge in science.
- I am interested in learning about science.

Figure 1 shows the index values for *enjoyment of learning science* for the OECD countries in order to show the results from the Nordic countries in an international perspective.

The figure consists of two parts, with the upper part showing the index value for both PISA 2006 and PISA 2015, while the lower part of the figure shows the difference between the index values in these two studies (PISA 2015 minus PISA 2006). Positive values therefore show that students responded more positively to these statements in 2015. We think it is important to see these two figures together. A positive value of the index shows that the country's students enjoy science more than the average for students in the OECD.

Figure 1 Index of students' enjoyment of learning science for PISA 2015 and PISA 2006

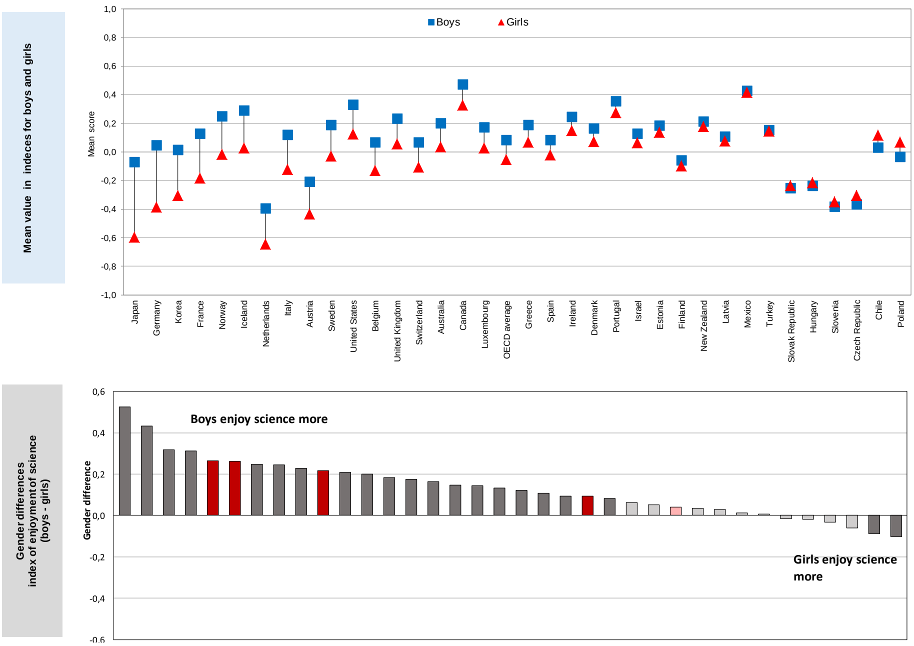


Note: Results are given for OECD countries. Statistically significant differences are shown in a darker tone. Countries are ranked in descending order of the change in the index of students' enjoyment of learning science between 2006 and 2015. The Nordic countries are highlighted in red.

The bottom part of figure 1 shows that in 16 of the OECD countries, the students' enjoyment of learning science improved significantly. The greatest increase of the index was in Ireland, followed by countries such as Poland, the US, the UK, and Canada.

However, when looking in the first part we see that despite the significant increase, the actual index value for Poland in PISA 2015 was still lower than in many of the other countries. In Portugal, there was no change between the two assessments, but the index value was still very high in both assessments compared to most other countries. The largest decreases were in Hungary, the Czech Republic, and Slovakia. Students in the Nordic countries, with the exception of Finland, reported becoming more interested in science. We will discuss the Nordic results in greater detail below.

Figure 2: Enjoyment of learning science in PISA 2015 by gender



Note: Results are given for OECD countries. Statistically significant differences are shown in a darker tone. Countries are ranked in descending order of the change in the index of students' enjoyment of learning science for boys minus girls. The Nordic countries are highlighted in red.

Figure 2 shows the index of *enjoyment of learning science* for girls and boys separately. This figure also consists of two parts. The upper part shows the index values for boys and girls in PISA 2015, and the lower part shows the differences between boys and girls. Positive values on the lower part mean that boys responded more positively than girls to the statements in this construct. Figure 2 clearly shows that boys reported enjoying learning science more than girls in most of the OECD countries. The greatest gender except Finland. Only in Poland and Chile did girls report enjoying science significantly more than boys.

In the following, we will focus more closely on the Nordic perspective and start with the same index about enjoyment that was described above. Table 1 shows the results for the index of *enjoyment of learning science* in the Nordic countries.

Table 1: Results for the index *enjoyment of learning science*. Changes from PISA 2006 to PISA 2015, gender differences, and changes in the science score per unit of this index

	Index of enjoyment of science				Change in science index	Gender differences		Change in the science score per unit of this index	
	PISA 2006		PISA 2015		2015–2006	PISA 2006	PISA 2015	PISA 2006	PISA 2015
	Mean	S.D.	Mean	S.D.		Boys-girls	Boys-girls		
Denmark	−0.07	0.98	0.12	1.14	*0.19*	0.01	0.09	*31*	*26*
Finland	0.11	0.89	−0.07	1.01	*−0.19*	*−0.18*	0.04	*32*	*30*
Iceland	−0.03	1.13	0.15	1.26	*0.18*	*0.20*	*0.26*	*40*	*24*
Norway	−0.01	1.08	0.12	1.20	*0.12*	*0.24*	*0.27*	*35*	*29*
Sweden	−0.10	1.04	0.08	1.26	*0.18*	0.01	*0.22*	*33*	*27*
OECD avg.	−0.00	1.03	0.02	1.17	*0.01*	*0.07*	*0.13*	*30*	*25*

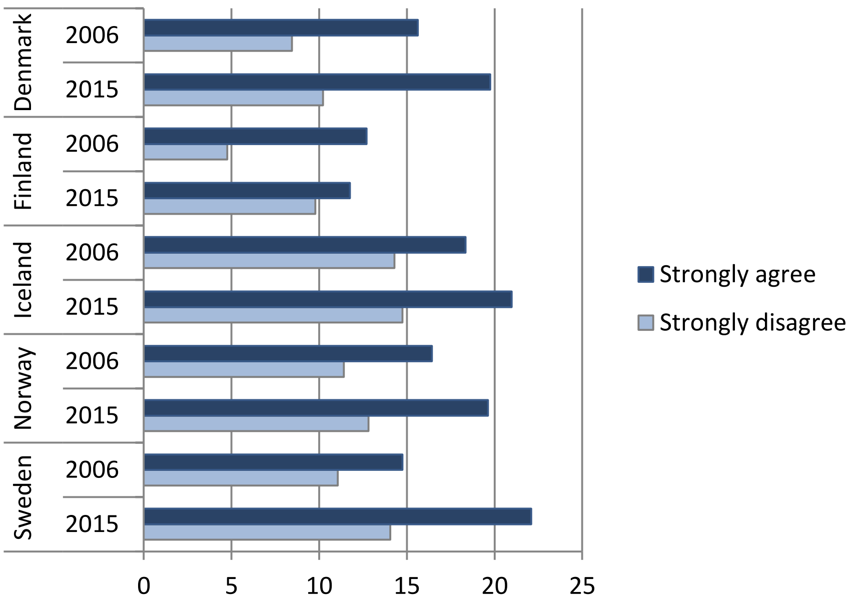
Note: Statistically significant values are indicated in italic. Standard error of the index for the Nordic countries is 0.02 except for Sweden where it is 0.03.

As we already have seen, the results show that students in the Nordic countries, except in Finland, reported higher enjoyment of learning science than the average in the OECD countries. In these countries, the students reported a significantly greater enjoyment of learning science in 2015 than in 2006, while there was a significant decrease in Finland. The results clearly show that enjoyment of learning science was positively related to the science score. In 2006, the change in science score per unit of this index was high for all the Nordic countries and was highest for Iceland. In 2015, the relationship with performance was weaker in all Nordic countries, and the decrease was particularly evident in Iceland, although the coefficients were still positive and significant. Enjoyment of learning science seems therefore to be an important aspect for learning science.

In all of the Nordic countries, the boys on average expressed a higher enjoyment of learning science compared to the girls, except in Finland where there was no significant difference. The gender differences were greatest in Norway, Iceland, and Sweden.

Table 1 also shows that the standard deviation was greater in 2015 than in 2006, which means that there was an increase in the variance of this index for the Nordic countries and on average for the OECD in 2015 compared to 2006. To determine if more students chose the two extreme categories of "strongly disagree" and "strongly agree", we made an investigation of students' answers to all the individual questions in this construct for all response categories. Figure 3 illustrates this, showing the percentage of students who strongly agreed and strongly disagreed, respectively, with the statement *I am interested in learning about science.*

Figure 3: Results for the statement "I am interested in learning about science" showing the percentage of students who answered "strongly agree" and who answered "strongly disagree". Statistically significant differences are described in the text



In all of the Nordic countries, except in Finland, more students responded "strongly agree" to all statements in 2015 than in 2006. The differences were statistically significant. The numbers of students who responded "strongly disagree" also increased in the Nordic countries but the differences were statistically significant only in Sweden and in Finland.

4.3.2 Instrumental motivation to learn science

Instrumental motivation to learn science is defined as students' motivation to learn science because it is useful for them and for their future education and careers (Wigfield & Eccles 2000). The statements in PISA are designed to measure the extent to which the students perceive science as relevant and useful for them in connection to their expectations about educational careers. The statements are:

- Making an effort in my school science subject(s) is worth it because this will help me in the work I want to do later on.
- What I learn in my school science subject(s) is important for me because I need this for what I want to do later on.
- Studying my school science subject(s) is worthwhile for me because what I learn will improve my career prospects.
- Many things I learn in my school science subject(s) will help me to get a job.

The students could respond "strongly agree", "agree", "disagree", or "strongly disagree" to these statements.

Table 2: The index of instrumental motivation to learn science. Changes from PISA 2006 to PISA 2015, gender differences, and the change in the science score per unit of this index

	Index of science self-efficacy				Change in science index	Gender differences		Change in the science score per unit of this index	
	PISA 2006		PISA 2015		2015–2006	PISA 2006	PISA 2015	PISA 2006	PISA 2015
	Mean	S.D.	Mean	S.D.		Boys-girls	Boys-girls		
Denmark	0.04	0.97	0.04	1.00	0.00	*−0.08*	−0.03	*18*	*12*
Finland	−0.22	0.89	0.16	0.92	*0.37*	*−0.13*	−0.04	*31*	*18*
Iceland	0.09	1.11	0.22	1.04	*0.14*	*0.09*	0.03	*28*	*9*
Norway	−0.16	0.95	0.11	0.94	*0.27*	0.02	−0.05	*22*	*13*
Sweden	−0.05	0.98	0.26	0.97	*0.31*	0.02	0.04	*26*	*14*
OECD avg.	0.01	0.98	0.14	0.98	*0.13*	*0.02*	*0.04*	*18*	*9*

Note: Statistically significant values are indicated in italic. Standard error (S.E.) of the index for the Nordic countries is 0.02.

The results in Table 2 show that there has been a noteworthy and significant increase in instrumental motivation from PISA 2006 in all of the Nordic countries except in

Denmark where the result was unchanged. Hence, students in the Nordic countries, except Denmark, seem to have perceived science as being more useful in PISA 2015 compared to how the students in PISA 2006 responded to the same statements. Gender differences for the index were small and not statistically significant in any of the Nordic countries. Instrumental motivation for science had a weak positive correlation with the science score in the Nordic countries and for the average in the OECD, especially when compared to 2006.

4.3.3 Self-efficacy in science

Science self-efficacy is defined as the extent to which students believe in their own ability to handle science tasks effectively and to overcome difficulties. For each statement in this scale, students are asked to rate whether they "can do this easily", "do it with a bit of effort", "would struggle to do it on their own", or "couldn't do it". The questions are expressed as follows:

- Recognize the science question that underlies a newspaper report on a health issue.
- Explain why earthquakes occur more frequently in some areas than in others.
- Describe the role of antibiotics in the treatment of disease.
- Identify the science question associated with the disposal of garbage.
- Predict how changes to an environment will affect the survival of certain species.
- Interpret the scientific information provided on the labelling of food items.
- Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars.
- Identify the better of two explanations for the formation of acid rain.

Table 3 summarizes the index, gender differences, changes from 2006 to 2015, and the correlation between the index and students performance in science.

Table 3: Index of science self-efficacy, gender differences, changes from 2006 to 2015, and changes in the science score per unit of this index

	Index of science self-efficacy				Change in science index	Gender differences		Change in the science score per unit of this index	
	PISA 2006		PISA 2015		2015–2006	PISA 2006	PISA 2015	PISA 2006	PISA 2015
	Mean	S.D.	Mean	S.D.		Boys-girls	Boys-girls		
Denmark	−0.08	1.02	0.08	1.24	*0.16*	*0.24*	*0.37*	*41*	*22*
Finland	0.02	0.93	−0.04	1.18	*−0.07*	*0.10*	*0.26*	*41*	*23*
Iceland	0.14	1.16	0.24	1.52	*0.11*	*0.34*	*0.50*	*38*	*15*
Norway	0.12	1.06	0.19	1.27	*0.07*	*0.18*	*0.23*	*33*	*22*
Sweden	−0.07	1.03	0.05	1.29	*0.12*	*0.21*	*0.39*	*38*	*21*
OECD avg.	0.00	1.04	0.04	1.30	*0.04*	*0.12*	*0.20*	*38*	*18*

Note: Statistically significant values are indicated in italic. Standard error of the index for the Nordic countries is 0.02 (0.03 for Sweden) in PISA 2006.

There was a statistically significant increase in this index for all countries except Finland, which means that the students in the four countries had gained confidence in their ability to answer scientific problems. There was also a growing difference between students in general. Table 3 shows that the standard deviation was greater in 2015 than in 2006. That means that there was an increase in variance of this index for the Nordic countries and on average for the OECD countries in 2015 compared to 2006, meaning increased differences between different students' answers. There was furthermore a large gender gap in this index, which means that girls reported feeling it hard to deal with science-related topics. There was a positive correlation between the index of *self-efficacy* and science scores for the Nordic countries and for the OECD on average, but as for the other two constructs, the correlations were weaker in 2015 compared to 2006.

4.3.4 Enjoyment, motivation, self-efficacy, and science performance

So far, we have described the three interest/motivation variables one by one and looked at the relationship with performance without considering the impact of other variables simultaneously. It is, however, reasonable to expect that these variables tap partly the same constructs and share common variance. To explore how the different interest variables might be related to each other and their relative importance in predicting science performance when modelled together rather than in isolation, as well as to look for possible changes in these relationships over time, we performed a correlation analysis and a multiple regression analysis.

The correlation analysis suggested that all three interest/motivation variables were positively and significantly related to each other in both 2006 and 2015. Hence, a student who enjoyed science was more likely to believe that he or she was competent in science and was more likely to be motivated to learn science. However, the strength of the correlations was low to moderate. As a background for interpreting these results, have in mind that two completely unrelated variables have a correlation of 0 and two perfectly related variables have a correlation of 1. The correlations between the different interest indices for the five Nordic countries for PISA 2006 and PISA 2015 ranged from $r = 0.2$ for the correlation between *instrumental motivation* and *science self-efficacy* in Denmark in 2015 to $r = 0.6$ for the correlation between i*nstrumental motivation* and *enjoyment of learning science* in Iceland in 2006. The pattern of correlations was similar for all Nordic countries, with somewhat weaker relationships between *science self-efficacy* and *instrumental motivation*, and somewhat stronger relationships between *enjoyment of learning science* and *self-efficacy* and between *enjoyment of learning science* and *instrumental motivation*. The pattern of correlations also looked similar over time, although correlations were generally lower in 2015 compared to 2006.

We next performed a multiple linear regression analysis to determine if and to what extent the different interest indices still predicted performance (as suggested by Tables 1–3) when modelled together and to study whether there were similarities between 2015 and 2006. The results of this analysis suggested that in both 2015 and 2006 in all Nordic countries *enjoyment of learning science* and *science self-efficacy* were positively and significantly related to science performance; however, a pattern could be seen where the effect of *self-efficacy* was weaker in 2015 than in 2006. In contrast, the effect of *instrumental motivation* on science performance was weak and in several cases not significant when controlling for the other two variables (Table 4). The pattern was similar for all Nordic countries in both 2015 and 2006. Overall, however, the model explained less of the variation in performance in 2015 compared to 2006. The amount of variance in performance that could be accounted for by the regression model was smaller for 2015 compared to 2006, as evidenced by the R^2 values, and this decrease was particularly visible in Iceland. Thus, even if students in general reported more interest and motivation for learning science, there might be aspects of this increased interest that were not as strongly related to performance in 2015 compared to 2006, and lower-performing students might, for example, have reported high levels of self-efficacy without this being evidenced in their performance on the PISA test, and there might be other variables that are also important for explaining the science performance of modern youth.

Table 4: Regression coefficients and amount of variance explained (R2), science performance as the dependent variable. Results from PISA 2006 and PISA 2015 for the Nordic countries

Predicted variable: *Science performance*	*Instrumental motivation*		*Enjoyment of learning science*		*Science self-efficacy*		All variables R^2	
Country and year	*2006*	*2015*	*2006*	*2015*	*2006*	*2015*	*2006*	*2015*
Denmark	0	−1	*14*	*19*	*34*	*15**	.22	.15
Finland	*13*	3*	*12*	*22**	*32*	*16**	.24	.14
Iceland	2	0	*25*	*20*	*24*	*8**	.28	.14
Norway	0	−4	*24*	*24*	*22*	*14**	.21	.17
Sweden	7	−2*	*15*	*23**	*27*	*13**	.21	.15

Note: Statistically significant regression coefficients are indicated in italic. When the difference between the coefficients within a country is significant between 2006 and 2015, this is indicated with an asterisk on the 2015 value.

Italic = $p < .01$, meaning that the value is statistically significant at the 1% level.

* = change in coefficient between 2006 and 2015 is significant.

The values in Table 4 are b-coefficients from the regression analysis, together with R^2, which is a measure of how much of the variation in the data can be explained by the regression model. The b-coefficients can be interpreted as follows. A one-step increase of the value in the respective interest index (for example, *enjoyment of learning science*) will result in a performance increase corresponding to the value of the b-coefficient. An increase in the *enjoyment of learning science* index in Denmark in 2006 from 0.5 to 1.5 will have a positive effect on science performance, which will increase the PISA score by around 14 points according to the model (an increase of 30–40 points on the PISA test is generally regarded as corresponding to one additional year of schooling). An increase in the *instrumental motivation* index in Denmark 2006, on the other hand, will not have any significant effect on science performance according to the model. Therefore, even if *instrumental motivation* is significantly related to performance when other interest variables are not accounted for, this relationship disappears when controlling for the other interest variables. This result was obtained for all Nordic countries and for both years, except for Finland in 2006. We also tested whether the regression coefficients were significantly different ($p < .01$) between the two years and in all Nordic countries, and the b-coefficient for *science self-efficacy* was significantly smaller (although still significant) in 2015 than in 2006. For Sweden and Finland, there were also significant changes in the coefficients for the variables *instrumental motivation* and *enjoyment of learning science*, while this was not the case in Denmark, Iceland, or Norway. Thus, in the Nordic countries, in particular the effect of *science self-efficacy* on science performance was significantly lower in 2015 than in 2006. At the same time, students in all Nordic countries except Finland reported higher levels of *science self-efficacy* in 2015 (Table 3).

In summary, a student who enjoys science and trusts in their ability to solve science-related questions (self-efficacy) tends to score better on the PISA test. This was true in 2006 as well as in 2015.

4.3.5 Science-related career expectations

Over the past several years, there has been an emphasis on the importance of recruiting more individuals to education and careers in science and technology. Therefore, the PISA background questionnaire asks about the students' expectations about careers in relation to science and technology. In both PISA 2006 and PISA 2015, the students were asked to respond to an open-ended question: *What kind of job do you expect to have when you are about 30 years old?*

The students' responses were given in their own words and could be any job title or description. All responses were classified according to the International Standard Classification of Occupations (ISCO-08). However, many students at this age are unsure of what profession they expect to have, and many students did not answer or indicated that they were undecided. In this context we focused only on student responses that were well-defined expectations of a science-related career, defined as those career expectations that require the study of science beyond compulsory education. These responses were categorized into the following major groups: science and engineering professionals; health professionals; science-related technicians and associated professionals; and information and communication technology professionals (see OECD, 2016a, Annex A1 for more details).

Figure 4: Percentages of students who in 2015 expect to work in different science-related professional and technical occupations when they are 30 years old. Results based on students' self-reports

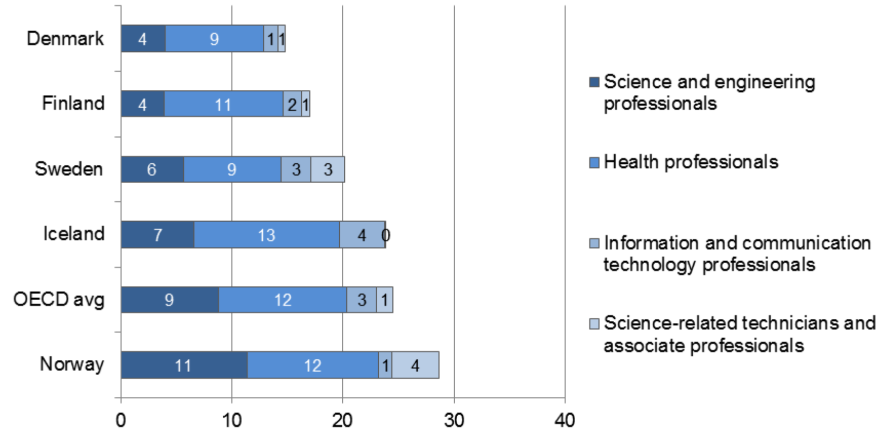
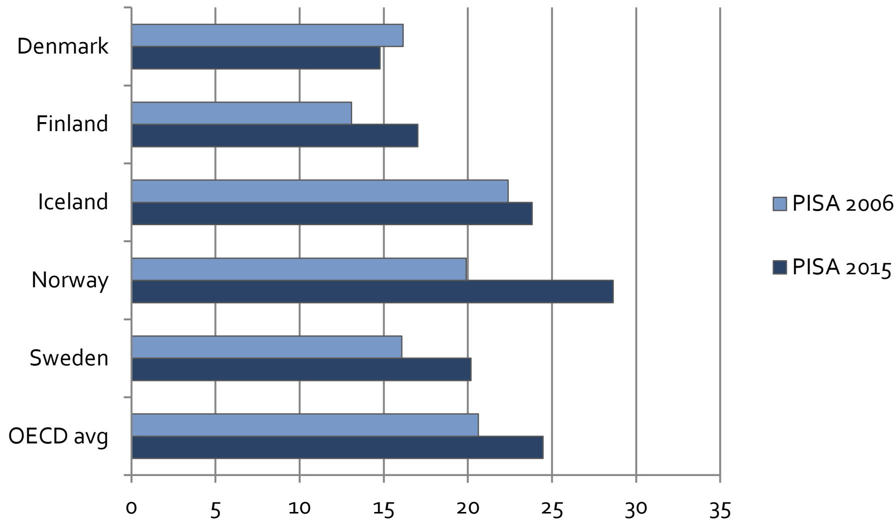


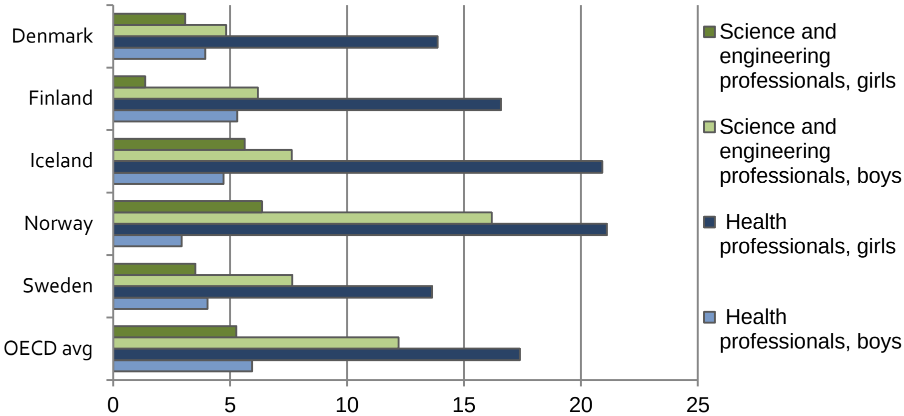
Figure 4 shows that on average almost 24% of the students across the OECD countries reported in PISA 2015 that they expected to work in a science-related occupation. Although all 15-year-olds do not have such clear ideas as to what they want to work with in the future, the expectation of a future with relation to science and technology might have an influence on students' commitment to science learning in the actual situation in the science classroom. One interesting finding is that relatively few students reported that they were aiming for a job as information and communication technology professionals. This might be because the ISCO codes are quite detailed in this area, and many students answer more generally such as "engineer" and not a specific occupation like "computer scientist", "software developer", "applications programmer", etc. Another plausible explanation might be that students' interest in information and communication technology is more related to entertainment than to a future job.

Figure 5: Percentage of students in PISA 2006 and PISA 2015 who expected to work in science-related professional and technical occupations when they are 30 years old. Results based on students' self-reports



The percentage of students who expected to work in science-related professional and technical occupations when they are 30 increased from PISA 2006 to PISA 2015 in all Nordic countries except Denmark (Figure 5).

Figure 6: Percentage of students expecting to work as science and engineering professionals and health professionals by gender in 2015



There were large gender differences in students' occupational preferences. Figure 6 shows that there in PISA 2015 were gender differences in students' choices of science careers in all of the Nordic countries. Boys more than girls expected to choose careers as engineers, and girls more than boys preferred health education. The same gender stereotype pattern was seen in PISA 2006 (Kjærnsli & Lie 2011).

4.3.6 Enjoyment, motivation, self-efficacy, and science career.

To further explore how the different interest variables (*enjoyment of learning science*, *instrumental motivation*, and *science self-efficacy*) impact the likelihood that a student indicated that he or she imagined a science-related career, a logistic regression was run with the interest indices as independent variables and interest in a science career as the dependent variable. Because the career variable only had two values (yes or no/undecided), ordinary least squares linear regression was not suitable and therefore logistic regression, which is a more proper analysis for this kind of outcome variable, was performed. Even if the method was somewhat different from the regression analysis presented earlier, the purpose was the same – to determine the impact of different independent variables (the interest indices) on a dependent variable (in this case future occupation in science). The results of this analysis show that in contrast to science performance, and not very surprisingly, *instrumental motivation* seemed to be a highly important variable with regard to students imagining a future science-related occupation. Students scoring high on *instrumental motivation* were much more likely to indicate they imagined a science career when they are in their 30s than students scoring low on *instrumental motivation*. The values in Table 5 are odds ratios (not the actual regression coefficients, but derived from these and shown in the table for the sake of simplicity), and these can be interpreted as follows. A value of 1 indicates that regardless of the value for the interest variables, the students were as likely to indicate a future scientific-related occupation. An odds ratio of 2, on the other hand, indicates that with a one-step increase in the interest variable, the student was twice as likely to indicate a future science-related occupation. For now, it can suffice with reading the table as follows. Significant coefficients above 1 mean that with a higher value on the interest index, the student was more likely to indicate that he or she considered a science-related occupation in the future. In line with findings from the previous regression analysis, the impact seems to have been stronger in 2006 than in 2015.

Table 5: Odds ratios from logistic regression with interest variables as independent variables and career expectation as the binary dependent variable for PISA 2006 and PISA 2015 in the Nordic countries

	Instrumental motivation		Enjoyment of learning science		Science self-efficacy	
Country and year	2006	2015	2006	2015	2006	2015
Denmark	*2.8*	*2.3*	*1.5*	*1.4*	*1.2*	*0.9**
Finland	*2.8*	*2.4**	*1.1*	*1.3*	1.1	1.1
Iceland	*2.6*	*1.5**	*1.5*	*1.4*	1.1	1.0
Norway	*2.2*	*1.8**	*1.3*	*1.3*	1.1	1.0
Sweden	*2.3*	*1.8**	*1.4*	*1.6*	1.0	1.0

Note: Statistically significant coefficients are indicated in italic, and when the difference between the coefficients within a country is significant between 2006 and 2015 this is indicated with an asterisk on the 2015 value.

Italic = p < .01, meaning the value is statistically significant at the 1% level, * = change in coefficient between 2006 and 2015 is significant. Values are rounded to the nearest decimal to ease readability, and this is why the same value might be significant in one cell but not in another.

Both in 2006 and 2015, and in all Nordic countries, both *instrumental motivation* and *enjoyment of learning science* were significantly associated with the likelihood of students indicating a future occupation within the fields of science. *Science self-efficacy* had little impact on the likelihood that the student would be interested in a future career in science, and the effect of this variable was non-significant in all Nordic countries but Denmark (and here the effect was not very strong even if the value was significant), and the odds ratio was around 1 for most countries in both years.

We compared the size of the coefficients across the years, and in all countries except Denmark the coefficient for the *instrumental motivation* variable was significantly weaker ($p < .01$) in 2015 than in 2006, although this variable was still the most important of the variables that were tested. In Denmark, on the other hand, the *science self-efficacy* coefficient was significantly weaker in 2015, although in both years this coefficient was rather weak. For the other variables and the other countries, there were no significant differences between the logistic regression coefficients for 2006 and 2015.

In summary, a student who enjoys science and recognises the value of science for their coming job is more likely to indicate a science occupation when asked what kind of job they think they will have in the future. This can be compared with the previous analyses where a student who enjoys science and trusts in their ability to solve science-related questions (self-efficacy) is more likely to score well on the PISA test.

4.4 Discussion

Students' attitudes and motivation in relation to science are important for their understanding of the role of science and technology in a democratic society. They are also important for the students' choice of profession. It is therefore important to study how different factors such as instrumental motivation, enjoyment of learning science, and science self-efficacy depend on each other and if these variables can predict students' performance and their future science-related career expectations. This chapter will discuss the results and correlations between these variables and science performance in PISA 2006 and PISA 2015 for the Nordic countries and for the OECD countries.

The PISA definition of science literacy recognizes that a student's response to a science-related issue requires more than just skills and knowledge, and it also depends on how able and willing the student is to engage with the issue. The PISA 2015 assessment evaluated students' attitudes towards science in three areas through a questionnaire, including interest in science and technology, environmental awareness, and valuing scientific approaches to enquiry, all of which are considered core to the construct of scientific literacy. This means that the PISA study recognises the need for students to both understand and value science. In a democratic society, it is important that all citizens, independent of gender, age, or ethnicity, are competent to differ between fact and opinions and have the possibility to participate in well-informed discussions about social challenges.

The PISA 2015 data show that students' reported enjoyment of learning science has increased in all Nordic countries but Finland. In all Nordic countries except Denmark, there has also been an increase in students' instrumental motivation to learn science. There was also a smaller increase in students' science self-efficacy. Students in the Nordic countries seem to perceive science as more enjoyable and more valuable than they did just over a decade ago. There has also been an increase in enjoyment in several of the English-speaking countries such as the US, the UK, Australia, Ireland, and Canada. At the same time, the enjoyment was lower in 2015 than in 2006 in many countries in eastern and central Europe like Germany, France, the Netherlands, Slovenia, Slovakia, the Czech Republic, and Hungary. It seems as if students look at science in different ways in different parts of Europe, and one question is what the Nordic countries share with English-speaking countries that might explain the increased engagement in science.

At the same time, the PISA data show increased differences in the Nordic countries between boys and girls in terms of interest, self-efficacy, and career aspirations. This is an interesting finding because the Nordic countries are usually considered to be the most gender equal in the world (OECD, 2016a; Sørensen & Dohn, 2016a). In addition, even though more students reported high self-efficacy and great interest in 2015, there were increasing differences between students and there were a growing number of students reporting low enjoyment. We have on the one hand a growing number of students who are more engaged in school science, and on the other hand a growing number of students who feel more alienated.

The results of our analysis show that there were positive relationships in the Nordic countries between all of the three different interest variables and science performance when related to test score one by one in 2015. The correlations were, however, weaker for all interest variables in 2015 compared to 2006. This means that even though there was an increase in enjoyment of learning science and science self-efficacy, these variables explained less of the results. There seemed to be new patterns in students' answers and a change in the reasons for why the students were interested in science. The regression analysis counted all variables together and showed that *self-efficacy* and *enjoyment of learning science* are important for science performance. Even though *instrumental motivation* was positively related to science performance itself, this effect disappeared when modelled together with enjoyment and self-efficacy. Students with high instrumental motivation performed better compared to lowly motivated students, but this can be explained by higher science self-efficacy and higher enjoyment of learning science among those students.

Expressed career expectations differed a lot among the OECD countries, but in all Nordic countries except Denmark there were more students who expected a science-related career in 2015 compared to 2006. There might be differences in whether 15-year-old students have made up their minds about possible careers. For example, in Denmark the majority of 15-year-old students aim for a common start in upper secondary school, whereas Swedish students of that age choose between different theoretical and vocational programmes. The increase in students considering a science career in most Nordic countries is nevertheless important in the light of recent discussions about recruitment into the fields of science and technology and the concern about shortages of science graduates in the labour market (Gago et al., 2004). Students' instrumental motivation to learn science is clearly associated with an expectation of a science-related career. Enjoyment of learning science also seems to have some importance for students' interest in pursuing a science career. Self-efficacy, on the other hand, appears to have little or no impact on the likelihood that students expect a science-related career.

Although there are some differences across the Nordic countries and over time, the patterns generally look the same and suggest that *enjoyment of learning science* and *science self-efficacy* together are important for achievement in science, whereas *instrumental motivation* and *enjoyment of learning science* together are important reasons behind expecting a science-related job, which is in line with previous research (Areepattamannil, Freeman, & Klinger, 2011; Wigfield & Eccles, 2000). Consequently, all three aspects of interest assessed here appear to be important to monitor and to encourage. If society wants students who perform well in science as well as being interested in a science-related occupation, students' enjoyment of learning science seems especially important.

Another interesting finding is that for the 2015 data the models with all interest variables counted together explained less of the variation in performance and less of the variation in students who want to choose a science-related profession compared to 2006. This is in line with the results above where the change in the science score per unit of each index also decreased between 2006 and 2015. As shown here, there appear to be increasing differences between genders and between those with high and low interest in science. One example is boys who are more interested in science and have higher self-efficacy but who perform on the same level as girls on the PISA test in most of the Nordic countries. There seem to be new components behind students' attitudes that might explain why the models in general explain less of the variation.

4.4.1 Implications

There are positive signs in the Nordic countries when it comes to students' enjoyment of learning science and their science self-efficacy, and there an increased number of students expecting a career in science or technology-related occupations. The large differences between students when it comes to interest, motivation, and career expectations are nevertheless a challenge for modern societies. Research shows that students' interests and career expectations are deeply connected with their identity construction. It is not primarily about what to be, but about whom to be (Schreiner, 2005; Oskarsson, 2011; Teknikdelegationen, 2010). Socialisation into groups with those who think and act alike are strong and are reinforced by social media that is designed to help us to get "likes" and find "friends". Different groups of students seem to perceive science differently, and this could be one important explanation for the increasing differences between genders and between other groups of students that cannot be explained based on the available data, and thus further research is required.

Modern society is experiencing rapid changes with new technology influencing a greater part of our lives. Many jobs that students might have later in life do not even exist today, which means that students must be open to change and new trends in education and the labour market. The emerging picture is that enjoyment of learning science, instrumental motivation to learn science, and science self-efficacy have increased among many students in the Nordic countries. This indicates that the image of science is shifting. In the discussions about sustainability, the environment, and climate change, the focus is in many ways on new technologies like solar panels, windmills, and electric cars. Engineers, inventors, and entrepreneurs behind the brands that produce smartphones, electric cars, rockets, computer games, and social media platforms are well known from the media and in popular culture. In 2017, it was ten years since the iPhone was presented, and smartphones and social networks now provide new opportunities to retrieve information and new ways to hang out with and acquire friends. Students in the Nordic countries as well students in the English-speaking countries are diligent users of computers, smartphones, and social networks (OECD, 2017). All together, these developments might give students easier access to science and to technological achievements and the value of science may be more obvious in the everyday life of young people. This could be one explanation behind what seems like a shift in Nordic students' interest and self-belief in science and thus contribute to an understanding of why more students are thinking that science might be something for them.

4.5 References

Areepattamannil, S., Freeman, J. G., & Klinger, D. A. (2011). Influence of motivation, self-beliefs, and instructional practices on science achievement of adolescents in Canada. *Social Psychology of Education, 14*(2), 233-259. doi:10.1007/s11218-010-9144-9

Bandura, A. (2010). Self-Efficacy. In *The Corsini Encyclopedia of Psychology*. John Wiley & Sons.

Bybee, R., & McCrae, B. (2011). Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. *International journal of Science Education, 33*(1), 7 – 26.

Davidsson, Eva and Helene Sørensen (2010), Naturvidenskab: In Niels Egelund (ed.) PISA 2009 *Danske unge i en international sammenligning Bind 1 – Resultatrapport,* København, DPU.

Gago, J. M., Ziman, J., Caro, P., Constantinou, C., Davies, G., Parchmannn, I., . . . Sjøberg, S. (2004). *Europe Needs More Scientists: Report by the High Level Group on Increasing Human Resources for Science and Technology.* Retrieved from http://europa.eu/rapid/press-release_IP-04-444_en.htm

Jidesjö, A. (2012). *En problematisering av ungdomars intresse för naturvetenskap och teknik i skola och samhälle: innehåll, medierna och utbildningens funktion*. Norrköping: Department of Social and Welfare Studies, Linköping University.

Jidesjö, A., Oscarsson, M., Karlsson, K., & Strömdahl, H. (2009). Science for all or science for some: What Swedish science students want to learn about in secondary science and technology and their opinions on science lessons. *Nordina, 5*(2), 213-229.

Kjærnsli, M., & Lie, S. (2011). Students' preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education, 33*(1), 121-144.

Kjærnsli, M & Jensen, Fredrik (red.) (2016). *Stø kurs, Norske elevers kompetanse i naturfag, matematikk og lesing i PISA 2015.* Oslo, Universitetsförlaget

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger, S. Hidi & A. Krapp (Eds.) *The Role of Interest in Learning and Development* (pp. 3-25). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Krapp & Prenzel (2011) PISA 2006 science. *International Journal of Science Education, 33*, 7-26. DOI: 10.1080/09500693.2010.518644

Lindahl, B. (2003). *Lust att lära naturvetenskap och teknik? En longitudinell studie om vägen till gymnasiet.* (Ph D), Göteborg: Acta Universitatis Gothoburgensis.

Lyons, T. (2006). Different countries, same science classes: Students' experiences of school science in their own words. *International Journal of Science Education, 28*, 591-613. DOI: 10.1080/09500690500339621

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. Perspectives on psychological science, 1(2), 133-163.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Science Report.

Martin, M. O., Mullis, I. V. S., Foy, P., & Stanco, G. (2012). *TIMSS 2011 International Science Report*.

OECD. (2007). *PISA 2006: Science Competencies for Tomorrow's World, Vol. 1 Paris:* OECD.

OECD. (2009). *PISA 2006 Technical report*. Paris: OECD.

OECD. (2016a). *PISA 2015 Results (Volume I)* Paris: OECD.

OECD. (2016b). *PISA 2015 Technical Report*. Paris: OECD.

OECD (2017), PISA 2015 *Results (Volume V): Collaborative Problem Solving*, OECD Publishing, Paris.

Osborne, J. and Dillon, J. 2008. Science education in Europe: Critical reflections. A report to the Nuffield Foundation, London: King's College

Oscarsson, M., Jidesjö, A., Karlsson, K. G., & Strömdahl, H. (2009). Science in society or science in school: Swedish secondary teachers' beliefs about science and science lessons compared with what their students want to learn. *Nordina, 5*(1), 18-34.

Oskarsson, M. (2011). *Viktigt – men inget för mig: ungdomars identitetsbygge och intresse för naturvetenskap.* Linköpings universitet, Norrköping.

Oskarsson, M., Eliasson, N., & Karlsson, K. G. (2017). Verkliga vardagssammanhang i årskurs 4 eller kontextlös kunskap i årskurs 8? Everyday life context in grade 4 or knowledge without context in grade 8. *Nordic Studies in Science Education, 13*(1), 36-51.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education: Theory, research and applications* (2nd Ed.). New Jersey, NJ: Merril Prentice Hall.

Schreiner, C. (2005). *Exploring a ROSE-garden. Norwegian youth's orientations towards science-seen as signs of late modernity.* University of Oslo, Oslo.

Skaalvik, E.M., Valas, H. (1999) Relations among Achievement, Self-Concept, and Motivation in Mathematics and Language Arts: A longitudinal study, Journal of Experimental Education, 67, 135–149.

Skolverket. (2016). *TIMSS 2015: svenska grundskoleelevers kunskaper i matematik och naturvetenskap i ett internationellt perspektiv*. Stockholm: Skolverket.

Sjøberg, S., & Schreiner, C. (2010). *The ROSE project: An overview and key findings*. Retrieved from Oslo:

Sørensen, H. (2008). Piger og drenge svarer forskelligt – hvilke konsekvenser har det for undervisningen. I: Troelsen, R. P. & Sølberg, J. (red). *Den danske ROSE-undersøgelse,* 41- 54. København: Institut for curriculumforskning, Danmarks Pædagogiske Universitetsskole, Århus Universitet.

Sørensen, H. & Dohn, N. B. (2016a). Elevernes holdninger og forventninger i forhold til naturfag. In V. T. Christensen (Ed.), *PISA 2015 Danske unge i en international sammenligning* (pp. 83-110). København: KORA.

Sørensen, H. & Dohn, N. B. (2016b). Om naturfagene i PISA og i folkeskolen. In V. T. Christensen (Ed.), *PISA 2015 Danske unge i en international sammenligning* (pp. 17-46). København: KORA.

Teknikdelegationen. (2010). *Vändpunkt Sverige: ett ökat intresse för matematik, naturvetenskap, teknik och IKT*. Stockholm: Fritzes.

Wigfield, A, & Eccles, J. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield & J. Eccles (Eds), *Development of achievement motivation* (pp 92-120). New York, NY: Academic Press.

Wigfield, A. & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*, 68-81.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American educational research journal, 29*(3), 663-676.

5. PISA, reading literacy, and computer-based assessment

Maria Rasmusson
The Department of Education, Mid Sweden University (and the Department of Education, Uppsala University).

Ulf Fredriksson
The Department of Education, Stockholm University.

5.1 Abstract

The OECD changed the mode of delivery in PISA 2015 from a paper-based assessment (PBA) to a computer-based assessment (CBA). In PISA 2015, four countries improved their results in all three tested domains, and three of these were Nordic countries (Denmark, Norway, and Sweden). Students in the Nordic countries have significant experience in using digital devices, and this might be a plausible explanation for the improved results. In line with this, the results from South Korea and Turkey, two countries with a low average use of computers and the Internet, deteriorated when the test mode was changed. It is of vital importance to ensure that the PISA results from the computer-based test are comparable to previous results if countries are to make policy decisions based on the results. The overarching purpose of the present study is to explore how the change from PBA to CBA was associated with the Nordic students' performance on the reading tasks in PISA in relation to item format, gender, computer experience, and the three reading aspects that were assessed. The analysis was performed on the overall results in reading literacy as well as the specific items that were used in both PISA 2012 and PISA 2015, in all the Nordic countries.

Among the Nordic countries, there were no dramatic changes in the results on reading literacy in PISA 2015 compared to PISA 2012. Only in Sweden was a major

change observed. Our findings show that in comparison with the other Nordic countries, Sweden seems to have a larger group of students who spend large amounts of time on the Internet, and this group improved their results on PISA 2015 compared with PISA 2012 more than other groups in Sweden and more than similar groups in the other Nordic countries. All of the Nordic countries showed a decreased proportion of students who gave no response to items and who did not complete all items in the PISA test on reading literacy. At the same time, the proportion of full-credit responses and no-credit responses increased.

Keywords: Reading literacy, computer-based assessment, paper-based assessment, test mode, gender difference.

5.2 Background

In PISA 2015, the mode of delivery of the test was changed from a paper-based assessment (PBA) to a computer-based assessment (CBA). Is it possible to change the mode of delivery for a test without any implications for the results? This paper sought to determine whether there are reasons to believe that the change of test mode had an impact on the results.

Recent years have witnessed the increased use of computers in all parts of daily life. This has had an impact both on reading habits and on how student achievement in general, and reading skills in particular, can be measured (OECD, 2010a; Skolverket, 2011). Today, much reading takes place on computers, tablets, and smartphones, and this has created a discussion about whether the reading itself has changed as a result of the move from reading texts on paper to reading texts on different types of screens. This discussion centers around two questions: 1) Whether it makes a difference to read on paper or on screen and 2) whether there is a difference between reading texts structured for the type of linear reading that is normally performed when reading printed texts in books, newspapers, journals, etc., or when we read texts structured for non-linear reading on the Internet. The answers to these questions have implications for computer-based testing of reading skills.

The first issue has been investigated in several studies. The results have been inconsistent, however, with some studies indicating that there is no difference and that students' results on reading comprehension tests when reading on paper or on screen are basically the same (Baker, 2010; Kim & Huynh, 2010), while other studies have shown better results on tests for reading on paper compared with reading on a screen (Mangen, Walgermo & Brønnick, 2013; Rasmusson, 2015). If differences between texts in terms of length, structure, and content are taken into consideration, the picture becomes even more complex (Kerr & Symmons, 2006). The studies mentioned above have been performed on student populations that are more or less homogenous in respect of their cultural background and their familiarity with using computers.

The other question about the relation between reading traditional texts on paper and reading texts on the Internet has also been studied. Texts on the Internet are different than printed texts because on the Internet texts, pictures, video, etc., are combined, which creates new types of communication (Bolter & Gromala, 2003). A book is normally read in a linear way from the first to the last page, but an Internet text often has links that make it possible to move in different ways in the text (Kamil, Mosenthal, Pearson, & Barr, 2000). These differences have been explored in studies about reading on the Internet (Coiro, 2003; Kamil, Mosenthal, Pearson, & Barr, 2000; Kress & van Leeuwen, 1996, Aarseth, 1997). Rowsell and Burke (2009) argue that Internet reading demands another type of understanding of the text design. Rasmusson and Eklund (2013) identified four special skills that they consider to be necessary for Internet reading, including multimodal literacy, navigation, digital skills, and information skills. There are several similarities between the two types of reading, but there are also differences. In addition to the skills needed for all reading (decoding and language comprehension), the Internet reader also needs to know how to navigate (Frønes, 2017), which has been defined in an OECD report (2011) as the way in which students move around in a digital text in order to orient themselves and to find the information they need.

There has been an increased interest in what is referred to as e-tests – using computers for tests – and these have made it possible to organize testing in new and more efficient ways. Digitalized tests simplify data collection and the handling of data after that the students have completed the test (Scheuermann, & Guimarães Pereira, 2008), and although there are obvious administrative advantages linked to computer-based testing, an important issue is whether the same skills are tested. As discussed above, there are good reasons to assume that reading a text with hyperlinks on the screen is not identical to reading a printed text on paper. At the same time, studies where the CBA has been performed in a way as close to paper reading as possible show that there might be only small differences between the two modes of reading if it is just a matter of reading on paper and on screen and no navigation is involved (Mangen, Walgermo & Brønnick, 2013; Rasmusson, 2015).

5.2.1 Reading literacy in PISA

The PISA framework for the reading literacy assessment is organized into three aspects of reading that might be regarded as mental strategies. The aspects imply that the students should be able to (1) access and retrieve information in the text, (2) integrate and interpret what they read, and (3) reflect and evaluate, that is to say, to stand back from the text and relate it to their own experience (OECD, 2010b, p. 38). Each reading item is designed primarily to assess one of these three aspects.

The PISA framework refers to a text-display space that can be on paper sheets or on digital screens, and it specifies that texts are typically "fixed" in a paper space but can be "fixed" or "dynamic" in a digital space. The adjective "dynamic" refers to hypertexts, i.e. texts that use navigation tools and certain features that make it possible to use, and even require, non-sequential reading (OECD, 2016b). The PISA 2015 reading test was delivered on paper or computer, but it only used fixed-text formats, and hypertexts that included links or other navigation features were not used. The intention was that the CBA should be comparable with the PBA.

5.2.2 PISA, computer based testing, and reading

PISA has been faced with the challenge of new reading habits and new ways to test reading. Already in PISA 2006 an attempt was made to use CBA. In connection to PISA 2006, where science was the main domain, an opportunity was offered to take a digital science test called Computer Based Assessment of Science (CBAS). Only three countries participated – Denmark, Iceland, and South Korea (Björnsson, 2008). A conclusion drawn from the CBAS was that boys seemed to have an advantage when they were tested using a digital format. A possible explanation given was that the use of computers is more attractive to boys than to girls because of the more dynamic use of videos and animations, which in many cases might be more familiar to boys than girls (Martin, 2008).

In PISA 2009, a digital test of reading was organized, which was referred to as the Electronic Reading Assessment (ERA). In this test, digital texts with hypertexts and navigation instruments were used, and a clear distinction was made between this test and the ordinary PISA reading test that was still on paper. The ERA intended to look at reading of Internet texts, while the ordinary PISA test was considered to test reading in the same way as in earlier PISA tests of reading. This distinction was made in the framework for PISA 2009 (OECD, 2010b).

Traditionally, girls tend to read better than boys, and this was also the case in the digital reading test in PISA 2009. The gap betw een boys and girls in reading, which has been reported in all PISA studies that have included traditional reading, turned out to be smaller when digital reading was tested. Another finding in the ERA was that those who used computers the most were not necessarily those who had the highest scores on the digital reading test (OECD, 2011; Skolverket, 2011). In general, the correlation between digital reading and traditional reading was 0.83. In some countries students performed better on the digital reading test, while in others they performed better on the traditional test. The Swedish and the Icelandic students had significantly better results on the digital test than on the traditional test, while the Danish students had significantly better results on the traditional test. For the Norwegian students, there was no significant difference between the two tests (OECD, 2011).

Digital reading was tested again in PISA 2012 (OECD, 2013; Skolverket, 2013), and the same general observations as in 2009 were confirmed. Among the Nordic countries, Denmark, Norway, and Sweden participated in PISA 2012. While the Swedish average score on the reading test decreased between 2009 and 2012, the average scores for Denmark and Norway remained at about the same level (OECD, 2013).

The PISA study in 2015 was planned to be a CBA. The intention was not to test what has been referred to as digital reading above, but to test traditional reading using computers. The reasons for moving to a CBA was to make the test more similar to the reading habits of young people outside the classrooms and because CBA provides many practical advantages in organizing the test, collecting the data, and working with the data (OECD, 2016c). Because the intention was to make the results comparable with the results from earlier PISA tests, and because it could not be expected that all countries would be able to organize the test on computers, it had to be a test that tested knowledge and skills in the same way as in earlier tests. In the case of reading, this meant that what had earlier been described as reading on paper now had to be tested on a computer screen.

In order to further explore the validity of the test, a field trial was organized in 2014. Half of the students were tested on computers and the other half on paper. The two versions of the test were made as similar as possible. The evaluation of the field trial led to some adjustment of the test, and the PISA Governing Board concluded: "At aggregate levels, the influence of the mode of assessment on student scores is considered to be negligible" (OECD, 2016c, p. 5). It was assumed that it should be possible to regard the PISA 2015 as part of the series of earlier PISA studies.

Jerrim (2018) examined the field trials in Germany, Ireland, and Sweden and argued that the results "show that pupils completing the computer-based test performed substantially worse than pupils completing the paper-based test in all three countries. The difference is most pronounced in Germany (up to 26 PISA points), followed by Ireland (up to 18 PISA points) and Sweden (up to 15 PISA points)" (p. 2). He argued that after they applied the method used to account for mode effects in PISA 2015 "the differences decrease in all three countries. However, there is important heterogeneity in this respect. Whereas no statistically significant differences in performance remain in Sweden, pupils sitting the computer-based test in Ireland and Germany still perform 11 and 15 points lower in science, respectively" (p. 2).

One important constraint in the inquiry of Jerrim, which the author highlights, is that due to limitations in data from the field trial it has not been possible to look at mode effects for specific groups other than boys and girls. This issue is also partly mentioned in the notes from the PISA Governing Board (OECD, 2016c). In an article by Helen Ward (2018), Yuri Belfali from the OECD responds to Jerrim´s paper and argues that due to "the large statistical uncertainty associated with country-specific results, and of the non-representative nature of PISA field-trial samples, conclusions about the influence of the mode of assessment on individual countries' trends should not be drawn from this research."

5.2.3 PISA 2015

In PISA 2015, the test was for the first time given digitally in most of the participating countries. Although measures had been taken to ensure that the results from 2015 would be comparable with the results from earlier PISA studies, some of the results raised questions.

Generally, in PISA 2015, four countries improved their results in all the three tested domains, and among these were three Nordic countries (Denmark, Norway, and Sweden). Among the Nordic countries, the Swedish results raised most questions. The Swedish PISA results improved for the first time in 2015 (OECD, 2016a; Skolverket, 2016) after what can be described as an accelerated decline from 2000 to 2012 (OECD, 2013). Presently, we do not know whether this improvement was an actual change away from a long-term trend or just a break in what might be a continued downward trend. There has been intensive discussion about possible reasons behind this improvement.

While the results in 2015 improved in Sweden, some other countries experienced the opposite. South Korea, which has traditionally been one of the top performers in PISA, had lower results in all three domains in 2015. Turkey, which has had a low average, but an average that has been improving, had lower results in 2015 than ever before. South Korea and Turkey are not the only countries where PISA 2015 showed decreased performance. Generally, 16 out of the 35 OECD countries had poorer results in all domains in PISA 2015 compared with PISA 2012. One difference between the Nordic countries and South Korea and Turkey is that Nordic students generally have more computer experience than students in these two countries.

In more or less all countries that have participated in PISA, the differences between boys and girls in reading decreased in 2015 compared with 2009. This was a break in a general trend towards bigger differences. Obviously, there could be many reasons behind this, but it is difficult to think of any global change that might have had an impact on this development.

These results and discussions related to PISA 2015 raise the question of whether, in spite of the measures taken, there has been an impact on the results due to the change in testing mode.

5.3 Research questions

The overarching question for the study was how the change from PBA to CBA was associated with the Nordic students' performance on the reading tasks in PISA. More precisely, this study asked whether this change was associated with:

- open-ended items versus multiple-choice items;
- boys versus girls;
- students with more or less experience of using computers; and/or
- the three different types of tasks (reflect and evaluate, access and retrieve, and integrate and interpret).

The overarching question and the different aspects of this question were the starting point for the analysis of the data for Denmark, Finland, Iceland, Norway, and Sweden.

5.4 Method

We assumed that a change in test mode could influence the test results and a possible effect of this change could be that those who were used to the new test mode, in this case using computers, found the test easier than in previous PBA. Thus it was reasonable to look at differences between students who have more or less computer experience and at differences between boys and girls. Earlier studies showed that boys tend to use computers to a greater degree than girls (OECD, 2011; Rasmusson & Åberg-Bengtsson, 2015), and if some students find it easier to take the test in its new mode it might also be possible that the number of items to which no response is given and the number of items not reached should decrease.

The method used in the present study was three-fold. The first part was an analysis of the overall results in reading literacy in PISA 2012 and PISA 2015. This analysis included the amount of time the students spent on the Internet in relation to their performance on the reading literacy test as expressed in plausible values, in total, and by gender. The second part was a comparison of the proportion of response categories (full credit, no credit, no response, and not reached) for the sub-samples of students answering the 44 items that were included in both PISA 2012 and PISA 2015. The proportion of responses has been used to analyze differences in performance related to whether the test was based on PBA or CBA. The third part was the same as the second part, but instead of response categories we analyzed items measuring the three reading aspects of reflect and evaluate, access and retrieve, and integrate and interpret. All of the Nordic countries were included in the analysis, namely Denmark, Finland, Iceland, Norway, and Sweden.

5.4.1 *Sample*

The total sample in each country in 2012 and 2015 is described in Table 1. Due to a rotated design,[29] all students did not answer all items (OECD, 2017). In Appendix A, the table shows the number of students who answered each item in the five countries in 2012. The median value of the number of students answering each item in 2012 ranged from 1,061 in Iceland to 2,247 in Denmark. The table in Appendix B shows the sample distributed on the 44 items in 2015. In 2015, the median value of the number of students answering each item ranged from 406 in Iceland to 847 in Denmark, thus fewer students responded to each item in 2015 than in 2012.

Table 1: Total national PISA samples and the numbers of girls and boys, by country

Country	2012			2015		
	n	n girls	n boys	n	n girls	n boys
Denmark	7,481	3,777	3,704	7,161	3,602	3,559
Finland	8,766	4,307	4,459	5,882	2,863	3,019
Iceland	3,508	1,739	1,769	3,371	1,741	1,630
Norway	4,686	2,291	2,395	5,456	2,706	2,750
Sweden	4,736	2,378	2,358	5,458	2,731	2,727

Note: In Finland, in PISA 2012, students with an immigrant background were oversampled.

Table 2: Mean results for reading literacy on PISA 2012 and PISA 2015

	Finland	Norway	Denmark	Sweden*	Iceland
2012	524	504	496	483	483
2015	526	513	500	500	482

Note: * Difference between 2012 and 2015 is statistically significant ($p = .02$)

[29] Rotated design means that the students do not all respond to the test items in the same order, and all students only have a sample of the items in their test. Thus there are different tests with different subsets of items in different orders. To avoid a large number of students not reaching the same items, the items are placed in different parts of the test for different students.

5.4.2 Data

In the first part of the analysis, we used the plausible values in the reading literacy test as a measure of the students' performance. The plausible values are calculated using all of the reading items in PISA 2012 and PISA 2015, respectively. We chose data from 2012 to compare with 2015 in order to minimize the time elapsed between the tests and thus also, at least to some extent, other factors that might influence the test results such as educational reforms or societal changes. As mentioned previously, the PISA surveys in 2012 and 2015 differed with regards to delivery mode, and moreover the tests did not have all items in common. Changes in the students' results between 2012 and 2015 were analyzed by gender. From the student questionnaire, the question "During a typical weekday, for how long do you use the Internet outside of school?" was used as an indicator of digital experience. This question had seven response options, ranging from "no time" to "more than six hours per day". The results from the Norwegian students for this question were missing in PISA 2015, and thus Norway was omitted from the analysis.

In the second part of the analysis, we looked into the items that were included in both surveys. There were 44 reading literacy items in PISA 2012 (OECD, 2013), and all 44 of these were included among the 103 items in PISA 2015 (OECD, 2016a). In the analysis we used the 44 reading literacy items from PISA 2012 and the same 44 items from PISA 2015. Of these, 20 items were multiple choice and 24 items were open-ended (see description in Appendix C). Results for the open-ended question, Children's Future Q10, were missing in Finland in 2012 and thus this item was removed in all countries in order to get comparable results. Results from Summer Job Q06 (open-ended) and Narcissus Q06 (multiple choice) were also left out in the analysis for all countries due to missing results from Iceland in 2012 (see Appendix D). Thus, 22 open-ended items and 19 multiple-choice items remained in the analysis for a total of 41 items. The raw scores on the item level are coded as full credit, no credit, no response, or not reached. "Full credit" means that the student has given an answer that is considered correct. "No credit" means that the student has given an answer that is not considered to be correct. "No response" means that the student has not given an answer. "Not reached" means that the student has not gotten to the item within the time allocated for the test. The difference between "no response" and "not reached" is that a no response answer is followed by an item where the student have given a response, while "not reached" is an item either at the very end of the test or an item with a response followed by other items where the student has not given a response. Due to the rotated design, not all of the 41 items were distributed to all students.

The items in PISA 2012 and PISA 2015 were not coded in the same manner in the two surveys. Thus, the first step in our analysis was to recode the items in order to get comparable variables. In 2012 the students' responses were coded as invalid if they ticked several options in a multiple-choice question with only one right answer. In the CBA in 2015, it was not possible to tick more than one option due to the computerized test design. Consequently, we have chosen to recode the invalid responses in 2012 as "no credit" to get more comparable categories. In PISA 2015, a "missing response prior to a valid response is defined as an omitted response and is treated as a wrong response" in the item response theory (IRT) scaling in order to establish common item parameters across countries and surveys (ETS, 2016, p. 2). This treatment of missing responses in the IRT scaling procedures did not affect the analysis on the item level and was thus not taken into account in the second part of the analysis in the present study.

In the third step of the analysis, we performed the same type of analysis as with the question types described above, but instead with regard to the three reading aspects separately. The reading literacy assessment part was organized into three broad aspects of reading that might be regarded as mental strategies. The aspects imply that the students should be able to *access and retrieve* information in the text, *integrate and interpret* what they read, and r*eflect and evaluate*, in other words, to take a step back from the text and relate it to their own experience (OECD, 2016b). Each reading item is designed primarily to assess one of these three aspects.

5.4.3 Analysis

In the first part, an analysis of the overall results in reading literacy in PISA 2012 and PISA 2015 was performed using the PISA data explorer. This analysis included the students' time spent on the Internet in relation to their performance on the reading literacy test. The second part was a comparison of the proportion of response categories on the item level for those sub-samples of students answering the items that were included in both PISA 2012 and PISA 2015. All student answers to these items received a code for each item (full credit, no credit, no response, or not reached). The proportion of full credit, no credit, no response, and not reached was calculated for each item in each country for 2012 and 2015. For example, in Denmark, 30% of the students' answers to the item South Pole Q02 were coded as full credit in 2012, and in 2015 the corresponding proportion was 43%. The difference in proportion from 2012 to 2015 was thus 13 percentage points. This was a multiple-choice item, and the sum of all percent differences (the changes in percentage points) for the multiple choice full-credit responses was computed for the Danish students. In other words, the percentage points between 2012 and 2015 were calculated for each item and each code (full credit, no credit, no response, and not reached) separately, and thereafter the sum of all item percentage points for multiple-choice and open-ended items was computed by country. The sum of the percentage points for all items measuring each aspect was divided by the number of items in each of the two item formats. The values reported were thus the average change in percentage points per item by item format. This gave us a value for the difference in the average of the percentage points between 2012 and 2015 of answer codes for the multiple-choice items and open-ended items for each country.

The same type of analysis on the item level was also performed for the three reading aspects of reflect and evaluate, access and retrieve, and integrate and interpret. The sum of the percentage points for all items measuring each aspect was divided by the number of items in each of the three aspect categories, and the values reported were thus the average change in percentage points per item by aspect.

5.5 Results

First, the results from the analysis of the overall performance on the reading literacy test, gender differences, and time spent on the Internet in PISA 2012 and PISA 2015 are presented. The second part of this section gives a more detailed analysis of the percentage points for each code for the shared items in PISA 2012 and PISA 2015.

5.5.1 Changes in boys' and girls' results between 2012 and 2015

An interesting approach is to look at how much the average scores for boys and girls changed between 2012 and 2015. In Table 3 it becomes evident that Sweden differed from the other Nordic countries with regards to the boys' difference in performance between the PBA in 2012 and the CBA in 2015. That is, only the Swedish boys showed a significant improvement between 2012 and 2015. Generally, the results of the boys improved in all countries, but only significantly in Sweden. The changes in the girls' results were more modest and also more mixed. While the results for the girls in Sweden and Norway showed some improvements, Finland and Iceland had results in the opposite direction, and the Danish results did not change much at all.

Table 3: Change in average results for boys and girls on the PISA reading literacy test from 2012 to 2015

Country	Boys	Girls
Denmark	8	−1
Finland	10	−5
Iceland	3	−7
Norway	12	6
Sweden	23*	11

Note: *A significant increase p = .0035.

5.6 Internet usage and performance on the reading literacy test

The students answered the question "During a typical weekday, for how long do you use the Internet outside of school?" (IC002801). The results from the Norwegian students on this question were missing in PISA 2015, and thus Norway was omitted from the analysis. In 2012, 61% of the Swedish students, 59% of the Danish students, 57% of the Icelandic students, and 43% of the Finnish students spent two hours or more on the Internet every weekday. In 2015, the corresponding share of students had increased by 17 percentage points in Sweden, 16 percentage points in Finland,10 percentage points in Denmark, and 6 percentage points in Iceland. The Swedish students reported spending more time on the Internet than students in the other Nordic countries. Almost half of the Swedish girls and boys spent more than four hours on the Internet outside of school during a typical weekday in 2015. This can be compared with the Finnish students where only 27% of the girls and 29% of the boys spent this much time on the Internet (see Table 4). The largest increase from 2012 to 2015 in the share of students using the Internet for more than four hours was among Swedish girls where the share increased by 23 percentage points.

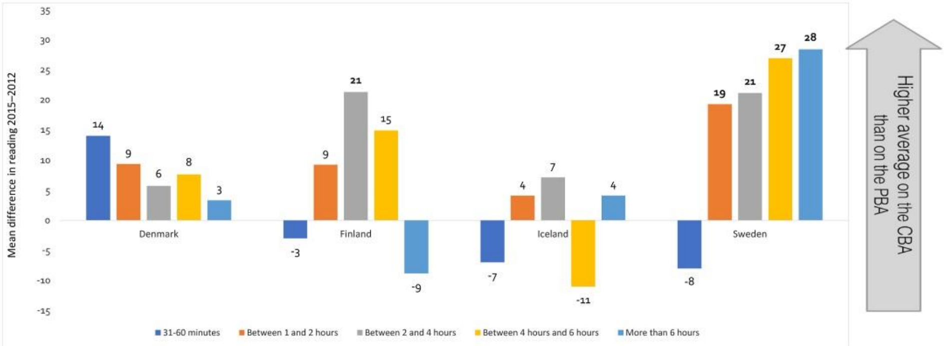
Table 4: Percentages of students per country and gender by time spent on the Internet outside of school

	Denmark		Finland		Iceland		Sweden	
	2012	2015	2012	2015	2012	2015	2012	2015
Girls								
31–60 minutes	11	8	18	12	13	9	13	5
Between 1 and 2 hours	29	24	34	25	30	28	28	14
Between 2 and 4 hours	32	36	28	31	33	31	31	33
Between 4 hours and 6 hours	16	18	7	16	13	17	15	25
More than 6 hours	6	11	3	11	5	10	8	21
Total %	94	97	90	95	94	95	95	98
Boys								
31–60 minutes	10	7	13	11	8	9	8	5
Between 1 and 2 hours	22	17	32	23	26	21	18	14
Between 2 and 4 hours	31	33	33	33	35	32	31	29
Between 4 hours and 6 hours	22	23	12	17	16	20	20	25
More than 6 hours	13	18	6	12	11	14	18	23
Total %	98	98	96	96	96	96	95	96

Note: The response categories "no time" and "1–30 minutes per day" have been left out due to no answers or a very small number of answers, and consequently the total is not 100%.

The results in reading literacy were examined in relation to time spent on the Internet. The Swedish students that used the Internet more than six hours on a typical weekday improved the most on the reading test from 2012 to 2015 among the Nordic countries when the test mode changed from PBA to CBA (see Figure 1). This group of Swedish students that spent a lot of time on the Internet improved on average from 455 points to 484 points on the reading test. The Swedish students using the Internet for four to six hours a day improved the second most from 485 points to 512 points (see table 4). This might indicate that the change in test mode favored Swedish students with a lot of Internet experience.

Figure 1: Differences in averages for PISA reading scores between 2012 and 2015 by time spent on the Internet outside of school. The shares of students were very small in the categories "no time" and "0–31 minutes", and these have been left out of the figure

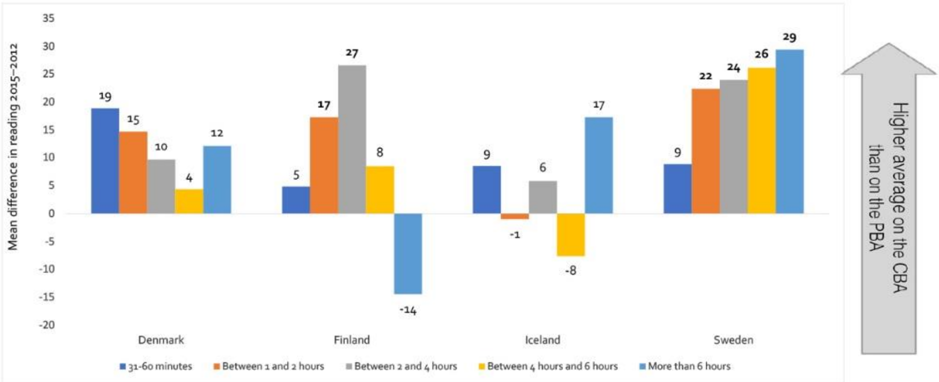


Note: Statistically significant values ($p \leq .05$) are in bold.

However, the Finnish students using the Internet for more than six hours did not improve. This group was smaller in Finland than in the other Nordic countries, and a closer examination of the students' activities on the Internet is called for. Moreover, the students spending a lot of time on the Internet at the time when the PISA assessment was made might still have different amounts of Internet experience from previous years. If, for example, a large share of the Swedish students have been using computers and the Internet for many hours for several years, this might affect their performance in a different way from students in another country who started using the Internet to a large extent more recently, even though they used the Internet to the same extent when they answered the questionnaire. It is well known that the reading performance at this age is affected by all the reading activities over the student's whole life. This group might also have different backgrounds in the Nordic countries. There are small numbers of students in some of the categories, and thus these results should be interpreted with caution.

When the analysis of students' time spent on the Internet was performed for girls and boys separately, it became evident that the largest increase from 2012 to 2015 in performance on the reading literacy test was among the Swedish boys using the Internet for more than six hours (see Figure 2). This group of boys performed better in reading literacy on the CBA in 2015 than the corresponding group on the PBA in 2012. This might indicate that this group benefited from their Internet experience and/or computer habits when the delivery mode changed from PBA to CBA. This group might also have been more motivated when the test was delivered on computer than on paper, and this probably accounted for a part of this improvement in reading literacy.

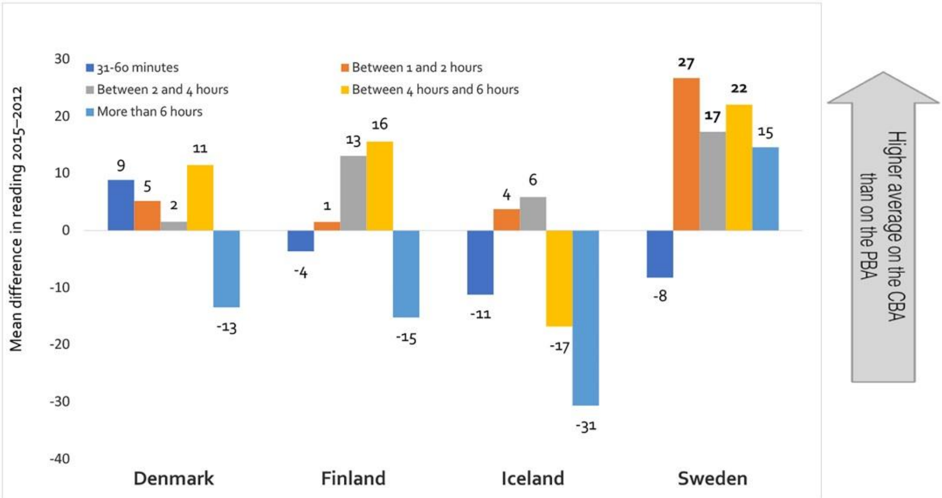
Figure 2: Differences in boys' average scores for the PISA reading scale between 2012 and 2015 according to time spent on the Internet outside of school. Because the shares of students were very small in the categories "no time" and "0–31 minutes", these have been left out of the figure



Note: Statistically significant values ($p \leq .05$) are in bold.

The pattern for the girls is more difficult to interpret (see Figure 3). In Denmark, Finland, and Iceland the girls using the Internet more than six hours performed better on the PBA in 2012 than on the CBA in 2015. In these countries, the groups of girls using the Internet for more than six hours was smaller than in Sweden, and in Denmark in 2012 this was 6% of the girls and in 2015 was 11% of the girls, and the corresponding shares of girls in Finland were 3% and 11% and in Iceland were 5% and 10%. In Sweden, 8% of the girls in 2012 and 21% of the girls in 2015 reported using the Internet for more than six hours. The Swedish girls using the Internet for more than six hours performed better on the CBA than on the PBA. On the whole, time spent on Internet did not seem to be as related to reading performance for girls as for boys in the Nordic countries. This is in line with much other research on differences between boys and girls in terms of reading skills. Girls' results on reading tests seem generally to be less affected by what they read, while boys' results seem to be more dependent on what they read (see for example Asher & Markell, 1974; Scott, 1986; Taube & Munck, 1996). The averages and standard errors are reported in detail in the tables in appendix D.

Figure 3: Differences in girls' averages for PISA reading scale from 2012 to 2015 by time spent on the Internet outside of school. Because the shares of students were very small in the categories "no time" and "0–31 minutes", these have been left out of the figure



Note: Statistically significant values ($p \leq .05$) are in bold.

As touched upon previously, these analyses raise questions that would be interesting to pursue in further studies, for example, whether the Swedish students are using the Internet for activities including reading to a greater extent than the other Nordic countries and whether the improvement from 2012 to 2015 in reading literacy in Sweden was a true improvement of the students reading comprehension and not related to Internet experience or to the change in test mode. Due to limitations in the available information in the PISA data and due to the time frames of the present study, all of these questions will not be possible to answer.

5.7 Analysis on the item level by response format

This section presents the results of the students' answers to the items that were shared in PISA 2012 and PISA 2015 (41 items were included in the analysis, see Appendix C). The proportions of student answers that were coded as full credit, no credit, no response, and not reached were compared between 2012 and 2015. The averages of the percentage points for items, are reported separately for open-ended items and multiple-choice items. This analysis shows whether there was a change in the proportions of codes from 2012 to 2015. The different codes are of course related to each other; if an item has fewer not reached and fewer no-response codes, then it must have a larger proportion of either full credit or no-credit codes or both. It might be suspected that students with a lot of computer experience find it easier to answer items using a computer than pencil and paper, and thus we might have a smaller proportion of no response and not-reached codes on the CBA than on the PBA.

5.7.1 Full-credit responses

A full-credit response is a response considered to be correct. The change in percentage points for full credit responses followed the same pattern for all five Nordic countries regarding multiple-choice items (see Figure 4). The bars in Figure 5 represent the change in full-credit codes from 2012 to 2015 and are the sum of percentage points of all full-credit codes for the open-ended and multiple-choice items, respectively. A positive difference represents a larger proportion of full-credit responses in 2015 than in 2012, and negative bars indicate a higher proportion of full-credit responses in 2012 than in 2015. All five countries had a larger share of full-credit responses in 2015 than in 2012. The largest percentage point increase from 2012 to 2015 for multiple-choice items coded as full credit was found in Finland followed by Sweden, Norway, and Denmark. In other words, there was a larger share of full-credit responses on multiple-choice items on the CBA in 2015 than on the PBA in 2012. This result might be related to a decrease in item difficulty due to test mode, an increase in student motivation due to test mode, or a higher average student proficiency level in 2015 than in 2012. However, out of the five Nordic countries, only the Swedish overall reading literacy results improved in 2015, and thus, the change regarding full-credit multiple-choice responses for these selected items does not reflect a generally higher student proficiency as estimated using the overall plausible values in the other Nordic countries.

The open-ended items followed the same pattern as the multiple-choice items except in Denmark. The Danish students had a larger proportion of full-credit responses on the PBA in 2012 than on the CBA in 2015. Finland had the largest percentage point increase for full-credit responses on multiple-choice items of all the Nordic countries, and the same held true for Norway for open-ended items.

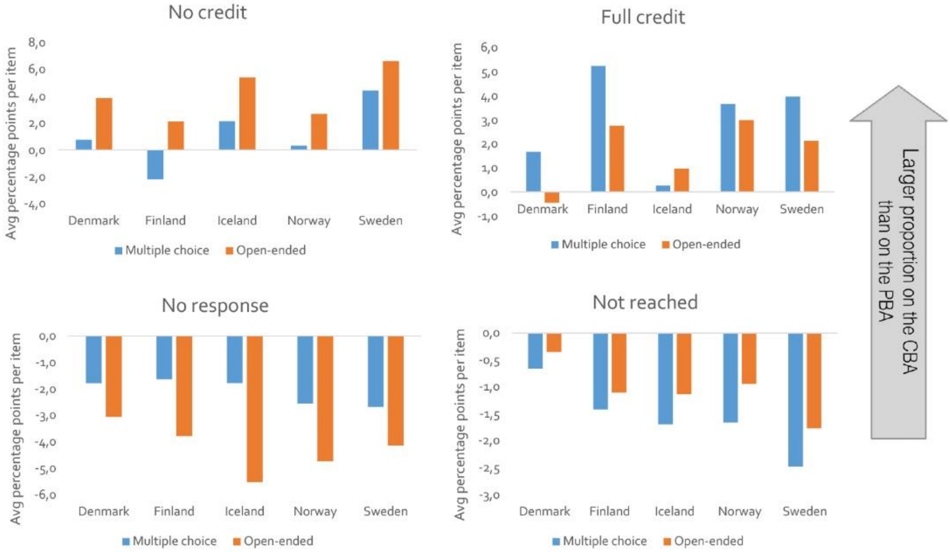
5.7.2 No-credit responses

The responses coded with "no credit" include all items where students have attempted to answer but where the answer was assessed as incorrect. The difference in the share of no-credit responses followed the same pattern for all Nordic countries regarding open-ended items but not multiple-choice items (see Figure 4).

The Finnish students incorrectly answered a larger share of multiple-choice items on the PBA in 2012 than on the CBA in 2015. In Sweden, Norway, Iceland, and Denmark on the other hand, the share of multiple-choice items that received no credit was somewhat larger on the CBA than on the PBA. The students in the latter four countries attempted to answer more multiple-choice items in the CBA than the PBA, although they answered more of these incorrectly. The multiple-choice items might thus be perceived as easier to attempt to answer on a computer than on paper.

The open-ended items followed the same pattern in all Nordic countries, with the largest difference between the PBA and the CBA in Sweden and the smallest in Finland. The share of no-credit answers to open-ended items was larger in 2015 than in 2012. This could be interpreted as a tendency to answer more items, especially open-ended questions, when the test is on a computer than on paper and accordingly with a larger risk of getting an incorrect answer instead of a no-response code.

Figure 4: The mean differences in proportions from 2012 to 2015 as the average of percentage points per item for full credit, no credit, no response, and not-reached codes by item format



5.7.3 No response

This response code is used when the students skip a question for some reason and is only used when there is no visible evidence of an attempt to answer a question. In all countries, a larger share of the answers were coded as "no response" in 2012 than in 2015. The students thus skipped a larger share of items on the PBA than on the CBA, this holds true for both response formats (see Figure 4). Iceland had the largest difference of skipped open-ended items of all the Nordic countries.

Sweden had the largest difference of skipped multiple-choice items of all Nordic countries. In line with the reasoning above for no-credit responses, the Nordic students answered more items when the assessment was computer-based than when it was paper-based, and thus they skipped fewer items on the CBA. When the students answered more items, the proportions of both correct and incorrect responses increased as described above.

5.7.4 Not reached

The last category of responses is the items that the students did not reach. It is only items at the end of a test that can be coded as not reached, and if students have skipped items in the middle of the test these would be coded as "no response". The design in PISA is rotated and thus different students have different items at the end of their test.

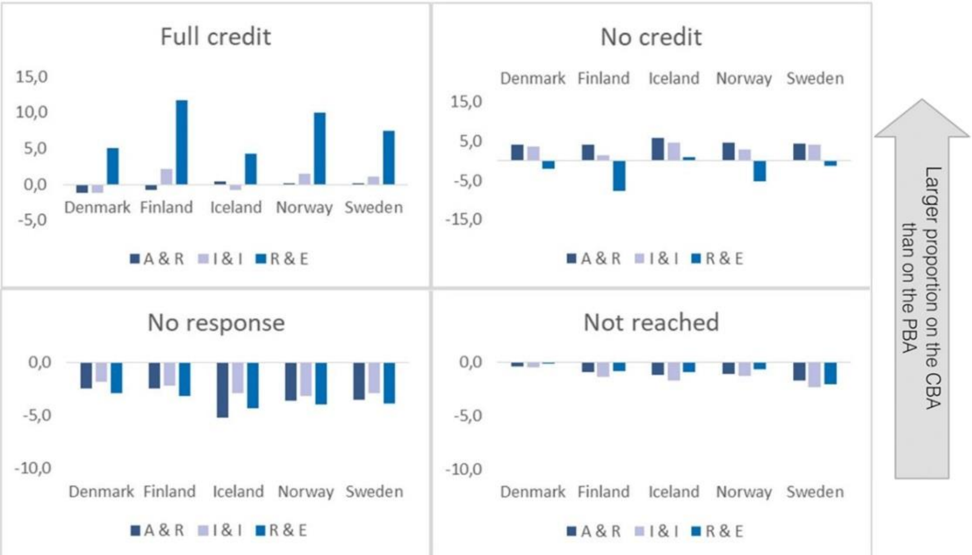
In this category, all Nordic countries had a larger share of not-reached items in the PBA in 2012 than in the CBA in 2015. The Swedish results showed the largest difference in favor of the PBA in 2012. In all Nordic countries, there was a smaller proportion of "not reached" items for the multiple-choice questions in the CBA (see Figure 4). Consequently, a greater number of students finished the test when it was delivered on computer than on paper. Tentatively, it can be assumed that the students in the Nordic countries, who are often very experienced computer users, find it easier to write answers on a computer than with a pencil and paper. It might also be the case that the more experienced Nordic students answer a test on a computer faster than a test on paper and thus a larger proportion of students are able to finish the test when it is delivered on a computer.

5.8 Analysis on item level by reading aspect

The items shared between PISA 2012 and PISA 2015 were analyzed with regards to the reading aspects of reflect and evaluate, access and retrieve, and integrate and interpret (see Appendix C). The items measuring the reflect and evaluate aspect had the largest difference for full-credit answers between the PBA in 2012 and the CBA in 2015 (see Figure 5).

Students in all Nordic countries earned full credit to a greater degree on the reflect and evaluate items on the CBA than on the PBA. The reflect and evaluate items were also answered to a greater degree on the CBA than the PBA. A possible explanation for the improved performance on the reflect and evaluate items might be that these items demanded a more involved answer than other items and the students with a lot of computer experience found it easier to write longer answers using a keyboard than pencil and paper.

Figure 5: The mean differences in proportions between 2012 and 2015, the average of percentage points per item, for full credit, no credit, no response, and not reached codes by reading aspect. A & R, access and retrieve; I & I, integrate and interpret; R &E, reflect and evaluate



5.9 Discussion

The purpose of this article was to explore how the change from PBA to CBA in PISA might be associated with the Nordic students' performance on the reading test. We assumed that such an association with student performance could be explored in two ways. When a new test mode is used, one possible effect could be that those who are used to that mode, in this case using computers, will find the test easier than using the earlier mode. This gives a reason to look at differences between students who have more or less computer experience and at differences between boys and girls. If some students find it easier to take the test in its new mode, it might also be possible that the number of items to which no response is given and the number of items not reached would decrease. The students' response patterns might also be analyzed in relation to open-ended items and multiple-choice items and in relation to the three different types of tasks (reflect and evaluate, access and retrieve, and integrate and interpret) that require written answers.

Generally, we can see that when the results on the reading literacy tests in PISA 2012 and 2015 are compared for boys and girls there are no statistically significant changes between the years except for the Swedish boys who improved on the CBA in 2015.

When students with more or less experience of using computers were compared in the Nordic countries, there was no obvious pattern (see Figure 1). Independently of the time spent on the Internet, most of the groups in Figure 1 showed average results from 2015, which were better than those from 2012, but in Finland those who spent the most time and the least time on the Internet actually had better results in 2012. The Icelandic results were better in 2012 for those who spent less time on the Internet and for those who spent 4 to 6 hours a day on the Internet. In Denmark, all students had better results in 2015, but those who spent the least time on the Internet seem to have improved their results the most. However, in Sweden all groups spending from one to more than six hours per day on the Internet had significant differences between the two tests. Among the other countries, only one of the groups (1–2 hours) in Finland had a significant change. The assumption that those who spent the most time on Internet would have improved their results the most seems to be valid only in Sweden.

In the next step in the analysis, when time spent on the Internet was split up and analyzed separately for boys and girls, the pattern for the boys was much the same as the general pattern described above. In Sweden, the boys who spent the most time on the Internet were those who improved their results the most, while in Denmark it was almost the other way around (see Figure 2). The results from Finland and Iceland did not seem to follow any obvious pattern. The results for the girls (see Figure 3) were even more incongruent. It is difficult to see any general pattern between the Nordic countries, and it is even difficult to find a pattern within the countries due to the scattered results. Significant differences can only be found among the Finnish boys and for both boys and girls in Sweden. One conclusion that can be made is that when the students were split up based on gender and time spent on the Internet the group that showed the biggest improvement between 2012 and 2015 were the boys in Sweden who spent the most time on the Internet. Those who showed the biggest decline in results were the girls in Iceland who spent the most time on the Internet.

The results for the Swedish boys confirmed the assumption that those who spend the most time on the Internet are those who benefitted the most from the change of test mode, while the results from the other countries and from girls in general do not support this assumption. When this comparison is made, it should also be kept in mind that the group of boys who spend more than 6 hours a day on the Internet is not equally large in all countries. In Sweden, 23% of the boys belong to this group, while in the other countries the similar groups are about 10 percentage points smaller than the Swedish group (see Table 3). The same is also true when girls from the different countries are compared. Obviously, the group who spend much time on the Internet is larger in Sweden than in the other countries, but this does not explain why the pattern of more time spent on the Internet and higher scores on the CBA than on the PBA is only seen in Sweden.

The reason for why boys in Sweden who spend 6 hours or more a day on the Internet had more improved results compared to the other groups might be related to their motivation to use computers, but it could also be related to this group finding it easier to take tests on computers than with paper and pen. What is interesting to note is that we do not see the same pattern in the other Nordic countries. If motivation plays a role, just the use of computers cannot be the only reason for the motivation of the Swedish boys. If that would have been the case, we should have seen more similar patterns in at least some of the other Nordic countries. If we believe that motivation has been an important reason for the improved results, the use of computers could have contributed to that, but it cannot be the only reason for the increased motivation.

Obviously, the measure used for this analysis is not the best. What really is of concern is not the time spent on the Internet, but the students' experiences of reading on a computer. We do not know what the students do on the Internet. Their activities could be more or less oriented towards reading, and there could be systematic differences in Internet use between the students in the different countries that we do not know about. We are also missing information about other computer-based activities. Because the question about time spent on the Internet was the best available questions that was asked in both PISA 2012 and PISA 2015, this is what can be analyzed with the support of the available PISA data.

The analysis of the type of responses is to some extent easier than the analysis above. Generally, the percentage of no responses and items not reached decreased when PISA 2015 was compared with PISA 2012 (see Figure 4). This means that the proportions of responses given a full credit and given no credit increased. This seems to be true both for multiple-choice items and open-ended items in Sweden, Norway, and Iceland (see Figures 4 and 5), while in Denmark the percentage of full-credit responses to open-ended items decreased slightly and in Finland the percentage of no-credit responses to multiple-choice items decreased.

When the percentages of full-credit responses for multiple-choice items are compared with the full-credit responses for open-ended items, it can be seen that the percentage has increased in all countries, with the exception of Denmark, for multiple-choice items (see Figure 4). In the case of the no-credit responses, it is the other way around – the percentage of open-ended responses increased more than the responses to the multiple-choice items. This is probably related to it being easier for students to respond both to multiple-choice items and open-ended items on a computer, but the greater readiness to give a response might more easily pay off when they only need to indicate an answer from a list of choices. When they have to write an answer themselves, the readiness to just write something might in many cases not be enough to get a full credit. A possible explanation is that the students find writing on a computer easier than writing with a pencil.

This development, although it is not equally strong in all five countries, seems to indicate that the students to a higher degree have responded to the items and managed to respond to more items on the CBA than the PBA. When more items are answered, it is obvious that the percentage of full-credit responses and no-credit responses will increase. This development might in most cases lead to higher scores. If only some of the responses given are correct, that will in most cases lead to a higher score. Even if none of the responses are correct when the number of responses increases, this will not lead to lower scores. The only possible scenario in which an increased response rate could actually lead to lower scores would be if less time is spent in general on the items in the test and as a result the responses on items that earlier had a high percentage of full-credit responses would then have a lower percentage of full-credit responses. This situation was not explored in this study.

When the three different types of reading aspects (reflect and evaluate, access and retrieve, and integrate and interpret) are analyzed separately, the clearest result is that the proportion of full-credit answers to tasks that demand the students to reflect and evaluate increased in all countries. At the same time, the results do not seem to indicate that the students responded to or managed to perform reflect and evaluate items to a greater degree than the other two aspects. The reflect and evaluate items can be regarded to be to some extent more demanding than the other two aspects, and the students need to express in writing their reflections about a text and/or an evaluation of the text. One possible explanation could be that when students find it easier to write an answer on a computer, they write longer answers to these questions and the likelihood to get credit for the answer will then increase, but because we have not explored the length of the answers this is still just an assumption.

One of the findings is that Sweden is sticking out in comparison with the other countries. The analysis of the results for the students, and in particular boys, who use the Internet a lot shows that this group in Sweden improved their results, but a similar development was not seen in the other countries. If this has had an impact on the Swedish results, the question remains why this has only happened in Sweden despite the fact that students in all of the Nordic countries are highly accustomed to using computers. The Swedish students seem to spend more time on the Internet than in the other countries, but is that enough to create the improvement of the Swedish results? In the analysis of the response patterns, the Swedish results are more similar to the other countries. Looking at Sweden individually, it could be argued that the findings in this study might give some support to the hypothesis that the change of test mode has had a positive impact on the test results. If the analysis is expanded to include all of the Nordic countries, however, the evidence to support this hypothesis is much weaker. Obviously, something has happened to the response pattern that might have improved the results, but we know that only in Sweden did a significant improvement in the overall reading results occur.

Finally, it is also necessary to relate our results to those presented by Jerrim (2018) that showed that Swedish students actually performed better on the PBA than the CBA. Jerrim's result is to some extent in line with Rasmusson´s study (2015) and Mangen, Walgermo & Brønnick's study (2013) comparing students' reading on paper and reading on screen, but is contradicted by some other studies (see for example Baker, 2010; Kim & Huynh, 2010). Even if the study presented in this article cannot prove that the improvements of the Swedish results are related to the change of test mode, it gives some hints that it could have been a contributing factor. According to Jerrim, the improvement of the Swedish students becomes even more of a puzzle. The Swedish students not only improved their results in 2015, but they improved their results in a test mode that was more difficult than the test mode that had been used in 2012. One important issue to keep in mind when Jerrim´s results are discussed in relation to PISA 2015 and compared to the analysis made in this article is that they do not cover the same students and that the students who participated had been selected in different ways. In Jerrim's study, based on the PISA field trial in 2014, a representative sample of students was not selected from Sweden (or from any of the participating countries). This means that the risks for biases related to the sample are bigger in the field trial than in the main study. The purpose of the field study was not to present results on a national level, but to try out the test instruments. As pointed out in Jerrim´s paper, it has not been possible to perform a more detailed analysis of the results of different groups of students as has been done in this article. Having said this, it is still strange that this article comes to partly different results as Jerrim. This adds to the uncertainty of how to regard the impact of the change of test mode in PISA 2015.

5.10 Conclusions

There has been much discussion about the results from PISA 2015. This article has not managed to provide any conclusive evidence that can be used to say that the change of test mode has had an impact on students' results, but hopefully the results shed some light on the issue. Among the Nordic countries, there were no dramatic changes in the results on reading literacy in PISA 2015. Only in Sweden was a significant change in results observed. The findings in this article show that in comparison with the other Nordic countries, Sweden seems to have a larger group of students who spend a lot of time on the Internet, and this group improved its results on PISA 2015 compared with PISA 2012 more so than other groups in Sweden and more so than similar groups in the other Nordic countries. Generally, all of the Nordic countries showed a decreased proportion of students who gave no response to items and who did not reach all of the items in the PISA test on reading literacy. At the same time, the proportion of full-credit responses and no-credit responses increased. The general result of this is most likely that there has been at least some improvement in the scores even if that improvement might have been counterbalanced by other changes.

Even if it is not possible to clearly show whether the change of test mode in PISA 2015 has influenced the results, this article contributes together with other results to raise at least a word of caution. There are more reasons than before to be careful when comparisons are made between PISA results from 2015 and results from earlier PISA studies, and the comparative link between the PISA studies from different years might be weaker in 2015 than before.

What is also important to note when digital testing is discussed is to highlight the advantages with this testing mode. Digitalized testing, such as PISA 2015, moves the test practice closer to the everyday practice of many students. As reading increasingly takes place on screens, it is relevant to also test reading digitally. In addition, it should also be mentioned that the administration of digital tests has several advantages compared with paper and pencil tests. Digital testing will most likely, and rightly, become more and more the dominant testing mode. Students with more experience from using computers will probably gain a lot from CBA when taking tests that demands a great deal of writing, as in the PISA reading literacy test. The only concern that will be important is to consider the comparability to older paper and pencil tests. When countries consider moving from PBA to CBA, there is much that can be gained, but it might at the same time be more problematic to make straightforward comparisons between results from the new CBA with the old results collected through PBA.

5.11 References

Aarseth, E. J. (1997). *Cybertext: Perspectives on ergodic literature*. Baltimore: The Johns Hopkins University Press

Asher, S. R. & Markell, R.A. (1974) Sex differences in comprehension of high and low interest reading material. *Journal of Educational Psychology* 66, 680 – 687

Baker, R. D. (2010). *Comparing the readability of text displays on paper, e-book readers, and small screen devices* University of North Texas, Ann Arbor: ProQuest

Björnsson, J. K. (2008). Changing Icelandic national testing from traditional paper and pencil based tests to computer based assessment: Some background, challenges and problems to overcome. In F. Scheuermann & A. Guimarães Pereira, (Eds.) *Towards a research agenda on computer-based assessment. Challenges and needs for European educational measurement* (pp. 10-14). Ispra, Italy: JRC Scientific and Technical Reports.

Bolter, J. D. & Gromala, D. (2003) *Windows and mirrors: interaction design, digital art and the myth of transparency*. Cambridge, Mass.: MIT

Coiro, J. (2003). Exploring literacy on the internet: reading comprehension on the internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher*, 56(5), 458–464

ETS (2016). *PISA 2015 Main Survey – Notes on IRT Scaling and Analysis.*

Frønes, T. S. (2017) *Å lese og navigere på nettet En studie av elevers navigasjonsstrategier.* Oslo: Universitet i Oslo.

Greiff, S., Wüstenberg, S., Hardt, K. & Avvisati, F. (2015). *Computer-generated log-file analyses as a window into student's minds? A showcase study based on the PISA 2012 assessment of problem solving.* Paris: OECD.

Jerrim, J. (2018) *A digital divide? Randomised evidence on the impact of computer-based assessment in PISA.* CfEE Research Brief January 2018. London: Centre for Education Economics. Webb document http://www.cfee.org.uk/sites/default/files/CfEE%20Digital%20Divide_1.pdf retrieved 30 January 2018

Kamil, M. L., Mosenthal, P. B., Pearson, D., & Barr, R. (2000). *Handbook of reading research, volume III.* London: Routledge.

Kerr, M. A., & Symons, S. E. (2006). Computerized presentation of text: Effects on children's reading of informational material. *Reading and Writing*, 19(1), 1-19.

Kim, D. & Huynh, H. (2010). Equivalence of paper and pencil and online administration modes of the statewide English test for students with and without disabilities. *Educational Assessment*, 15, 107-121.

Kress, G., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London: Routledge.

Mangen, A., Walgermo, B. R., & Brønnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. International Journal of Educational Research, 58, 61-68.

Martin, R. (2008). Utilising the potential of computer delivered surveys in assessing scientific literacy. Sydney: Australian Council for Educational Research

OECD (2010a). *PISA 2009 results: What students know and can do–student performance in reading, mathematics and science (Volume I).* Paris: OECD

OECD (2010b). *PISA 2009 Assessment Framework Key Competencies in Reading, Mathematics and Science.* Paris: OECD

OECD (2011). *PISA 2009 results: Students on line: Digital technologies and performance (Volume VI).* Paris: OECD

OECD (2013). *PISA 2012 results: What students know and can do. Student performance in mathematics, reading and science (Vol. I).* Paris: OECD

OECD (2016a). PISA 2015 *Results (Volume I) Excellence and Equity in Education.* Paris: OECD

OECD (2016b) PISA 2015 *Assessment and Analytical Framework: Science, Reading; Mathematic and Financially Literacy*. Paris: OECD

OECD (2016c). *Governing Board: Interpreting PISA Results: Draft Frequently Asked Questions about PISA 2015.* Paris: OECD

OECD (2017). *PISA 2015 Technical Report*. Paris: OECD.

Rasmusson, M. (2015). Reading paper – reading screen. A comparison of reading literacy in two different modes. *Nordic Studies in Education*, 35(01), 3–19

Rasmusson, M., & Eklund, M. (2013). "It's easier to read on the Internet—you just click on what you want to read..." Abilities and skills needed for reading on the Internet. *Education and Information Technologies*, 18(3), 401-419.

Rasmusson, M & Åberg-Bengtsson, L. (2015). Does performance in digital reading relate to computer game playing? A study of factor structure and gender patterns in 15-year-olds´reading literacy performance. *Scandinavian Journal of Educational Research*, Volume 59, 2015, Issue 6

Rowsell, J. & Burke, A. (2009). Reading by Design: Two Case Studies of Digital Reading Practices*. Journal of Adolescent & Adult Literacy* 5 3(2)

Salmerón, L. & García, V. (2011). Reading skills and children's navigation strategies in hypertext. *Computers in Human Behavior*, 27(3), 1143–1151

Scheuermann, F & Guimarães Pereira, A (2008). *Towards a research agenda on computer-based assessment. Challenges and needs for European educational measurement* Ispra, Italy: JRC Scientific and Technical Reports.

Scott, K.P. (1986) Effects of sex-fair reading material on pupils´ attitudes, comprehension and interest. *American Educational Research Journal*, Spring 23, (19, 105 – 106)

Skolverket (2011) *Eleverna och nätet. PISA 2009 om 15-åringars förmåga att söka, läsa och värdera digital information.* Stockholm: Skolverket

Skolverket (2013) *15-åringars kunskaper i matematik, läsförståelse och naturvetenskap.* Stockholm: Skolverket

Skolverket (2016) *PISA 2015: 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik.* Stockholm: Skolverket

Taube, K. & Munck, I. (1996) Gender differences at the item level. In H. Wagemaker (Ed.) *Are girls better readers? Gender differences in reading literacy in 32 countries. A*msterdam: IEA

Ward, H. (2018) *More doubts cast on reliability of Pisa scores.* London: TES. Web document: https://www.tes.com/news/school-news/breaking-news/more-doubts-cast-reliability-pisa-scores retrieved 30 January 2018

5.12 Appendix A

Table 5: Number of students who answered each item per country in PISA 2012

	Denmark		Finland		Iceland		Norway		Sweden	
Item	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*
About a book Q01	2,261	5,220	2,696	6,133	1,080	2,428	1,433	3,253	1,457	
About a book Q05	2,261	5,220	2,696	6,133	1,080	2,428	1,432	3,254	1,457	3,279
About a book Q06	2,261	5,220	2,696	6,133	1,080	2,428	1,432	3,254	1,457	3,279
Biscuits Q01	2,342	5,139	2,743	6,086	1,081	2,427	1,433	3,253	1,457	3,279
Biscuits Q02	2,342	5,139	2,743	6,086	1,080	2,428	1,433	3,253	1,458	3,278
Biscuits Q06	2,342	5,139	2,743	6,086	1080	2,428	1,433	3,253	1,457	3,279
Children's Futures Q02	2,247	5,234	2,678	6,151	1,060	2,448	1432	3,254	1,418	3,318
Children's Futures Q06	2,247	5,234	2,678	6,151	1,060	2,448	1,432	3,254	1,418	3,318
Children's Futures Q09	2,247	5,234	2,678	6,151	1,060	2,448	1,432	3,254	1,418	3,318
Children's Futures Q10	2,247	5,234	0	8,829	1,060	2,448	1,432	3,254	1,418	3,318
Chocolate and Health Q02	2,325	5,156	2,751	6,078	1,059	2,449	1,410	3,276	1,432	3,304
Chocolate and Health Q03	2,325	5,156	2,751	6,078	1,059	2,449	1,410	3,276	1,432	3,304
Chocolate and Health Q04	2,325	5,156	2,751	6,078	1,061	2,447	1,410	3,276	1,431	3,305
Chocolate and Health Q05	2,325	5,156	2,751	6,078	1,061	2,447	1,410	3,276	1,431	3,305
Fair Trade Q02	2,144	5,337	2,612	6,217	987	2,521	1,331	3,355	1,356	3,380
Fair Trade Q03	2,058	5,423	2,499	6,330	951	2,557	1,315	3,371	1,328	3,408
Fair Trade Q07	2,246	5,235	2,704	6,125	1,061	2,447	1,410	3,276	1,431	3,305
Job Vacancy Q03	2,261	5,220	2,696	6,133	1,080	2,428	1,433	3,253	1,457	3,279
Job Vacancy Q06	2,261	5,220	2,696	6,133	1,080	2,428	1,433	3,253	1,457	3,279
Kokeshi Dolls Q01	2,246	5,235	469	8,360	1,059	2,449	1,410	3,276	1,432	3,304
Kokeshi Dolls Q02	2,245	5,236	2,704	6,125	1,059	2,449	1,410	3,276	1,432	3,304
Kokeshi Dolls Q05	2,245	5,236	2,704	6,125	1,059	2,449	1,410	3,276	1,432	3,304
Narcissus Q01	2,244	5,237	2,677	6,152	1,063	2,445	1,432	3,254	1,418	3,318
Narcissus Q06	2,243	5,238	2,677	6,152	0	3,508	1,432	3,254	1,417	3,319
Narcissus Q07	2,243	5,238	2,677	6,152	1,059	2,449	1,432	3,254	1,418	3,318
Sleep Q03	2,246	5,235	2,704	6,125	1,061	2,447	1,410	3,276	1,431	3,305

	Denmark		Finland		Iceland		Norway		Sweden	
Item	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*
Sleep Q06	2,246	5,235	2,704	6,125	1,061	2,447	1,410	3,276	1,431	3,305
Sleep Q07	2,246	5,235	2,704	6,125	1,061	2,447	1,410	3,276	1,432	3,304
Sleep Q10A	2,246	5,235	2,704	6,125	1,059	2,449	1,410	3,276	1,432	3,304
Sleep Q10B	2,246	5,235	2,704	6,125	1,059	2,449	1,410	3,276	1,432	3,304
South Pole Q01	2,260	5,221	2,696	6,133	1,079	2,429	1,432	3,254	1,457	3,279
South Pole Q02	2,260	5,221	2,696	6,133	1,080	2,428	1,431	3,255	1,455	3,281
South Pole Q04	2,260	5,221	2,696	6,133	1,080	2,428	1,431	3,255	1,455	3,281
Summer Job Q01	2,245	5,236	2,678	6,151	1,064	2,444	1,432	3,254	1,417	3,319
Summer Job Q04	2,245	5,236	2,678	6,151	1,060	2,448	1,432	3,254	1,418	3,318
Summer Job Q05	2,245	5,236	2,678	6,151	1,064	2,444	1,432	3,254	1,418	3,318
Summer Job Q06	2,245	5,236	2,677	6,152	0	3,508	1,432	3,254	1,418	3,318
Work Right – Q03	2,261	5,220	2,696	6,133	1,081	2,427	1,433	3,253	1,456	3,280
Work Right Q02	2,261	5,220	2,696	6,133	1,080	2,428	1,433	3,253	1,457	3,279
Work Right Q06	2,261	5,220	2,696	6,133	1,080	2,428	1,433	3,253	1,457	3,279
World Languages Q01	2,245	5,236	2,677	6,152	1,064	2,444	1,432	3,254	1,418	3,318
World Languages Q05	2,244	5,237	2,677	6,152	1,064	2,444	1,432	3,254	1,418	3,318
World Languages Q06	2,244	5,237	2,677	6,152	1,063	2,445	1,432	3,254	1,418	3,318
World Languages Q08	2,244	5,237	2,677	6,152	1,059	2,449	1,432	3,254	1,418	3,318

Note: * Missing in this context is the students who did not receive the item in their version of the test due to the rotated test design.

5.13 Appendix B

Table 6: Number of students who answered each item per country in PISA 2015

	Denmark		Finland		Iceland		Norway		Sweden	
Item	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*	Valid	Missing*
About a book – Q01	835	6,326	702	5,180	398	2,973	653	4,803	667	4,791
About a book – Q05	835	6,326	703	5,179	398	2,973	653	4,803	668	4,790
About a book – Q06	835	6,326	703	5,179	399	2,972	653	4,803	669	4,789
Biscuits – Q01	1,181	5,980	742	5,140	411	2,960	658	4,798	686	4,772
Biscuits – Q02	1,181	5,980	741	5141	409	2,962	658	4,798	685	4,773
Biscuits – Q06	1,181	5980	741	5,141	409	2,962	658	4,798	685	4,773
Children's Futures – Q02	849	6,312	727	5,155	410	2,961	666	4,790	695	4,763
Children's Futures – Q06	849	6,312	727	5,155	410	2,961	667	4,789	695	4,763
Children's Futures – Q09	849	6,312	727	5,155	410	2,961	667	4,789	695	4,763
Children's Futures – Q10	849	6,312	727	5,155	410	2,961	667	4,789	695	4,763
Chocolate and Health – Q02	1,165	5,996	750	5,132	397	2,974	660	4,796	645	4,813
Chocolate and Health – Q03	1,165	5,996	750	5,132	397	2,974	660	4,796	645	4,813
Chocolate and Health – Q04	1,165	5,996	752	5,130	398	2,973	661	4,795	650	4,808
Chocolate and Health – Q05	1,165	5,996	749	5,133	397	2,974	660	4,796	646	4,812
Fair Trade – Q02	849	6,312	730	5,152	411	2,960	677	4,779	681	4,777
Fair Trade – Q03	848	6,313	730	5,152	411	2,960	676	4,780	681	4,777
Fair Trade – Q07	848	6,313	730	5,152	411	2,960	676	4,780	681	4,777
Job Vacancy – Q03	839	6322	708	5,174	406	2,965	655	4,801	677	4,781
Job Vacancy – Q06	839	6,322	708	5,174	405	2,966	655	4,801	676	4,782
Kokeshi Dolls – Q01	834	6,327	726	5,156	399	2,972	665	4,791	665	4,793
Kokeshi Dolls – Q02	835	6,326	726	5,156	399	2,972	665	4,791	665	4,793
Kokeshi Dolls – Q05	835	6,326	726	5,156	399	2,972	665	4,791	665	4,793
Narcissus – Q01	840	6,321	713	5,169	393	2,978	648	4,808	667	4,791
Narcissus – Q06	838	6,323	709	5,173	386	2,985	647	4,809	657	4,801
Narcissus – Q07	836	6,325	709	5,173	385	2,986	647	4,809	655	4,803

Item	Denmark		Finland		Iceland		Norway		Sweden	
	Valid	Missing*	Valid	Missing*	Valid	Item	Valid	Missing*	Valid	Missing*
Sleep – Q03	847	6,314	730	5,152	410	2,961	676	4,780	679	4,779
Sleep – Q06	847	6,314	730	5,152	410	2,961	676	4,780	679	4,779
Sleep – Q07	847	6,314	730	5,152	410	2,961	675	4,781	679	4,779
Sleep – Q10A	845	6,316	730	5,152	408	2,963	673	4,783	676	4,782
Sleep – Q10B	845	6,316	730	5,152	408	2,963	673	4,783	676	4,782
South Pole – Q01	813	6,348	713	5,169	403	2,968	679	4,777	624	4,834
South Pole – Q02	809	6,352	709	5,173	397	2,974	674	4,782	621	4,837
South Pole – Q04	807	6,354	708	5,174	392	2,979	672	4,784	616	4,842
Summer Job – Q01	850	6,311	727	5,155	410	2,961	667	4,789	690	4,768
Summer Job – Q04	847	6,314	726	5,156	406	2,965	664	4,792	684	4,774
Summer Job – Q05	849	6,312	726	5,156	410	2,961	667	4,789	688	4,770
Summer Job – Q06	847	6,314	726	5,156	406	2,965	664	4,792	685	4,773
Work Right – Q02	837	6,324	708	5,174	407	2,964	655	4,801	680	4,778
Work Right – Q03	839	6,322	708	5,174	409	2,962	656	4,800	681	4,777
Work Right – Q06	838	6,323	708	5,174	408	2,963	655	4,801	679	4,779
World Languages – Q01	848	6,313	724	5,158	402	2,969	660	4,796	682	4,776
World Languages – Q05	846	6,315	724	5,158	400	2,971	657	4,799	680	4,778
World Languages – Q06	843	6,318	719	5,163	399	2,972	652	4,804	675	4,783
World Languages – Q08	841	6,320	718	5,164	398	2,973	651	4,805	674	4,784

Note: * Missing in this context is the students who did not receive the item in their version of the test due to the rotated test design.

5.14 Appendix C

Table 7: Description of items included in both PISA 2012 and 2015

Item Code	Unit Name	Item Format	Situation	Text Format	Text Type	Aspect
R432Q01	About a book	Closed Constructed Response	Personal	Continuous	Argumentation	Integrate and interpret
R432Q05	About a book	Open Constructed Response	Personal	Multiple	Argumentation	Reflect and evaluate
R432Q06	About a book	Complex Multiple Choice	Personal	Continuous	Argumentation	Integrate and interpret
R456Q01	Biscuits	Multiple Choice	Personal	Continuous	Narration	Access and retrieve
R456Q02	Biscuits	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R456Q06	Biscuits	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R420Q02	Children's Futures	Short Response	Educational	Non-continuous	Exposition	Access and retrieve
R420Q06	Children's Futures	Open Constructed Response	Educational	Non-continuous	Exposition	Reflect and evaluate
R420Q09	Children's Futures	Closed Constructed Response	Educational	Non-continuous	Exposition	Access and retrieve
R420Q10	Children's Futures	Open Constructed Response	Educational	Non-continuous	Exposition	Integrate and interpret
R455Q02	Chocolate and Health	Open Constructed Response	Personal	Continuous	Description	Reflect and evaluate
R455Q03	Chocolate and Health	Short Response	Personal	Continuous	Description	Access and retrieve
R455Q04	Chocolate and Health	Multiple Choice	Personal	Continuous	Description	Integrate and interpret
R455Q05	Chocolate and Health	Complex Multiple Choice	Personal	Continuous	Description	Integrate and interpret
R424Q02	Fair Trade	Complex Multiple Choice	Educational	Non-continuous	Argumentation	Integrate and interpret
R424Q03	Fair Trade	Multiple Choice	Educational	Non-continuous	Argumentation	Reflect and evaluate
R424Q07	Fair Trade	Multiple Choice	Educational	Continuous	Argumentation	Reflect and evaluate
R446Q03	Job Vacancy	Closed Constructed Response	Occupational	Non-continuous	Description	Access and retrieve
R446Q06	Job Vacancy	Open Constructed Response	Occupational	Non-continuous	Description	Reflect and evaluate
R406Q01	Kokeshi Dolls	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R406Q02	Kokeshi Dolls	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R406Q05	Kokeshi Dolls	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R437Q01	Narcissus	Multiple Choice	Personal	Continuous	Narration	Integrate and interpret
R437Q06	Narcissus	Multiple Choice	Personal	Continuous	Narration	Integrate and interpret
R437Q07	Narcissus	Open Constructed Response	Personal	Continuous	Narration	Integrate and interpret
R404Q03	Sleep	Multiple Choice	Public	Continuous	Exposition	Integrate and interpret
R404Q06	Sleep	Multiple Choice	Public	Non-continuous	Exposition	Integrate and interpret
R404Q07	Sleep	Complex Multiple Choice	Public	Non-continuous	Exposition	Integrate and interpret

Item Code	Unit Name	Item Format	Situation	Text Format	Text Type	Aspect
R404Q10A	Sleep	Open Constructed Response	Public	Non-continuous	Exposition	Reflect and evaluate
R404Q10B	Sleep	Open Constructed Response	Public	Non-continuous	Exposition	Reflect and evaluate
R220Q01	South Pole	Short Response	Educational	Mixed	Exposition	Access and retrieve
R220Q02B	South Pole	Multiple Choice	Educational	Mixed	Exposition	Integrate and interpret
R220Q04	South Pole	Multiple Choice	Educational	Continuous	Exposition	Integrate and interpret
R453Q01	Summer Job	Multiple Choice	Occupational	Continuous	Instruction	Integrate and interpret
R453Q04	Summer Job	Open Constructed Response	Occupational	Continuous	Instruction	Reflect and evaluate
R453Q05	Summer Job	Complex Multiple Choice	Occupational	Continuous	Instruction	Access and retrieve
R453Q06	Summer Job	Open Constructed Response	Occupational	Continuous	Instruction	Reflect and evaluate
R466Q02	Work Right	Open Constructed Response	Occupational	Continuous	Argumentation	Access and retrieve
R466Q03	Work Right	Complex Multiple Choice	Occupational	Mixed	Argumentation	Integrate and interpret
R466Q06	Work Right	Closed Constructed Response	Occupational	Continuous	Argumentation	Access and retrieve
R412Q01	World Languages	Multiple Choice	Educational	Non-continuous	Exposition	Access and retrieve
R412Q05	World Languages	Multiple Choice	Educational	Continuous	Exposition	Integrate and interpret
R412Q06	World Languages	Complex Multiple Choice	Educational	Continuous	Exposition	Integrate and interpret
R412Q08	World Languages	Open Constructed Response	Educational	Mixed	Exposition	Integrate and interpret

5.15 Appendix D

Table 8: Units and items included in PISA 2012 and 2015, by type of question

Units	Multiple choice	Open question
South Pole R220	Q02, Q04	Q01
Sleep R404	Q03, Q06, Q07	Q10A, Q10B
Kokeshi Dolls R406		Q01, Q05, Q02
World Languages R412	Q01, Q05, Q06	Q08
Children's Futures R420		Q02, Q10*, Q06, Q09
About a book R432	Q06	Q01, Q05
Narcissus R437	Q01, Q06*	Q07
Job Vacancy R446		Q03, Q06
Summer Job R453	Q01, Q05	Q04, Q06*
Chocolate and Health R455	Q04, Q05	Q02, Q03
Work Right R466	Q03	Q02, Q06
Biscuits R456	Q01	Q02, Q06
Fair Trade R424	Q02, Q03, Q07	

Note: * These three items are excluded due to missing results in Finland and Iceland.

Table 9: Averages for girls' reading performance by time spent on the Internet outside of school [IC002801]

		Denmark		Finland		Iceland		Sweden	
Internet outside of school [IC002801]		2012	2015	2012	2015	2012	2015	2012	2015
No time	Average	‡	‡	542.0	‡	‡	‡	‡	‡
	SE	†	†	19.6	†	†	†	†	†
1-30 minutes	Average	511.1	506.5	579.0	534.6	526.3	488.8	512.6	501.7
	SE	17.0	13.8	7.0	9.5	10.0	12.5	9.4	501.7
31-60 minutes	Average	525.6	534.5	572.7	569.1	534.1	522.9	528.3	520.1
	SE	6.1	6.7	4.4	5.3	6.6	8.5	6.3	10.7
Between 1 and 2 hours	Average	527.1	532.3	559.8	561.3	516.9	520.6	523.2	549.9
	SE	3.3	4.9	2.9	3.9	4.0	4.8	3.9	6.4
Between 2 and 4 hours	Average	522.0	523.5	551.1	564.1	505.2	511.1	517.2	534.5
	SE	3.7	3.9	2.9	4.0	4.5	4.3	4.1	4.3
Between 4 hours and 6 hours	Average	500.6	512.1	529.9	545.4	500.1	483.3	504.7	526.7
	SE	5.3	4.5	6.2	5.3	6.2	6.4	6.3	5.4
More than 6 hours	Average	486.3	472.8	536.7	521.4	486.1	455.5	480.2	494.8
	SE	6.5	7.1	10.7	6.5	13.9	7.7	7.2	4.6

Note: † Not applicable. ‡ Reporting standards not met.

Table 10: Averages for boys' reading performance by time spent on the Internet outside of school [IC002801]

Internet outside of school [IC002801]		Denmark 2012	Denmark 2015	Finland 2012	Finland 2015	Iceland 2012	Iceland 2015	Sweden 2012	Sweden 2015
No time	Avg	‡	‡	‡	‡	‡	‡	390.0	‡
	SE	†	†	†	†	†	†	18.9	†
1-30 minutes	Avg	464.7	464.0	501.6	469.8	434.6	415.5	451.6	427.5
	SE	11.6	18.8	10.9	11.2	18.1	19.8	12.3	15.6
31-60 minutes	Avg	476.3	495.1	502.2	507.0	470.7	479.2	460.8	469.6
	SE	7.0	7.5	5.7	6.1	8.9	8.5	8.3	12.5
Between 1 and 2 hours	Avg	494.7	509.4	500.3	517.6	475.3	474.3	472.4	494.8
	SE	5.1	5.5	3.8	4.4	4.8	6.5	6.2	7.5
Between 2 and 4 hours	Avg	495.5	505.2	502.0	528.5	470.5	476.3	486.4	510.4
	SE	4.2	4.2	3.5	3.3	4.6	4.8	5.7	5.3
Between 4 hours and 6 hours	Avg	496.0	500.4	493.9	502.4	461.0	453.3	470.8	496.9
	SE	4.2	4.0	6.1	5.9	6.4	6.0	4.9	5.8
More than 6 hours	Avg	467.7	479.9	492.6	478.1	431.1	448.4	444.0	473.4
	SE	5.2	5.2	9.2	7.0	8.5	6.6	7.3	4.9

Note: † Not applicable. ‡ Reporting standards not met.

6. Feedback for everybody? – Variations in students' perception of feedback

Bent Sortkær, Danish School of Education, Aarhus University and VIA University College, Denmark

6.1 Abstract

All Nordic countries claim to advocate for equal opportunities for all students, irrespective of their gender or their social or ethnic backgrounds. However, the results presented in this study indicate that such equality is not always the case in practice. Using data from PISA (Programme for International Student Assessment) 2015, I have examined teacher-student feedback from a student perspective and found that boys in all Nordic countries perceive significantly more feedback than girls do. In Finland, Norway, and Sweden, immigrant students perceive more feedback than non-immigrant students do. Finally, in all five Nordic countries, high-performing students perceive less feedback than low-performing students do. Implications for research, policy, and practice are discussed.

6.2 Introduction[30]

This study is concerned with how teacher feedback is perceived by individual students in Nordic science classrooms in lower secondary schools. More specifically, the aim was to investigate the differences in the amount of feedback perceived by the students with respect to gender and to different social and ethnic backgrounds.

In Nordic school systems, equal opportunities for all students irrespective of their gender and socioeconomic and ethnic backgrounds are advocated for (Nordisk Ministerråd, 2014). Nevertheless, there are still significant differences with respect to science performance in relation to gender and to social and ethnic background (OECD, 2007, 2016). *Gender* differences in science performance are small in Denmark, Iceland, Norway, and Sweden, whereas girls perform significantly better than boys in Finland (OECD, 2016). However, the results from PISA (Programme for International Student Assessment) 2006 indicate that there are quite significant gender differences in all Nordic countries both in relation to the three different science competencies and in relation to the three different content areas tested in PISA 2006 (Nordisk Ministerråd, 2009). In, for example, the competency to *identify science issues*, girls perform significantly better than boys, whereas boys perform significantly better than girls in the competency to *explain phenomena scientifically* and in the content area of *physical systems*. Also in relation to students' *social background*, there are differences in science performance – here in favor of the students with a higher socioeconomic background (Nordisk Ministerråd, 2009 see chapter 4.2). Finally, there are differences in performance between immigrant and non-immigrant students in all five Nordic countries in favor of the latter (Nordisk Ministerråd, 2009 see chapter 4.1). Although these between-group differences in performance have been a political concern for years, the differences seem to be persistent.

[30] A special thanks to Christian Christrup Kjeldsen for excellent comments on the manuscript.

6.3 Agenda and research question

There is a body of research measuring and describing the magnitude of the differences in performance between different groups of students, but there is very little research examining whether teacher-student interactions and communication inside the classroom might give rise to and/or maintain these between-group differences in performance. In this study, I examine teacher-student communication inside the science classrooms in the Nordic countries by investigating whether there are between-group differences in the amount of teacher-provided *feedback perceived* by the students. The assumption is that between-group differences in the perception of feedback might lead to, maintain, or even increase the described inequality in science performance across the Nordic countries. Based on this introduction, the *research question* is:

> Is there a relationship between students' gender and social and ethnic background and the amount of feedback perceived by the students in science classrooms?

To answer this question, I used data from the OECD (Organisation for Economic Co-operation and Development) PISA (Programme for International Student Assessment) 2015 in order to examine whether student gender or social and ethnic background are related to how much teacher-student feedback the individual student perceives in the context of science lessons in the Nordic countries. Hence, I did not examine the distribution of feedback but rather how much feedback the individual students themselves reported to have perceived in the context of science lessons in the Nordic countries. Differences in the amount of perceived feedback could then either stem from different treatment by the teachers if teachers give more feedback to some students than others *or* it could stem from a difference in how the individual students perceive the feedback *or* it could be a combination of both.

The chapter is organized as follows. First, I present the theory and the empirical literature on feedback and explain what I expected to find in the analysis and why. Then I present the data, explain the methods, and describe the construction of the perceived feedback scale. Finally, I present and discuss the results and conclude by discussing policy implications and implications for research and practice.

6.4 Feedback and student progress

Feedback is an essential part of teacher-student interactions in everyday teaching and has been found to be a key determinant for student learning and achievement (Hattie, 2009; Hattie & Timperley, 2007; Meyer, 2005; Muijs et al., 2014; among others). In their review from 2007, Hattie and Timperley reported an average effect size of feedback of 0.79, which places feedback among the most effective factors in relation to student progress (Hattie & Timperley, 2007). Furthermore, Black and Wiliam (1998) concluded that "the gains in achievement [as a result of feedback and the overlapping concept of formative assessment] appear to be quite considerable, and [...] amongst the largest ever reported for educational interventions" (p. 61). Also in the Nordic countries, the concept of feedback has received a considerable amount of attention in recent years (Christensen, 2015; Gamlem, 2014).

6.5 Feedback as an active process

Because feedback is crucial for learning, it is important that all students in the classroom receive feedback regardless of their background. Nevertheless, there has been little attention paid to between-student differences in how the students perceive such feedback (see Blair, 2009; Gamlem & Smith, 2013; Havnes, Smith, Dysthe, & Ludvigsen, 2012 for exceptions). Feedback is defined as information about the learning process that is made available to an active learner (Black & Wiliam, 1998; Shute, 2008). Nevertheless, the prerequisite of being an active learner as the receiver of the teacher-provided feedback has been neglected or overlooked in most empirical studies into feedback in educational research. It has been assumed that teacher-provided feedback is actually perceived as such by the student, and between-student differences in the perception of feedback have therefore not been considered. In this study, the students' perspectives were used when examining teacher-student communication and, in this way, between-student differences in the perception of teacher-student feedback were examined. This strategy allowed me to examine how the individual student perceives teacher-student feedback in science classrooms in the Nordic countries.

6.6 Variations in students' perception of feedback

Students' gender, social background, and ethnic background influence both how students understand and perceive the social world and how the surroundings interact with the student (Palincsar, 1998). Bourdieu and Passerson (1990) used the term "habitus" to describe how a student's interactions are shaped by his or her unique experiences and socialization, while Bruner (1996) described how the student's culture influences the construction of meaning and how the student's characteristics, such as their social background, affect the student's learning experience. Building on these theories, the main argument in this study is that students' characteristics shape the teacher-student interaction in science lessons in the Nordic countries. I therefore hypothesized that there is a relationship between the student's characteristics and the amount of teacher-student feedback the student perceives. Very few studies have examined between-student differences in the perception of feedback, and no study that I know of has examined how gender, ethnicity, and socioeconomic background are related to between-student differences in the perception of feedback. In this study, I addressed this empirically by looking at how much teacher-student feedback the students perceive in science classrooms and whether this relates to the students' background characteristics.

6.7 Previous research

Overall, there exists very little literature on between-student differences in the perception of feedback, and several papers have called for further research into how the students' characteristics might mediate the relationship between feedback and learning outcomes (Black & Wiliam, 1998; Hattie & Gan, 2011; OECD, 2008; Perrenoud, 1998). The literature I have been able to locate deals with feedback in very broad terms, including different kinds of teacher-student communication and attention. Below, gender differences in the perception of feedback are discussed, and then the literature on the differences based on both the social background and ethnic background of the student is presented.

6.7.1 *Influence of gender*

The literature on gender differences in perceived feedback has yielded inconclusive results. Some studies have found that girls are getting more attention from their teachers than boys (Carvalho, Santos, Conboy, & Martins, 2014; Mike Younger & Warrington, 1996). In Portugal, a study by Carvalho et al. (2014) showed that girls perceive a larger amount of effective feedback than boys do, and through student interviews Younger & Warrington (1996) found that year 10 and year 11 students in England report that girls receive more attention and support in the classroom than boys do. However, other studies have found the opposite, namely that boys are the ones getting the most attention (Francis, 2000; Havnes et al., 2012). One study involving 14–16-year-old students in the United Kingdom found that boys both ask and are being asked more questions than girls are and that they draw more attention from their teachers (Francis, 2000, p. 31). Also, a Norwegian study in upper secondary schools found that boys report perceiving a higher quality of feedback than girls do (Havnes et al., 2012). While all of the above-reported results were based on student response, a study by Younger, Warrington, and Williams (1999) used classroom observation and found that boys are the ones getting the most attention.

As presented above, the literature on gender differences in perceived feedback is ambiguous and does not give a clear indication of what I could expect to find in my analysis. The contradictory findings indeed highlight that more research into this particular field is very much needed.

6.7.2 *Influence of social and ethnic background*

I have not been able to locate any research on how the student's social background or ethnic background relates to the perception of feedback. If feedback is considered as pedagogical communication in broader terms, the theory by Bourdieu (1990) on reproduction in education and the theory on implicit pedagogical communication by Bernstein (1975) are helpful in predicting what differences such an analysis might find. Bourdieu states that the language of the school is the language of the middle-class and, therefore, is inefficient for students not belonging to that class. Bernstein follows the same line of thinking in describing the language of the school as being implicit and invisible and only being available to middle-class students. Drawing on these theories, I expected high-ESCS[31] students to perceive more feedback than low-ESCS students because the former are most likely to have the prerequisites for understanding the implicit pedagogical language of feedback. Likewise, I expected non-immigrant students to perceive more feedback than immigrant students because the latter will tend to have a harder time understanding the school's implicit pedagogical language.

6.8 Data

The data used in this study were a subset of PISA 2015 with the five Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden). In 2015, PISA had a special focus on science. PISA 2006 also had a focus on science; however, what was new in the 2015 edition was five questions in the student questionnaire concerning how much teacher-student feedback the student perceives in the science lessons. In addition to these items, the student questionnaire had student-level information on gender and social and ethnic background. These properties make the PISA 2015 data suitable for answering the research question.

In total, there were 27,328 students from 1,056 schools across the five Nordic countries in the data. The numbers for each country were:

- Denmark – 7,161 students from 333 schools.
- Finland – 5,882 students from 168 schools.
- Iceland – 3,371 students from 124 schools.
- Norway – 5,456 students from 229 schools.
- Sweden – 5,458 students from 202 schools.

The data collection procedure was different in Iceland compared to the other Nordic countries. In Iceland, all students aged 15 years old were included in the assessment, whereas in the other Nordic countries a two-step sampling strategy was used, where schools were randomly selected in the first phase and a number of 15-year-old students within the selected schools were randomly selected in the second phase. Following the

[31] The index of economic, social and cultural status (ESCS) is a composite measure designed by the OECD that I used in the analysis as a measure of social background (see section on 'Measures' below).

guidelines from the PISA 2015 technical manual, I treated the Icelandic data no differently than the data from the rest of the Nordic countries (OECD, 2017, Chapter 8). Students within schools do not necessarily attend the same class or grade, and therefore I examined the between-student differences in the perception of feedback *within-schools* and not *within-classrooms*. However, because there is very little sorting in the Nordic countries in relation to gender,[32] and because most schools have a relatively heterogeneous student body and uptake in terms of ethnicity[33] and socioeconomic[34] background, I argue that the data make it possible to answer the research question.

6.9 Methods

6.9.1 Feedback as a latent construct

The central measure in this study is the "*perceived feedback*" scale, which is a latent construct and cannot be measured directly. Therefore, I used the five feedback-related survey questions from the student questionnaire to capture different aspects of teaching that are theoretically related to the concept of perceived feedback and to construct a scale representing the amount of feedback perceived by the students in the context of science lessons in the Nordic countries. If a student did not answer at least one of the five questions, the student was assigned a missing value for the perceived feedback variable. The questions are about how often certain things happen in the science lessons and were answered on a four-point Likert scale (*Never or almost never; Some lessons; Many lessons; Every lesson or almost every lesson)*. The five questions are:

---

[32] Only very few students attended all-boy or all-girl schools in the data (38 boys and 27 girls).

[33] Less than 2% of the students in the data were enrolled in schools where less than 50%of the students in the sample spoke the language of the test at home or were born in the test country.

[34] In spite of an increased segregation in the Nordic countries in recent years (see Holmlund, 2015), the ESCS intraclass correlations (ICCs) in the Nordic countries are the lowest among all OECD countries participating in PISA 2015 (.16 in Denmark, .14 in Finland, .11 in Iceland, .13 in Norway, and .10 in Sweden). The average ESCS ICC for the rest of the OECD countries is .26 (lowest in Canada (.18) and the UK (.18) and highest in Chile (.61)). A low ICC indicates that most of the between-student variation in ESCS is within-school variation (as opposed to between-school variation). Therefore, a low ICC, as we see in the Nordic countries, indicates a heterogeneous student body when it comes to the students' ESCS.

1. The teacher tells me how I am performing in this course.
2. The teacher gives me feedback on my strengths in this <school science> subject.
3. The teacher tells me in which areas I can still improve.
4. The teacher tells me how I can improve my performance.
5. The teacher advises me on how to reach my learning goals.

Before using the constructed scale on "perceived feedback" in my analysis, I tested the properties of the constructed scale empirically using confirmatory factor analysis with the statistical package Stata 14.2. I ran the test separately for each country because the analyses were at the country level. The results of these tests are listed in Table A1 in the Appendix.

First, I tested whether all five questions captured the same construct. The scale had a reliability coefficient (Cronbach's alpha) in all five countries of around 0.93, which was very high and indicated that the five items were highly correlated and were measuring the same construct. *Second*, I tested the convergent validity of the scale to see whether all five questions were contributing with an acceptable level of variance to the construct. I used the rule of thumb that this number should be above .5 (Mehmetoglu & Jakobsen, 2017). The scale showed acceptable convergent validity in all five countries, with an average variance extracted of around 0.74, indicating that all five standardized factor loadings, on average, were above 0.7. *Third*, I tested whether the empirical data fit the proposed theoretical model with five questions. The model fit indices of the proposed model revealed some minor problems with the Chi-square test and the Root Mean Square Error of Approximation (RMSEA). The Chi-square test is known to be affected by large sample sizes such as those in PISA, so this explains the large Chi-square (Kline, 2013). The RMSEA should be below .1, but it was around .16 in all countries, and this indicated that the proposed model did not fit the data perfectly. To improve the model fit, I modified the models using an exploratory approach (modification indices in Stata 14.2) by allowing the errors in the models to correlate. After this modification, the RMSEA was acceptable in each country. Overall, in spite of the large Chi-square, I considered the scale to have acceptable properties and I proceeded using all five questions to create the perceived feedback scale. After testing and modifying the scale, I standardized it within countries with a mean of zero and a variation of one.

A consequence of the country-specific modifications of the models was that not all Nordic countries were using the exact same empirical model. Denmark and Norway used one model, Finland and Iceland used a second model, and Sweden used a third model (see Table A1). This had no implication for answering the research question as to the relation between students' characteristics and perceived feedback *within* countries. However, cross-country comparisons of the results must be interpreted with this in mind.

6.9.2 Group variables

To answer the research question, I included a variable on gender, namely the dummy variable *girl*. I also included two dummy variables on ethnicity – *born in the test country*, indicating whether the respondent was born in the test country, and *speaks the test language at home*, indicating whether the respondent speaks the language of the test at home most of the time. To measure the social background of the student, I used the index of economic, social, and cultural status (*ESCS*), which is a composite variable generated by the PISA consortium on the basis of the following variables: the highest occupational status of the parents (according to the International Socio-Economic Index of occupational status), the highest level of education completed by the parents, and a list of possessions in the home such as a car or television (OECD, 2017, Chapter 16). The students were divided into four equal-sized groups based on their relative positions on the ESCS scale within each country. Group 1 was the 25% of the students with the lowest ESCS score, group 2 was the next 25% of the students, and so forth. In the analyses, group 4 was the group of students with the highest ESCS score and was the reference group.

In addition to the above-described variables, I controlled for student performance. I used the average of the ten plausible values for the student score from the PISA science test and grouped the students into quartiles in accordance with their position in the score distribution.[35] Group 1, being the 25% lowest-performing students, was the reference group in the analysis. I then controlled for school-average ESCS (measured as the average ESCS of the students participating in PISA 2015 from each school) and standardized this measure within the country to have a mean of zero and a variation of one. An overview of the means, standard deviations (*SD*), and ranges is shown in Table 1.

---

[35] As a robustness check, I grouped the students using every one of the ten plausible values at a time, and the results were almost identical with no substantive differences.

Table 1: Descriptive statistics

Variable	All Nordic countries		Denmark		Finland		Iceland		Norway		Sweden	
Dependent variable	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*
Perceived feedback (*mean* = 0, *SD* = 1 in all countries)	−1.22	2.78	−1.25	3.09	−1.20	3.08	−.97	2.93	−1.36	2.57	−1.30	2.45
Independent variables	*mean*	*SD*	*mean*	*SD*	*mean*	*SD*	*mean*	*SD*	*mean*	*SD*	*mean*	*SD*
Girl (0/1)	.51	.50	.52	.50	.50	.50	.53	.50	.51	.50	.52	.50
ESCS (0–4)	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12
Born in the test country (0/1)	.93	.26	.93	.26	.97	.18	.91	.28	.92	.27	.91	.29
Speaks test language at home (0/1)	.91	.29	.88	.33	.95	.22	.95	.22	.92	.28	.86	.34
Performance (1–4)	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12	2.50	1.12
Girl (0/1)	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*	*min*	*max*
School average ESCS (*mean* = 0, *SD* = 1 in all countries)	−6.01	4.01	−3.12	2.35	−2.16	3.82	−6.01	1.92	−3.18	4.02	−4.86	2.75
Number of students		27,328		7,161		5,882		3,371		5,456		5,458
Number of schools		1,056		3330		168		124		229		202

6.9.3 Missing values

Compared to most other surveys, PISA 2015 had a very high response rate (around 90% in the Nordic countries), which is an indicator of high-quality data. However, not all respondents (students) answered all questions, and here I will describe how missing data were handled.

A total of 3,130 students did not answer at least one of the five perceived feedback questions, and a closer look shows an overrepresentation among these students of boys, students who do not speak the language of the test at home, students not born in the country of the test, and low-ESCS students. Because answers were not missing completely at random, a simple deletion of these students might create biased estimates. Therefore, I used multiple imputations to keep all observations in the analysis (five imputations using Stata 14.2) (Enders & Peugh, 2004).

There were 649 missing values for the ESCS variable, and these were replaced by the school-average ESCS. Born in the test country had 680 missing values and speaks the test language at home had 522 missing values. For these two variables on ethnicity, I added a third category to the original dummy variables and thus kept the observations in the analysis. I do not report the estimates of this third category.

6.9.4 Multi-level regression model

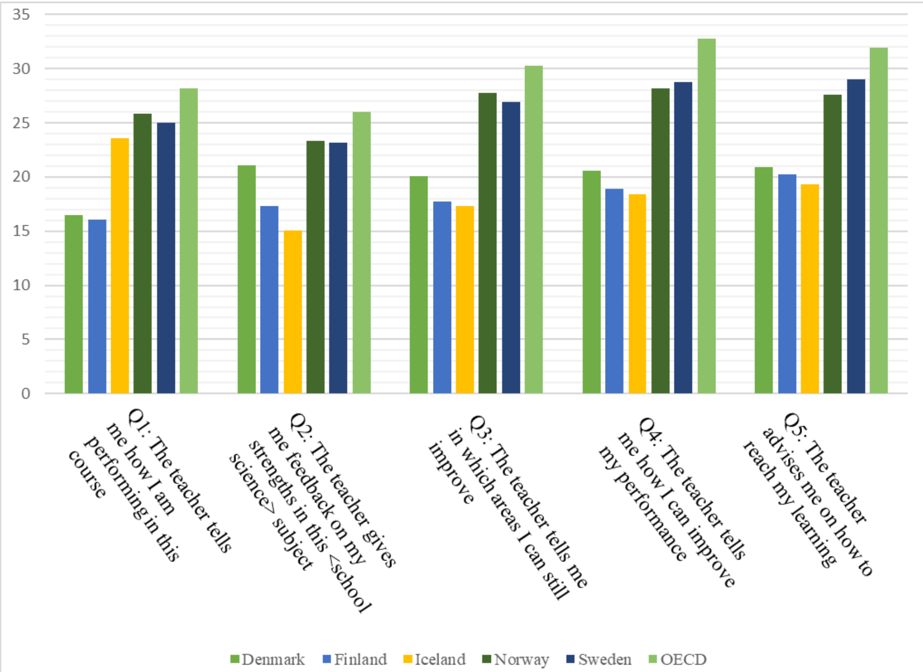
To answer the research question, I used the "perceived feedback" scale as the dependent variable in a multi-level regression model to examine variations in perceived feedback between groups (gender and social and ethnic background). By using a multi-level regression, the model took into account the unobserved shared characteristics of students attending the same school, and the standard error was adjusted accordingly. Furthermore, student and school weights were included in the model to take account of the sampling procedure.

The coefficient estimates of the model could thus be interpreted as the relative difference in the amount of perceived feedback between, for instance, boys and girls. In the multi-level model, I treated the student characteristics as level one and the school as level two. I calculated the model separately for each Nordic country.

6.10 Results

I start this section by presenting the raw numbers on how the students in the Nordic countries responded to the five feedback questions in the student questionnaire. Figure 1 shows the percentage of students in each of the Nordic countries that reported perceiving feedback in many or all lessons. I present this figure to give an impression of the amount of feedback 15-year-old students perceive in each of the five Nordic countries. I also include the average for the rest of the OECD countries in the figure for comparison.

Figure 1: Percentage of students who answer "many lessons" or "every or almost every lesson" to the five feedback questions



Note: The Nordic countries are not included in the OECD average.

Across all five Nordic countries, fewer than a third of the students answered "many lessons" or "every or almost every lesson" to any of the five feedback questions. Whether feedback is effective is not just a question of the amount of feedback, but elements such as timing, type, and function are also relevant (Sortkær, 2017). Thus, it is difficult to judge whether there should be more perceived feedback in the Nordic classrooms. Nevertheless, having more than two-thirds of the students reporting that they only perceive feedback in some lessons or not at all shows that there is room for improvement in relation to these students.

There was considerable variation across the five Nordic countries. Students in Norway and Sweden seemed to perceive feedback more often than students in Denmark, Finland, and Iceland. Figure 1 also shows that students in the five Nordic countries perceived less feedback than students in the rest of the OECD countries. In Iceland, students reported perceiving relatively more of the type of feedback related to Q1 (*The teacher tells me how I am performing in the course*) than of the other types of feedback. Q1 can be understood as the summative aspect of feedback. Denmark and Finland, on the other hand, were relatively low on this aspect of feedback (Q1). Students in Iceland, Norway, and Sweden reported perceiving relatively less of the type of feedback related to Q2 (*The teacher gives me feedback on my strengths in this <school science> subject*) compared to the other types of feedback.

Differences in student response between the Nordic countries as well as differences in relation to the rest of the OECD countries must be read with some caution due to a potential *cultural response bias* (Kjærnsli & Lie, 2011) and/or due to *different standards* (Ning, Van Damme, Van Den Noortgate, Yang, & Gielen, 2015). Kjærnsli and Lie (2011) described how cultural factors can influence the way questions are answered, and Ning et al. (2015) described how students in different countries have different standards for evaluating teaching.

The results from the multilevel regression model are shown in Table 2. In all of the Nordic countries, boys perceived significantly more feedback than girls. In Denmark, for instance, boys scored a third of a standard deviation higher on the perceived feedback scale than girls, and in Iceland the difference was .46 standard deviations in favor of the boys. By using a regression model, I was able to "control" for other characteristics. In other words, even if I compared boys and girls with similar social and ethnic backgrounds, with similar performance levels in science, and who came from schools with similar average ESCS, the boys still reported perceiving significantly more feedback than the girls did. The gender difference in perceived feedback is not an isolated Nordic phenomenon, and the boys in the other OECD countries, on average, also reported perceiving more feedback than the girls did (own calculations).

Table 2: Multilevel regression models on perceived feedback in science

	Denmark	Finland	Iceland	Norway	Sweden
Student level					
ESCS					
Level 1	−.04	−.03	−.05	−.03	.01
Level 2	−.03	−.05	−.09	−.02	.00
Level 3	−.03	−.02	−.09	.00	.00
Level 4 – reference					
Girl	−.35***	−.37***	−.46***	−.28***	−.37***
Speaks test language at home	−.07	−.13†	−.03	−.09	−.11*
Born in the test country	−.06	−.19*	−.01	−.17*	−.16*
Science performance					
Level 1 – reference					
Level 2	−.16**	−.16**	−.13*	−.06	−.08†
Level 3	−.27***	−.27***	−.29***	−.12*	−.26***
Level 4	−.34***	−.36***	−.45***	−.21***	−.24***
School level					
School-average ESCS	−.04*	.05*	−.09**	−.09***	−.05*
Number of students	7,161	5,882	3,371	5,456	5,458
Number of schools	333	168	124	229	202

Note: Models estimated by maximum likelihood.
† $p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$ (two-tailed tests).

The coefficients for the ESCS variable can be interpreted as relative to the 25% of the students with the highest ESCS score, namely, those students in level 4. None of the coefficients are statistically significant and provide no support for my hypothesis that high-ESCS students perceive more feedback than low-ESCS students do.

Looking at the coefficients for the two variables on ethnicity, there are some significant relationships. In Finland and Norway, students not born in the test country scored significantly higher on the perceived feedback scale, and in Sweden both students not speaking the language of the test and students not born in the test country scored significantly higher on the perceived feedback scale. In Denmark and Iceland, the coefficients are not significant. Contrary to my hypothesis, it seems that immigrant students perceive more feedback than non-immigrant students do in Finland, Norway, and Sweden.

In all five Nordic countries, students scored lower on the perceived feedback scale in proportion to their performance in the PISA science test, and high-performing students reported perceiving significantly less feedback than low-performing students did. The relationship showed an almost linear trend in all five countries, and this finding might suggest that teachers in all five Nordic countries give the most help, in the form of feedback, to the students having the hardest time grasping the science curriculum. The relationship was strongest in Iceland, followed by Denmark and Finland.

As for the school-average ESCS of the students, in Denmark, Iceland, Norway, and Sweden students scored significantly lower on the perceived feedback scale if they were enrolled in high-ESCS schools. The opposite was the case in Finland, where students enrolled in high-ESCS schools perceived significantly more feedback than students from low-ESCS schools. Although statistically significant, the coefficient estimates for the relationship between school ESCS and perceived feedback were rather small and should not be given too much attention.

6.11 Measurement invariance

The above analysis revealed significant differences in the amount of perceived feedback between boys and girls in all five countries. The underlying assumption in the analysis was that the perceived feedback scale was measuring the same underlying latent construct within each group, in other words, that the scale has the same meaning for all groups of students (Kline, 2013). In this section, I present the tests for measurement invariance across gender and discuss the findings. I will not present the full details of the tests, just the results.

The perceived feedback scale showed configurational (or dimensional) invariance across gender, meaning that the number of factors in the model was the same for boys and girls. The next level of invariance I tested for was metric invariance, meaning that the factor loadings were equal across gender. The scale showed metric invariance in Denmark, Finland, and Sweden, but not in Iceland or Norway. The next level of measurement invariance was strong invariance. On this level, in addition to having similar factor loadings across groups, I tested for similar intercepts across gender. All scales failed to show strong invariance, meaning that factor loadings and the intercepts were not similar across gender.

Table 3: Measurement invariance on gender

	Denmark	Finland	Iceland	Norway	Sweden
Configurational invariance	+	+	+	+	+
Metric invariance	+	+	–	–	+
Strong invariance	–	–	–	–	–

Note: + measurement invariance found, – measurement invariance not found.

These tests of measurement invariance revealed that the perceived feedback scale was only to some extent measuring the same construct for boys and girls because the scale failed to meet the strictest tests. Especially in Iceland and Norway, the perceived feedback scale seemed to measure slightly different constructs, and this might partly explain the difference found between boys and girls in perceived feedback in these two countries.

6.12 Conclusion, discussion, and policy implications

All Nordic countries claim to advocate for equal opportunities for all students, irrespective of their gender or their social or ethnic backgrounds. However, the results presented in this study indicate that such equality is not always the case in practice.

I found that boys reported perceiving much more feedback than girls in all five Nordic countries. Using a regression framework, I inferred that this difference was present even when looking at students who appeared the same in all other observables such as social background, ethnic background, and science performance. The difference was statistically significant in all countries, and the size of the coefficient was non-trivial. Unfortunately, it is not possible to say anything about *why* boys seem to perceive significantly more feedback than girls do in Nordic science classrooms. It is not possible to deduce from the data whether the gender difference stems from different treatment by the science teachers or from a difference in perception by the students of the feedback given or a combination of both. One part of the explanation might be that boys attract more attention from the teachers and thus get more feedback, as was suggested by Francis (2000). Another part of the explanation might be that similar information from the teachers is understood as feedback by some students but not by others. Finally, this difference might be due to different interpretations and understandings of the feedback questions in the feedback scale used in this study. The test of the perceived feedback scale failed to show strong measurement invariance, and this indicated that part of the difference between genders might be because the scale measured slightly different constructs when applied to boys and girls. To get more precise estimates about gender differences in the perception of feedback, future large-scale assessments should improve the feedback construct so that it works equally well across genders. Furthermore, it would be interesting and helpful to combine student questionnaires like the ones used in PISA 2015 with objective observations and interviews among the same students in order to learn more about the mechanisms that contribute to the gender difference in perceived feedback.

The analysis furthermore shows that, in Finland, Norway, and Sweden, students not born in the test country perceived more feedback than students born in the test country, and in Sweden, students not speaking the language of the test at home were the ones perceiving the largest amount of feedback. One explanation for this relationship might be that teachers are conscious of the potential language difficulties and thus are more explicit in their communication when they give feedback to immigrant students. In an observational study in a Danish kindergarten, Palludan (2004) found that the pedagogues more often used an educational language when communicating with immigrant children than they did in communication with non-immigrant children. Even though that study was conducted in a different context, the same mechanism might be at play in lower secondary classrooms in Finland, Norway, and Sweden.

The analysis did not reveal any difference in perceived feedback in relation to the social background of the students.

Finally, the analysis shows an almost linear relationship between science performance in the PISA test and the amount of feedback perceived. Low-performing students reported perceiving significantly more feedback than high-performing students. There is some intuitive logic to this result. In an everyday reality with 28 students in a science classroom, there is very little time for a teacher to give feedback to all students. A consequence of this might be that the students who are in most need of attention are the ones getting the feedback, and those students might very well be the ones who are struggling to understand the science curriculum. While this result is understandable from the scenario described, this might not be an ideal situation because the high-performing students will be left without much feedback and therefore might not reach their full potential.

The theory of Lev Vygotsky (1978) emphasizes that children's cognitive development is advanced through social interaction with more skilled individuals. Therefore, if high-performing students do not get much feedback from their teachers, their cognitive development might not be optimized. The purpose of a recent report from the Norwegian Ministry of Education and Research (2016) was to come up with suggestions on how to increase the group of students who perform at advanced levels. A suggestion from that study was that a redistribution of feedback might have the potential to stimulate the cognitive development of high-performing students.

This being said, high and low-performing students might not need the same amount of feedback to advance their cognitive development. Therefore, the relationship between performance and perceived feedback calls for an analysis of whether the distribution of feedback within the classrooms is *optimal* for both low and high-performing students. We need to know more about the type and quality of the feedback that the students report they perceive.

The relationship between student performance and the amount of perceived feedback is not an isolated Nordic phenomenon, and the trend in the rest of the OECD countries looks the same with comparable estimates.

Overall, the knowledge gained from this study has the potential to inform future research and practice by providing an insight into between-student differences in perceived feedback. Taking the perspective of the students will enable teachers to adjust and fine-tune the feedback information given in the context of science lessons in order to reach out to *all students* in the classroom and, therefore, to realize the full potential of feedback for learning outcomes (Hattie & Gan, 2011). Thus, the presented findings are relevant for teachers in practice, teachers to be, educators in teachers' colleges, and policymakers in general in order to focus on the relevance of the student perspective in feedback practices.

Further studies should look into differences in the quality of the perceived feedback. As a study from the UK suggests, boys report receiving more negative attention, because teachers expect them to misbehave, and girls report receiving more positive attention because they appear more attentive and ready to learn (Michael Younger et al., 1999). The same type of qualitative differences between genders might also be at play in the Nordic countries and in relation to ethnicity and performance level.

6.13 References

Bernstein, B. (1975). Class and pedagogies: Visible and invisible. *Educational Studies*, *1*(1), 23–41.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Blair, K. P. (2009). *The neglected importance of feedback perception in learning: An analysis of children and adults' uptake of quantitative feedback in a mathematics simulation environment*. Stanford University.

Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in education, society and culture*. London: Sage Publications.

Bruner, J. S. (1996). *The culture of education*. Cambridge: Harvard University Press.

Carvalho, C., Santos, J., Conboy, J., & Martins, D. (2014). Teachers' feedback: Exploring differences in students' perceptions. *Procedia – Social and Behavioral Sciences*, *159*, 169–173.

Christensen, V. (2015). *Nettekster fanger og fænger – Multimodale tekster, feedback og tekstkompetence i danskundervisningen i udskolingen*. Aalborg Universitet.

Enders, C. K., & Peugh, J. L. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, *74*(4), 525–556.

Francis, B. (2000). *Boys, girls and achievement: Addressing the classroom issues*. London: Routledge Falmer.

Gamlem, S. M. (2014). *Tilbakemelding som støtte for læring på ungdomssteget*. Universitetet i Stavanger.

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, *20*(2), 150–169.

Hattie, J. A. (2009). *Visible learning – A Synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie, J. A., & Gan, M. (2011). Instruction based on feedback. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 249–271). New York: Routledge.

Hattie, J. A., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*(1), 21–27.

Holmlund, H. (2015). Education and equality of opportunity: What have we learned from educational reforms. In *Nordic economic policy review: Whither the Nordic welfare model?* (pp. 133–170). Copenhagen: Nordic Council of Ministers.

Kjærnsli, M., & Lie, S. (2011). Students' preference for science careers: International comparisons based on PISA 2006. *International Journal of Science Education*, *33*(1), 121–144.

Kline, R. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171–207). New York: Routledge.

Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using stata*. London: Sage Publications.

Meyer, H. (2005). *Hvad er god undervisning?* København: Gyldendal.

Ministry of Education and Research. (2016). *Mer å hente – Bedre læring for elever med stort læringspotensial*. Oslo: 07 Xpress AS.

Muijs, D., Kyriakides, L., van der Werf, G., Creemers, B., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, *25*(2), 231–256.

Ning, B., Van Damme, J., Van Den Noortgate, W., Yang, X., & Gielen, S. (2015). The influence of classroom disciplinary climate of schools on reading achievement: A cross-country comparative study. *School Effectiveness and School Improvement*, *26*(4), 586–611.

Nordisk Ministerråd. (2009). *Northern lights on PISA 2006. Differences and similarities in the Nordic countries*. (T. Matti, Ed.). Copenhagen.

Nordisk Ministerråd. (2014). *Growth, welfare and values: Programme for the Danish presidency of the Nordic Council of Ministers 2015*. Copenhagen.

OECD. (2007). *PISA 2006 – Science competencies for tomorrow's world: Volume 1 – analysis*. Paris: OECD.

OECD. (2008). *Assessment for learning formative assessment*. Paris: OECD.

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: OECD.

OECD. (2017). *PISA 2015 technical report*. Paris: OECD.

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, *49*, 345–375.

Palludan, C. (2004). *Børnehaven gør en forskel: Et pædagogisk-antropologisk hverdagslivsstudie af differentieringsprocesser*. Aarhus University.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 85–102.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Sortkær, B. (2017). Feedback i matematik – et litteraturstudie. In R. G. Henriksen (Ed.), *Feedback i matematik* (pp. 9–20). Frederikshavn: Dafolo.

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Younger, M., & Warrington, M. (1996). Differential achievement of girls and boys at GCSE : Some observations from the perspective of one school. *British Journal of Sociology of Education*, *17*(3), 299–313.

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: Reality and rhetoric? *British Journal of Sociology of Education*, *20*(3), 325–341.

6.14 Appendix

Table A1: Scale test

	Denmar k	Finland	Iceland	Norway	Sweden
Covariance between:	Q1*Q2	Q1*Q2 Q3*Q5	Q1*Q2 Q3*Q5	Q1*Q2	Q1*Q2 Q4*Q5
Cronbach's Alpha (should be above .7)	.91	.93	.94	.94	.94
Standardized factor loadings					
Q1 The teacher tells me how I am performing in this course	.69	.71	.72	.76	.79
Q2 The teacher gives me feedback on my strengths in this <school science> subject	.74	.80	.84	.83	.86
Q3 The teacher tells me in which areas I can still improve	.88	.91	.95	.90	.93
Q4 The teacher tells me how I can improve my performance	.91	.90	.93	.93	.89
Q5 The teacher advises me on how to reach my learning goals	.84	.91	.93	.88	.87
Average Variance Extracted (should be above .5)	.67	.73	.77	.75	.76
Model fit indices					
Chi-squared	195.7	29.4	38.8	156.4	89.4
RMSEA (should be below .1)	.09	.04	.06	.09	.08
CFI (should be above .9)	.99	.99	.99	.99	.99
TLI (should be above .9)	.98	.99	.99	.98	.99
SRMR (should be below .1)	.01	.01	.01	.01	.01

Note: a) Root Mean Square of Approximation.
b) Comparative Fit Inde.
c) Tucker-Lewis Index.
d) Standardized Root Mean square Residual.

7. The urban advantage in education? Science achievement differences between metropolitan and other areas in Finland and Iceland in PISA 2015

Kari Nissinen[36]
Ragnar F. Ólafsson[37]
Juhani Rautopuro[36]
Almar M. Halldórsson[37]
Jouni Vettenranta[36]

7.1 Abstract

In Finnish and Icelandic PISA 2015 data, the average student performance was consistently higher in the metropolitan areas (i.e. Helsinki and Reykjavik, respectively) than in the other parts of the countries. This paper investigates variables that might explain the observed regional differences. By statistical modeling, it was found that in both countries the regional differences in performance were reduced to regional differences in a few background variables. Part of the regional differences was explained by the higher average socio-economic status and cultural capital of metropolitan homes, but it was also found that students' ambitions (Iceland) and

[36] Finnish Institute for Educational Research, University of Jyväskylä, Finland
[37] Directorate of Education, Iceland
Note: Findings and conclusions are the authors' and do not necessarily reflect the views of their respective institutions.

occupational expectations (Finland) played important roles, both being at higher levels among metropolitan students. This suggests that policymakers should pay specific attention to student counseling and career planning in regions that do not necessarily offer versatile employment opportunities. This might raise students' learning motivation and, consequently, their learning results.

7.2 Introduction

The aim of this paper was to identify variables that explain regional differences in PISA achievement in science in 2015 within Finland and Iceland, respectively. We compared the capital/metropolitan areas of those countries with other parts of the countries and attempted to identify variables that could account for observed differences in achievement. In Finland, we compared the capital area with four other regions. In Iceland, the capital area was compared with smaller towns and rural areas.

International research has accumulated evidence that living in capital cities or large urban areas presents various advantages for the inhabitants in terms of access to services, cultural events, health care, and educational supply. In Finland, the PISA test results have usually been fairly similar in all parts of the country. Therefore, the research dealing with urban areas has mainly concentrated on school choices and segregation (e.g. Armila, Käyhkö and Pöysä 2018; Bernelius and Vaattovaara 2016). Similar kinds of studies have also been recently conducted in Sweden (Bäck 2016). In Finland, some research has also been undertaken regarding how well the students' grades correspond with their learning results in different areas of Finland (Harju-Luukkainen, Vettenranta, Ouakrim-Soivio and Bernelius 2016).

The PISA 2015 was an exception because the Helsinki metropolitan area stood out from the rest of the country in all PISA assessment domains, i.e. scientific literacy, reading literacy, and mathematical literacy (Vettenranta, Välijärvi, Ahonen, Hautamäki, Hiltunen, Leino, Lähteinen, Nissinen, Nissinen, Puhakka, Rautopuro and Vainikainen, 2016; Bernelius and Kauppinen, 2011). In Iceland, PISA achievement has been persistently lower in areas outside the capital (Halldórsson and Ólafsson, 2016).

The analysis presented in this paper is not theory-driven. As a step towards understanding the nature of these differences, however, we analyzed PISA 2015 achievement in science and examined whether these differences could be statistically explained by variables assessed in the student questionnaire that was administered after the students had completed the PISA test and from the school questionnaires that were filled in by the principals in each school.

While PISA provides the opportunity for international comparisons, it is useful to examine assessments conducted at the national level in order to further establish whether the capital metropolitan area distinguishes itself from other parts of the country. In short, the national tests administered annually in Iceland to all pupils at the end of compulsory school show repeatedly in recent years an overall difference in favor of the capital area in comparison with other parts of the country, albeit with some variability within regions at different times. Finland does not have annual national standardized tests. Instead the tests are sample-based and have a certain cycle in different school subjects and occur at the end of basic education. Regional differences from the point of view of provinces or types of municipalities have been examined. The Finnish results vary significantly between different school subjects. Mathematics and mother tongue are assessed quite regularly. In a recent mathematics assessment (9th graders, 2015), the only statistically significant differences were detected between pupils in Southwest Finland and Eastern Finland. Differences between municipalities of different kinds were not detected. In contrast, in the mother tongue assessment pupils in Eastern Finland had the best results together with pupils from Southern Finland. The lowest achievers were from Lapland – especially the boys. Moreover, pupils from urban municipalities had better results than pupils from suburban and rural municipalities.

7.3 Predicting science achievement in PISA

The variables that were employed in this analysis to explain PISA achievement in science in rural and urban areas in Finland and Iceland can be roughly grouped into five categories. Overall, these variables have been shown in previous studies to be related to academic achievement and include assessments of *socio-economic status*, *cultural wealth* (e.g. number of books at home and other cultural possessions), the *situation at home* (e.g. home educational resources and emotional support provided by the parents), the parents' occupational status, and the parents' educational level, which one would expect to influence the learning environment of the children.

Bringing the focus to the students themselves, *motivational factors* were assessed, such as the students' self-reported expected educational level, their expected occupational status, and their achievement motivation. In addition, students' internal motivation (do they enjoy science) and external motivation (do they see learning science as useful) as well as the time spent learning were considered.

Another major group of variables centered around the *disciplinary climate* in the schools, e.g. how often students come late for school. The *educational level of the teachers* was assessed, i.e. the percentage of certified teachers in the school and the number of teachers certified specifically in science. Also, *educational resources in the school* in the form of the availability or shortage of staff, educational materials, and resources for science learning were assessed.

Research on rural vs. urban areas, or metropolitan vs. regional, will be discussed along with research on motivational factors, disciplinary climate, student-teacher ratio, etc.

7.4 Capital and urban versus rural or regional

When a difference between urban and rural schools is observed in international studies, it is generally in favor of the urban schools. Curtis et al. (2017) found that attending local urban schools is associated with a greater likelihood of graduation in New Zealand. Young (2006) examined the differences in student achievement between rural and urban schools in Western Australia, and after controlling for student background variables their study showed that students attending rural schools were not performing as well as students from urban schools. Mohammadpour and Ghafar (2014) pointed out that cross-nationally in TIMSS 2007 “a large and significant inequality was found between students from urban and rural schools in mathematics achievement” (p. 210).

In contrast, Howley and Gunn (2003) reported that there is no difference in the results of the mathematics test in the United States between rural area students and students living in cities. In Finland and Iceland, there usually are only small differences between schools, and there are usually small differences between regions or different kinds of municipalities in Finland.

As suggested by Birzea et al. (2006, cited by Smit et al. 2015), the existence of an urban-rural difference in education might be less pronounced in economically advanced countries because the problem of attracting qualified teachers might be greater in poorer countries. Spending on education in poorer rural areas might be perceived as a luxury (see Mussa, 2013).

If the difference between metropolitan and regional or rural achievement is in some way related to a shortage of qualified teachers, this difference should be less pronounced in the developed countries. While there are notable exceptions, this hypothesis did get some support in a study of a number of countries (Mohammadpour and Ghafar, 2014, p. 210). To address the gap in student achievement, Piyaman et al. (2017) suggested that action is required aimed at building the capacity of the principals and teachers who work with rural pupils.

Harmon, Henderson, and Royster (2003) argue that many science teachers are teaching in a secondary teaching role or "out-of-field" altogether (p. 55) both in terms of content and pedagogical qualifications. In Finland, however, teacher qualification requires a master's degree and pedagogical studies. In Finland approximately 95% of teachers in comprehensive schools are qualified (Finnish National Agency for Education, 2017).

A study by Young (2000, cited by Abrams and Middleton, 2017) indicated that rural students tended to have "weaker beliefs in their own academic ability and did not pursue additional educational opportunities compared to their suburban and urban peers" (p. 167). This has links to another set of variables addressed in this paper as motivational factors (see below). It is possible that different levels of motivation might explain at least some of the differences in achievement in capital vs. regional schools. The social environment manifests itself in different types of *habitus* (Bourdieu, 1979), which has been linked to specific career aspirations among young people in Iceland (Vilhjálmsdóttir and Arnkelsson, 2013).

In response to observed differences between rural and urban schools, Harmon, Henderson, and Royster (2003) suggest different avenues of research to investigate the differences in academic performance in the US and how to improve the situation. They cite the importance of having equally good instructional resources, including digital technology and textbooks, which might be lacking in rural areas. Another disadvantage of rural schooling lies in the cost effectiveness of education per pupil. A weakness might also lie in the quality of the leadership "if school and district leaders lack adequate knowledge of mathematics and science reform movements in general" (Harmon, Henderson and Royster, 2003, p. 54).

This point is relevant in Iceland where very small municipalities have taken over educational responsibilities from the state, with arguably limited resources to fulfill these responsibilities. In Finland, government cutbacks during 2011–2015 have led to a situation where municipalities have taken over extra financial responsibilities in terms of personnel and the classes that are offered. The costs per student have not developed in a unified manner for education providers of different sizes. For large providers, i.e. large cities, the costs per student have decreased, while for small providers the opposite is true.

7.5 Motivational factors

As pointed out above, rural students tend to have weaker beliefs in their own academic ability and did not pursue additional educational opportunities compared to their suburban and urban peers (Young, 2006). Such low evaluation of one's abilities is likely to be reflected in the motivation and ambitions that pupils in rural areas have in terms of their future, and this might be reflected in their grades.

Overall, on the topic of motivation, Harmon, Henderson, and Royster (2003) argue that there is a need to improve community expectations of youth to *achieve* in science. A recent report in Iceland shows that there is a large gap between urban and rural environments in terms of adult education. Around 56% of women and 43% of men in the capital region had tertiary education, but only 41% of women and 20% of men in other regions had similar levels of education. The share of people with only compulsory education was roughly twice as high outside the capital region as in it (Statistics Iceland, 2018).

Analysis of PISA data from participating countries indicates that "[c]ontrary to conventional wisdom about big city schools, PISA finds that students in these schools generally perform better than those attending schools in non-urban settings" (OECD, 2013, p. 4). Among explanatory variables are the socio-economic status of students, better resources, greater autonomy in how they allocate those resources, and an adequate supply of teachers (OECD, 2013).

The PISA 2009 data show that in all the Nordic countries except Denmark urban schools have a higher socio-economic background compared with rural schools. School size is larger in urban areas in most participating countries, including all of the Nordic countries. In Iceland, the proportion of qualified teachers is greater in urban areas. This is also observed in many non-OECD countries, but this variable is generally not important in OECD countries. It is also found that urban areas in certain countries tend to enjoy better disciplinary climates, which in turn leads to greater student achievement (OECD, 2013, p. 3). However, the opposite is true in many countries, i.e. the disciplinary climate is better in the rural areas.

Overall, one can argue that it is not urbanization per se that explains the better performance, but the presence of underlying variables in those urban areas, such as disciplinary climate, which, if they are favorable, lead to better performance. Any analysis of urban-rural differences, or metropolitan vs. regional differences, must therefore go deeper and identify those variables.

7.6 Research questions

The general question was if the observed differences in science achievement between the capital areas (Reykjavik in Iceland and Helsinki in Finland) and the rest of the countries can be explained by the selected background variables. Two specific research questions were considered that approach the general question from slightly differing perspectives. Both of the specific research questions could be analyzed by appropriate linear statistical modeling.

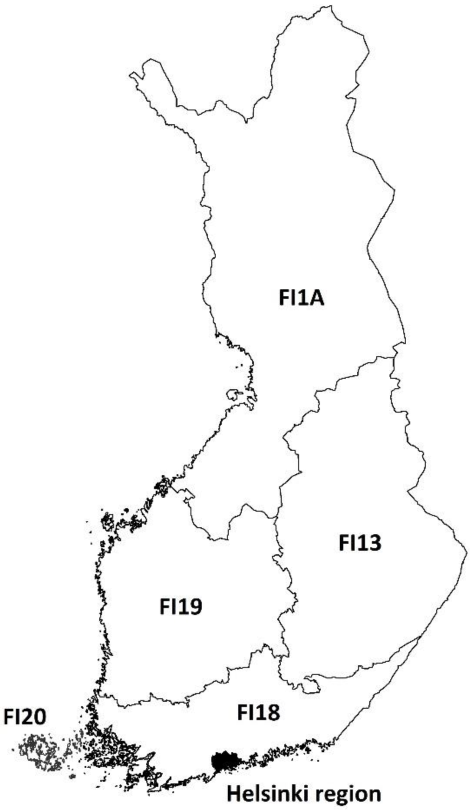
Research question 1. Can the performance differences between the capital area and the rest of the country be reproduced with background variables? In other words, are there background variables available that can predict the regional means of science achievement with high precision?

Research question 2. Can the significant regional mean differences be eliminated by controlling for some background variables?

7.7 Regional means of scientific literacy in Finland and Iceland

The Finnish PISA 2015 data set consisted of 5,882 students from 171 schools. For purposes of this study, the data set was split into five geographical regions as follows: (1) Helsinki metropolitan region, (2) Southern Finland, (3) Western Finland, (4) Eastern Finland and (5) Northern Finland. The basis of this regional classification was the 2006 version of European Union's Classification of Territorial Units for Statistics (NUTS) (http://ec.europa.eu/eurostat/web/nuts). The level 2 NUTS regions are consistently used in the Finnish PISA sampling design as explicit strata. The map in Figure 1 illustrates the Finnish NUTS 2 regional classification. Region FI18 is Southern Finland, FI19 is Western Finland, FI13 is Eastern Finland, and FI1A is Northern Finland. The small region FI20 is the autonomous Swedish-speaking archipelago province of Åland with some 25,000 inhabitants.

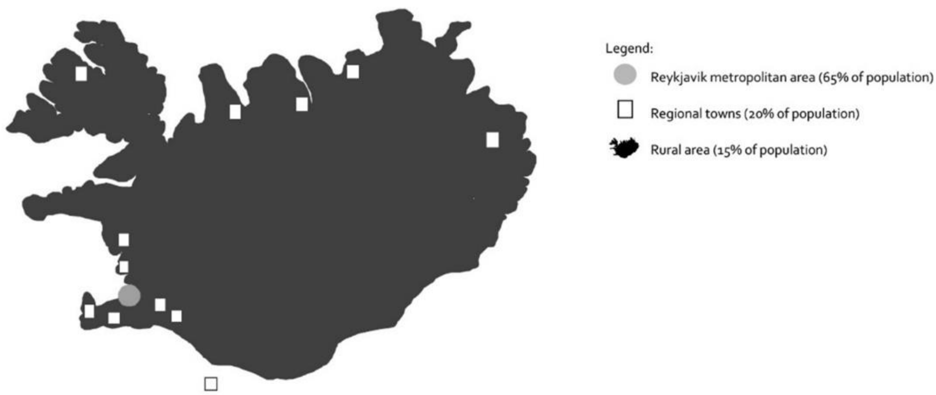
Figure 1: The Finnish NUTS 2 regional classification, 2006 version



The five-region classification employed in this study differed from Figure 1 in that Åland was joined to Southern Finland and, more importantly, the four municipalities which form the Helsinki metropolitan area (Helsinki, Espoo, Vantaa ,and Kauniainen) were separated from Southern Finland into a region of its own. This area is located in the middle of the southern coast of Finland, and it is indicated in the map with dark color. Considering the Helsinki area as a separate region is justified in many ways. The Helsinki region, with one million people, contains one fifth of the Finnish population and is the only metropolitan area in Finland. It is distinguished from the rest of the country (and even from the rest of Southern Finland) in the sense of socio-economic status, economic and cultural structure, and lifestyle. In the PISA 2015 sampling design, the Helsinki region was part of a larger geographical stratum, namely Helsinki-Uusimaa, which also contained smaller towns and countryside around the metropolitan area. As a result, the Helsinki region is slightly underrepresented in the data, but its sample size is still large enough to enable statistical analyses of sufficient power.

The Icelandic PISA 2015 data set consisted of 3,371 students from 124 schools. In principle, the Icelandic student data contained the whole 15-year-old population, i.e. the data were a census. In Iceland, a large part of the population is concentrated in Reykjavik and adjacent municipalities. No population center outside Reykjavik comes close in terms of the number inhabitants, the amount of services, etc. Thus, it made sense to contrast the capital area with the rest of the country. The situation in educational matters is, however, special for the smaller and perhaps isolated purely rural schools, and we therefore wanted to divide the area outside the capital area into two, i.e. smaller towns (with populations between 2,000 and 18,000 inhabitants) (see Figure 2) and rural areas with fewer inhabitants.

Figure 2: The Icelandic regional classification: Reykjavik metropolitan area, regional towns, and rural areas



The variable of interest was scientific literacy, the main assessment domain of PISA 2015. The Finnish regional mean scores for this assessment are shown in Table 1, and the Icelandic means are shown in Table 2. Additionally, the tables show the levels of significance of the pairwise differences between the capital areas (Helsinki, Reykjavik) and the other areas. These analyses followed the recommended PISA methodology, employing 10 plausible values of scientific literacy, student weights, and Balanced Repeated Replication (BRR) variance estimation (OECD 2009). The calculations were carried out using SAS® macros that were created specifically for PISA data analysis purposes by the Australian research institute ACER.

From Tables 1 and 2, it can be seen that in the both countries the mean score in the capital area was significantly higher than the mean score in any other area. In the Finnish PISA assessments, the regional differences have historically usually been very small, and PISA 2015 was the first cycle in which differences like this were seen. The distinction between the metropolitan area and the rest of the country has been more entrenched in Iceland, both in international studies and in standardized national tests, notwithstanding a considerable variability between years, regions, and disciplines.

Table 1: Regional mean scores for scientific literacy, Finland PISA 2015

	n	Mean score	Difference relative to Helsinki	Std error of the difference	Significance level of the difference
Helsinki region	846	553.9			
Southern Finland	2,088	532.3	−21.6	8.5	$p < 0.05$
Western Finland	1,467	517.2	−36.7	8.8	$p < 0.001$
Eastern Finland	620	522.1	−31.8	9.5	$p < 0.001$
Northern Finland	861	533.3	−20.6	9.5	$p < 0.05$

Table 2: Regional mean scores for scientific literacy, Iceland PISA 2015

	n	Mean score	Difference relative to Reykjavik	Std error of the difference	Significance level of the difference
Reykjavik	2127	478.5			
Towns (pop. 2,000–18,000)	748	463.9	−14.6	4.4	$p < 0.001$
Rural areas (pop. <2,000)	488	466.4	−12.1	4.3	$p < 0.01$

7.8 Background variables for predicting regional differences

It is commonly known that metropolitan areas differ from other kinds of areas, especially rural areas, in many ways. The overall standard of living is typically higher in big cities, as is the average level of education. The employment possibilities are usually better in large population centers, and the cultural life may be richer. The schools in metropolitan areas might also have better teachers and facilities than rural schools. As a consequence, the motivation and attitudes of students towards education might be more positive among young people living in cities. A number of background variables were selected from the PISA student questionnaire and school questionnaire to determine if they could be used in explaining and predicting the observed regional achievement differences. All of these variables are listed in Tables A1 and A2 of Appendix A.

First, seven variables related to the family's socio-economic status and cultural capital were chosen from the student questionnaire. These included the PISA index of economic, social, and cultural status (variable ESCS in the PISA data set), family wealth (WEALTH), cultural possessions at home (CULTPOSS), home educational resources (HEDRES), number of musical instruments at home (ST012Q09), number of books at home (ST013Q01), and ICT resources at home (ICTRES).

Variables that measure parental education included mother's educational level (MISCED) and father's educational level (FISCED). Parental occupational status was measured by the score on the International Socio-Economic Index (ISEI) scale (Ganzeboom et al. 1992). In the PISA data set, mothers' ISEI scores were in the variable BMMJ1 and fathers' ISEI scores were in the variable BFMJ2. After consideration, we decided to use the variable HISEI, the higher of these two, in the analyses.

It should be noted that the ESCS index is a composite score derived from home possessions, parental education, and occupation, and therefore it overlaps with many of the variables mentioned above (OECD 2018, Ch. 16). The correlations between these and other variables are given in Appendix B. The effect of these correlations on the reported analyses will be discussed briefly below.

In the PISA student questionnaire, the students were also asked which kind of profession they expected to have when they are approximately 30 years old. Their answers were then scored on the ISEI scale to form the variable BSMJ, the student's expected occupational status. Additional student variables that were considered were the student's expected educational level on the ISCED scale (ST111Q01), achieving motivation (MOTIVAT), enjoyment of science (JOYSCIE), and instrumental motivation to learn science (INSTSCIE).

Variables related to schoolwork were also considered. These included science learning time in minutes (SMINS), parents' emotional support (EMOSUPS), and how often student arrived late for school (ST062Q03).

From the PISA school questionnaire, eight variables were selected. These included student behavior hindering learning (STUBEHA), teacher behavior hindering learning (TEACHBEHA), percentage of certified teachers (of all teachers) in the school (PROATCE), percentage of certified science teachers (PROSTCE), shortage of educational staff (STAFFSHORT), shortage of educational material (EDUSHORT), science-specific resources of the school (SCIERES), and the ratio of total enrollment and number of teachers in the school (STRATIO). In addition, the disciplinary climate in science classes (DISCLISCI), as assessed by students, was considered.

Outside the variables listed above, there certainly are background variables that might be closely related to student's performance in science. A typical example is gender, and in Finland there is a statistically significant gender difference favoring girls (in Iceland, however, this difference is small and not significant). However, when the target is to examine and predict regional differences, gender is of no use because the gender mix is practically equal in all considered regions and regional differences cannot be reduced to gender imbalances in different areas. Nevertheless, we performed all main analyses separately for girls and boys, and the results were strikingly similar. Thus, in what follows we present results for the whole data only (i.e. girls and boys were analyzed together).

The Finnish regional means of the listed student and school questionnaire variables are given in Tables 3 and 4, respectively. The significances of pairwise differences between the capital and other areas were tested with asymptotic t-tests, and the standard errors of the differences were estimated using student weights and the BRR method (OECD 2009). The differences and their standard errors are given in Tables C1 and C2 of Appendix C.

It is noted that the Helsinki region stands out from other areas in almost every student-level aspect. The only variables that showed no statistically significant differences at all were the number of musical instruments and science learning time. The situation was different for the school-level variables, and there were few significant differences between Helsinki and the rest. The only difference worth mentioning is that the level of science-specific resources seems higher in Helsinki-area schools than elsewhere.

Table 3: The Finnish regional means of student and home-related background variables. The stars indicate the significance level of the difference relative to the Helsinki mean

Variable	Helsinki	Southern	Western	Eastern	Northern
Economic, social, and cultural status	0.61	0.28***	0.15***	0.09***	0.15***
Family wealth	0.19	0.24	0.16	0.01***	0.06*
Cultural possessions at home	0.43	0.15**	0.07***	−0.03***	0.16*
Home educational resources	−0.09	−0.30***	−0.32***	−0.48***	−0.36***
ICT resources at home	0.26	0.16*	0.07***	−0.09***	0.02***
Number of musical instruments at home	1.62	1.55	1.56	1.43	1.67
Number of books at home	3.86	3.47***	3.30***	3.14***	3.45**
Mother's educational level	5.43	5.02***	4.84***	4.91***	5.01***
Father's educational level	5.27	4.69***	4.58***	4.48***	4.58***
Parental occupational status	63.0	53.5***	50.4***	48.9***	49.6***
Student's expected occupational status	63.7	55.6***	53.9***	52.2***	54.9***
Student's expected educational level	3.51	3.02***	2.88***	2.80***	2.92***
Student's achieving motivation	−0.36	−0.59***	−0.68***	−0.85***	−0.72***
Enjoyment of science	0.09	−0.07	−0.12**	−0.16*	−0.12**
Instrumental motivation to learn science	0.30	0.14**	0.11***	0.11***	0.17*
Science learning time	177	169	163	163	183
Parents' emotional support	0.23	−0.06***	−0.05***	−0.06***	−0.20***
Student arriving late for school	1.67	1.55**	1.48***	1.36***	1.39***

Note: * p ≤ 0.05.
** p ≤ 0.01.
*** p ≤ 0.001.

Table 4: The Finnish regional means of school-related background variables. The stars indicate the significance level of the difference relative to the Helsinki mean

Variable	Helsinki	Southern	Western	Eastern	Northern
Student behavior hindering learning	0.29	0.24	0.13	0.47	0.47
Teacher behavior hindering learning	0.14	0.04	−0.24	0.22	0.25
% certified teachers in school	0.88	0.92	0.93	0.92	0.97*
% certified science teachers in school	0.94	0.96	0.92	0.94	0.99
Shortage of educational staff	−0.08	0.05	−0.06	0.19	−0.11
Shortage of educational material	0.08	0.17	−0.01	0.08	0.10
Science-specific resources of school	4.37	3.57*	3.61*	3.47	3.37*
Number of students per teacher in school	10.2	10.9	9.9	9.3*	10.5
Disciplinary climate in science classes	−0.01	−0.09	−0.17	−0.11	−0.07

Note: * p ≤ 0.05.
** p ≤ 0.01.
*** p ≤ 0.001.

All statistics were calculated at the student level. For the school questionnaire variables (Table 4), this means, for example, that the mean proportion of 0.88 of certified teachers in the Helsinki region is the average proportion over all students. It is not the direct school average (there is variation in school sizes). The interpretation is that in the Helsinki region, on average, 88% of a student's teachers are certified.

The student questionnaire background variables in the Icelandic data (Table 5) showed similar differences as in the Finnish case, and the Reykjavik means were repeatedly higher than the other groups' means, with only two exceptions (instrumental motivation, time spent learning). However, the results for the school questionnaire variables (Table 6) were different from the Finnish results. Reykjavik schools seem to have better teacher and other resources and less behavioral problems than the schools in other areas. The differences and their standard errors are given in Tables C3 and C4 of Appendix C. For the school questionnaire variables, even small differences can be highly significant. This is due to the fact that the Icelandic data cover practically all schools in the country, and the finite population correction, which is implicitly built into the BRR variance estimation method, makes the standard errors very small. Too much emphasis should not be given to the significance tests here.

Table 5: The Icelandic regional means of student and home-related background variables. The stars indicate the significance level of the differences relative to the Reykjavik mean

Variable	Reykjavik	Towns (pop. 2,000–18,000)	Rural areas (pop. <2,000)
Economic, social and cultural status	0.88	0.58***	0.41***
Family wealth	0.31	0.25*	0.18***
Cultural possessions at home	0.73	0.52***	0.58*
Home educational resources	0.63	0.50***	0.45***
ICT resources at home	0.44	0.34**	0.25***
Number of musical instruments at home	2.69	2.56*	2.81*
Number of books at home	4.01	3.65***	3.59***
Mother's educational level	5.00	4.68***	4.28***
Father's educational level	4.78	4.23***	3.94***
Parental occupational status	64.2	56.0***	52.7***
Student's expected occupational status	63.1	60.9**	54.2***
Student's expected educational level	4.42	4.23*	3.73***
Student's achieving motivation	0.49	0.25***	0.18***
Enjoyment of science	0.26	0.00***	−0.03**
Instrumental motivation to learn science	0.24	0.22	0.15
Science learning time	138	136	135
Parents' emotional support	0.28	0.27	0.03***
Student arriving late for school	1.81	1.69**	1.71*

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

Table 6: The Icelandic regional means of school-related background variables. The stars indicate the significance level of the differences relative to the Reykjavik mean

variable	Reykjavik	Towns (pop. 2,000–18,000)	Rural areas (pop. <2,000)
Student behavior hindering learning	−0.09	−0.65***	−0.80***
Teacher behavior hindering learning	0.23	−0.43***	−0.18***
% certified teachers in school	0.90	0.88***	0.70***
% certified science teachers in school	0.93	0.86***	0.81***
Shortage of educational staff	−0.12	−0.46***	−0.48***
Shortage of educational material	−0.37	−0.42***	−0.48***
Science-specific resources of school	3.66	2.97***	2.78***
Number of students per teacher in school	10.1	10.3***	8.7***
Disciplinary climate in science classes	0.05	−0.05*	−0.04*

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

7.9 Correlations between background variables and scientific literacy

Before tackling the actual research questions, some correlation analyses were carried out to determine how the selected background variables were related to students' scientific literacy scores. The correlations between student and home-related variables and scientific literacy are shown in Table 7. Due to the large samples, both in Finland and in Iceland almost all correlations were statistically significant, although they were not very strong. In Finland, the strongest correlations were observed with student's expected occupational status, number of books at home, student's expected educational level, socio-economic status, and enjoyment of science. In Iceland, the strongest correlations were observed with number of books at home, enjoyment of science, and student's expected educational level. Because these variables also showed significant regional differences (Tables 3 and 5), it could be anticipated that they would play a role in explaining the performance differences between capital area and the other regions.

Table 7: Pearson's correlation coefficients of student and home-related background variables with scientific literacy in Finland and Iceland. Stars indicate the significance level of the correlation

Variable	Correlation in Finland	Correlation in Iceland
Economic, social, and cultural status	0.32***	0.22***
Family wealth	0.02	−0.11***
Cultural possessions at home	0.26***	0.24***
Home educational resources	0.10***	0.15***
ICT resources at home	0.06**	0.01
Number of musical instruments at home	0.20***	0.18***
Number of books at home	0.34***	0.34***
Mother's educational level	0.22***	0.18***
Father's educational level	0.19***	0.14***
Parental occupational status	0.28***	0.18***
Student's expected occupational status	0.37***	0.25***
Student's expected educational level	0.32***	0.33***
Student's achieving motivation	0.22***	0.26***
Enjoyment of science	0.32***	0.34***
Instrumental motivation to learn science	0.18***	0.10***
Science learning time	0.23***	0.01
Parents' emotional support	0.11***	0.12***
Student arriving late for school	−0.18***	−0.17***

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

It is obvious that many of the background variables were correlated with each other. The correlation matrix of student and home-related background variables for Finland and for Iceland are shown in Tables B1 and B2 of Appendix B. The Finnish and Icelandic correlation matrices were very similar. It is natural that the socio-economic index ESCS is highly correlated with variables measuring families' resources and educational and occupational background, many of which are also correlated with each other, because ESCS is a combination of these variables (see Appendix A). Variables measuring student's motivation and expectations were also inter-correlated, and they also had some positive correlation with socio-economic status. The socio-economic status and student's motivation and expectations were associated rather weakly with student's attitudes and schoolwork.

Table 8: Pearson's correlation coefficients of school-related background variables with scientific literacy in Finland and Iceland. Stars indicate the significance level of the correlation

Variable	Correlation in Finland	Correlation in Iceland
Student behavior hindering learning	−0.02	−0.02
Teacher behavior hindering learning	0.01	0.01
% certified teachers in school	0.03	0.02
% certified science teachers in school	0.05	0.06***
Shortage of educational staff	−0.05*	−0.01
Shortage of educational material	0.01	−0.01
Science-specific resources of school	0.05	0.04*
Number of students per teacher in school	0.01	−0.07***
Disciplinary climate in science classes	0.11***	0.08***

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

The correlations of school-level variables with students' science scores were very low and often not significant (Table 8). In Finland, the staff shortage had a negative association with student performance, whereas in Iceland the percentage of certified science teachers had a positive association and student-teacher ratio had a negative association (that is, fewer students per teacher was associated with better results). It can be anticipated that the regional differences in school variables do not necessarily carry over to the regional score differences because they do not have much explanatory power.

The inter-correlations of the school-related variables are given in Tables B3 (Finland) and B4 (Iceland) of Appendix B. They are mainly very low. We also examined the correlations between school-related and student and home-related variables. These correlations were negligible.

In what follows, the regional differences in scientific literacy were analyzed by multiple regression methods, and under this methodology correlated background variables might cause multicollinearity problems. In our case, however, the methodology was not applied in the usual way, and multicollinearity did not become an issue. This is because the target of the analysis was to predict the regional means as precisely as possible, or alternatively, to determine if there is a group of background variables that can eliminate the regional differences. It is the group of predictors that is relevant here, and the individual regression coefficients and their comparative magnitudes are not of interest. In both approaches, forward selection was employed in choosing important variables, and the criterion for importance was the prediction precision instead of the statistical significance of individual variables.

7.10 Predicting regional mean scores with background variables

The answer for the first research question was obtained with the following approach. The target was to examine how well the regional mean scores of scientific literacy could be predicted (or be reproduced) with the selected background variables. A series of two-level linear regression models were fitted, where the science score was the response and each background variable served first as the single fixed covariate. The random effect of school was included in the model to account for the intra-cluster correlation of students within a school. After the model was estimated, it was used in predicting the science score of each student, and finally the regional means of these predictions were computed.

For each background variable, there was now a set of predicted or reproduced regional means of science scores. The best predictor was the variable that gave the most precise predictions of regional means in terms of relative prediction error (the smaller the error the better the predictor). In this analysis the mean of 10 plausible values of scientific literacy was chosen as the response of the two-level model. This decision simplified the analyses, but it underestimated the national variability of science proficiency. However, our purpose was to predict regional means instead of estimating variability. For this purpose, the simplified approach was valid because it did not introduce any bias in the point estimation of the regression coefficients or the prediction of the means. The computations were performed with the MIXED procedure of the SAS® software using REML estimation and student weights.

Next, the best predictor variable was kept in the model and the remaining background variables were added to the model one-by-one to determine if the prediction precision could be improved by including more variables in the predictive model. This was repeated in a loop until the relative prediction errors of the regional means could not be made smaller with additional variables.

The results for Finland are reported next, and the respective results for Iceland after that.

In the Finnish data, the best single predictor was the economic, social, and cultural status (ESCS). That is, it gave predictions that were the closest to the observed regional means. The predictions got even better by adding student's expected occupational status (BSMJ) first and then the number of books at home (ST013Q01). After this, no additional variables improved the predictions. All of the predictors in the models were statistically significant, and their effects on the science score were positive.

In what follows, Model 1 contained the socio-economic index as the sole predictor, Model 2 contained the socio-economic index and student's expected occupational status, and Model 3 contained the socio-economic index, the expected occupational status, and the number of books at home.

Tables 9 and 10 show the statistics of the predictions.

Table 9: The observed and predicted regional means of the PISA science score in Finland

Mean science score	Region				
	Helsinki	Southern	Western	Eastern	Northern
Observed	553.9	532.3	517.2	522.1	533.3
Predicted / model 1	543.3	531.2	526.6	524.1	526.4
Predicted / model 2	553.9	534.4	528.0	524.2	530.0
Predicted / model 3	554.8	534.4	527.8	521.8	531.5

Table 10: The relative errors (%) of the predicted regional means of the PISA science score in Finland

	Region					
	Helsinki	Southern	Western	Eastern	Northern	Average error (%)
Relative error / model 1	−1.91	−0.22	1.82	0.38	−1.29	1.12
Relative error / model 2	0.00	0.39	2.09	0.41	−0.62	0.70
Relative error / model 3	0.16	0.40	2.05	−0.06	−0.33	0.60

For some reason, the mean of Western Finland was the most difficult to predict, and the Western Finland students performed worse than expected on the basis of background variables. The prediction error was about 10 points on the original PISA science scale (Table 9), while in the other regions the prediction error was only 1–2 points.

In the Icelandic data, the best single predictor was the number of books at home (ST013Q01). However, the relative prediction error could be reduced by adding three more variables to the model. First, student's achieving motivation (MOTIVAT) was added to the model, then student/teacher ratio of the school (STRATIO) and finally ESCS, the economic, social, and cultural status. After this, additional variables did not improve the prediction precision. All of the explanatory variables in the models were statistically significant. Their effect on the science score was positive, except for STRATIO whose effect was negative.

Model 1 contained only the number of books at home, Model 2 contained the number of books at home and achieving motivation, Model 3 contained the number of books at home, achieving motivation, and student/teacher ratio, and Model 4 contained the number of books at home, achieving motivation, student/teacher ratio, and economic, social, and cultural status. Tables 11 and 12 show the statistics of the predictions.

The Reykjavik mean was most accurately predicted with Model 3 (i.e. no ESCS), but adding ESCS improved the predictions of the other regions and therefore reduced the overall (average) prediction error.

Table 11: The observed and predicted regional means of the PISA science score in Iceland

Mean science score	Region		
	Reykjavik	Towns (pop. 2,000–18,000)	Rural areas (pop. <2,000)
Observed	478.5	463.9	466.4
Predicted / model 1	477.2	469.2	467.9
Predicted / model 2	479.9	468.2	465.8
Predicted / model 3	479.6	467.4	468.1
Predicted / model 4	480.9	467.1	466.4

Table 12: The relative errors (%) of the predicted regional means of the PISA science score in Iceland

	Region			
	Reykjavik	Towns (pop. 2,000–18,000)	Rural areas (pop. <2,000)	Average error (%)
Relative error / model 1	0.27	1.14	0.32	0.58
Relative error / model 2	0.29	0.93	0.13	0.45
Relative error / model 3	0.23	0.75	0.36	0.45
Relative error / model 4	0.50	0.69	0.00	0.40

The conclusion is that the regional differences in science proficiency in both Finland and Iceland reduced to regional differences in a few given background variables. These variables grouped into socio-economic status, with some emphasis on cultural capital (number of books at home) and student's ambitions or motivation (expected occupational status, achieving motivation). In both countries, the regional differences in these background variables favored the capital area.

7.11 Regional differences when controlling for background variables

The second research question was answered using analysis of covariance. The starting point here was a one-way ANOVA model with science score (10 plausible values) as the response and the region as a categorical factor (in Finland five levels; in Iceland three levels). Then, in turn, each of the considered background variables were added to the model as covariates to control for their possibly confounding effect on the regional differences of the response. From each model, the regional means, adjusted for the covariates, were computed, and the significance levels of their differences were tested. The adjusted regional means were estimates for the "true" regional means in the case that the regions were equal with respect to the average level of the covariate. In this analysis, the usual PISA methodology (OECD 2009) was employed, and analyses were performed separately for each plausible value with sampling weights and the results were merged using the multiple imputation approach. The standard errors of the estimates were calculated by the BRR method, and the computations were performed with tailored SAS® macros.

Tables 13–15 show the regional means of the PISA science score when adjusted for the covariates that were found to be the most powerful in Finland. Controlling for the economic, social, and cultural status (ESCS) caused all other differences to lose their statistical significance except for in Western Finland (Table 13). However, controlling for student's expected occupational status (BSMJ) alone equalized the regional means even more (Table 14). Thus, in this sense BSMJ is a stronger covariate than ESCS.

Table 13: Unadjusted (observed) and adjusted regional means with ESCS as the covariate in Finland

	Observed mean	Difference relative to Helsinki	Significance level of the difference	Mean when adjusted for ESCS	Difference relative to Helsinki	Significance level of the difference
Helsinki region	553.9			540.9		
Southern Finland	532.3	−21.6	$p < 0.05$	531.6	−9.3	ns
Western Finland	517.2	−36.7	$p < 0.001$	521.5	−19.4	$p < 0.01$
Eastern Finland	522.1	−31.8	$p < 0.001$	529.7	−11.2	ns
Northern Finland	533.3	−20.6	$p < 0.05$	538.0	−2.9	ns

Table 14: Unadjusted (observed) and adjusted regional means with BSMJ as the covariate in Finland

	Observed mean	Difference relative to Helsinki	Significance level of the difference	Mean when adjusted for BSMJ	Difference relative to Helsinki	Significance level of the difference
Helsinki region	553.9			538.8		
Southern Finland	532.3	−21.6	$p < 0.05$	535.0	−3.8	ns
Western Finland	517.2	−36.7	$p < 0.001$	527.8	−11.0	ns
Eastern Finland	522.1	−31.8	$p < 0.001$	524.6	−14.2	ns
Northern Finland	533.3	−20.6	$p < 0.05$	538.1	−0.7	ns

If the three variables ESCS, BSMJ, and the number of books at home (ST013Q01), which were found to be important predictors in the analysis for the first research question, were used as covariates in the Finnish data, the results given in Table 15 were obtained. Controlling for these three variables simultaneously brought the adjusted means even closer to each other, and the mean of Northern Finland actually became the highest. So, if the level of economic, social, and cultural status, student's expectations, and the number of books at home were equal in the five regions in Finland, the regional differences in scientific literacy would practically disappear.

Table 15: Unadjusted (observed) and adjusted regional means with ESCS, BSMJ, and ST013Q01 as covariates in Finland

	Observed mean	Difference relative to Helsinki	Significance level of the difference	Mean when adjusted for ESCS, BSMJ, and ST013Q01TA	Difference relative to Helsinki	Significance level of the difference
Helsinki region	553.9			530.1		
Southern Finland	532.3	−21.6	$p < 0.05$	534.0	3.9	ns
Western Finland	517.2	−36.7	$p < 0.001$	530.7	0.6	ns
Eastern Finland	522.1	−31.8	$p < 0.001$	530.4	0.3	ns
Northern Finland	533.3	−20.6	$p < 0.05$	538.7	8.6	ns

Tables 16–19 show the regional means of the PISA science score when controlling for the covariates in the case of Iceland. According to Table 16, controlling only for the number of books was enough to make the differences between Reykjavik and the two other regions lose their statistical significance. That is, if in all regions of Iceland the homes had equal numbers of books, there would be no significant regional differences in science score.

Table 16: Unadjusted (observed) and adjusted regional means with ST013Q01 as the covariate in Iceland

	Observed mean	Difference relative to Reykjavik	Significance level of the difference	Mean when adjusted for ST013Q01TA	Difference relative to Reykjavik	Significance level of the difference
Reykjavik	478.5			476.0		
Towns (pop. 2,000–18,000)	463.9	−14.6	$p < 0.001$	469.5	−6.5	ns
Rural areas (pop. <2,000)	466.4	−12.1	$p < 0.01$	473.3	−2.7	ns

When student’s achieving motivation (MOTIVAT) was added to the model (Table 17), the adjusted regional means were even little closer to each other than in Table 16. It is interesting that the rural areas’ mean was now highest, although not significantly. This suggests that if the rural areas’ students were as motivated as those in Reykjavik, they might perform even better than their Reykjavik peer group.

Tables 18 and 19, where the student/teacher ratio (STRATIO) and the economic, social and cultural status (ESCS) were introduced as additional covariates, show only small changes compared to Tables 16 and 17. The regional differences are not significant in any of these tables. The differences between the controlled means are smallest in Table 18. We see again that controlling for ESCS (Table 19) increased the rural area mean score. Thus the students from rural areas seem to be "suffering" somewhat from their lower socio-economic status.

Table 17: Unadjusted (observed) and adjusted regional means with ST013Q01TA and MOTIVAT as covariates in Iceland

	Observed mean	Difference relative to Reykjavik	Significance level of the difference	Mean when adjusted for ST013Q01 and MOTIVAT	Difference relative to Reykjavik	Significance level of the difference
Reykjavik	478.5			475.3		
Towns (pop. 2,000–18,000)	463.9	−14.6	p < 0.001	471.2	−4.1	ns
Rural areas (pop. <2,000)	466.4	−12.1	p < 0.01	477.5	2.2	ns

Table 18: Unadjusted (observed) and adjusted regional means with ST013Q01, MOTIVAT, and STRATIO as covariates in Iceland

	Observed mean	Difference relative to Reykjavik	Significance level of the difference	Mean when adjusted for ST013Q01, MOTIVAT, and STRATIO	Difference relative to Reykjavik	Significance level of the difference
Reykjavik	478.5			476.1		
Towns (pop. 2,000–18,000)	463.9	−14.6	p < 0.001	472.6	−3.5	ns
Rural areas (pop. <2,000)	466.4	−12.1	p < 0.01	475.9	−0.2	ns

Table 19: Unadjusted (observed) and adjusted regional means with ST013Q01, MOTIVAT, STRATIO, and ESCS as covariates in Iceland

	Observed mean	Difference relative to Reykjavik	Significance level of the difference	Mean when adjusted for ST013Q01, MOTIVAT, STRATIO, and ESCS	Difference relative to Reykjavik	Significance level of the difference
Reykjavik	478.5			475.1		
Towns (pop. 2,000–18,000)	463.9	−14.6	p < 0.001	473.7	−1.4	ns
Rural areas (pop. <2,000)	466.4	−12.1	p < 0.01	478.6	3.5	ns

A comparison of the results of Finland and Iceland shows that the general socio-economic index ESCS and the number of books (which in fact is a component of ESCS) make an important covariate pair in both countries. The result that the number of books appears in the models together with ESCS indicates that ESCS alone does not sufficiently bring out the relevance of cultural capital. In Finland, however, the students' occupational expectations (i.e. what kind of job they think they will have as an adult) are even more important. They are correlated with the family's economic, social, and cultural status to some extent (Table B1 in Appendix B), but in the Finnish analysis this variable stands out in its own right. In Iceland, the occupational expectations are "replaced" with student's achieving motivation (MOTIVAT) and the school's student-teacher ratio (STRATIO). BSMJ and MOTIVAT measure at least partially the same phenomenon (they are positively correlated, although in Iceland this correlation is rather weak) because both have something to do with students' goal-setting. Achieving motivation is related to student's general ambitions, while the occupational expectations can be considered more concrete and practical. In Finland, STRATIO plays no role at all. So, in Iceland, unlike in Finland, there probably are some meaningful regional differences regarding this school characteristic.

7.12 Discussion

In PISA 2015 scientific literacy scores, both in Finland and Iceland the capital/metropolitan areas (that is, Helsinki in Finland and Reykjavik in Iceland) outperformed the other parts of the country. The empirical analyses of this paper suggest that in both countries the observed differences can essentially be explained by differences in certain background variables that primarily relate to families' socio-economic status and cultural capital as well as students' ambitions and expectations. All of these are at significantly higher levels among students in the capital areas than elsewhere.

Thus, the aims of the analysis were clearly attained, and the differences between the metropolitan and regional areas in both countries could be eliminated by controlling for a very restricted number of variables. A large part of the variables tested initially had significant correlations with science achievement, while the final models presented for Finland and Iceland had three and four variables, respectively, and the number could have been even more restricted. In Iceland, for example, controlling for only one variable (number of books at home) sufficed to eliminate the significance of the difference between the larger Reykjavik area and the smaller towns and rural areas. From a statistical point of view, one could say that it would be sufficient to increase the number of books in rural homes to eliminate these students' disadvantage. The number of books was also a useful predictor variable in Finland. Nevertheless, it is likely that other variables, correlated with number of books, contribute to the observed difference in academic achievement between the capital area and the regional areas.

There is a degree of similarity between the results for Finland and Iceland. In both countries, socio-economic status and number of books at home made a pair with predictive power. Similarly, two different aspects of student goal-setting appeared in the models for each country.

Student-teacher ratio was an important predictor in Iceland only. According to the model, one could statistically improve the educational attainment by reducing the class size in Iceland. The class sizes (number of pupils per teacher) are, however, already among the smallest in Europe. But this supports the findings of Wößmann and West (2006) that smaller class size in Iceland (and in Greece, but not in other countries in their study) improves achievement in TIMSS. That study and the present study might call for a further examination of class size in Iceland and how it affects achievement positively. However, Guðjónsdóttir and Karlsdóttir (2012) claim that the teaching quality is much more important than the actual class size.

Furthermore, goal-setting and motivational factors had strong effects in both countries. It is important for authorities to further explore what factors affect such motivation and how to improve it in areas that are lacking. In studying how a student's motivation to study science is situated within a wider cultural context – in metropolitan and regional contexts – sociological and cultural theories might be of use here in addition to purely psychological theories. These findings about motivational factors corroborate Gilbert and Yerrick's (2001) claim that low expectations in rural areas in science are a contributing factor to the achievement gap. The answer might not lie in further efforts to standardize curriculum and practices, but instead to take into account localized knowledge, values, and skills in the area and incorporate these into the classroom.

Research by Vilhjálmsdóttir (2008) on the *habitus* (Bourdieu, 1979) of Icelandic pupils at the upper-secondary level in relation to hobbies and past-times indicates that the pupils outside the capital area are less represented in a habitus group labeled *Arts*. The differences are not large, but they do indicate that there are regional differences in motivation and/or opportunities. The authors believe that habitus might affect career choice, i.e. that the choice of occupations is very much related to one's social milieu (Vilhjálmsdóttir and Arnkelsson, 2013). "Belonging to one of the habitus groups is based on pursuing certain activities and in turn that structures how occupations are perceived" (Vilhjálmsdóttir and Arnkelsson, 2013, p. 584). They found that "habitus measures were strongly linked to career variables, such as occupational perception and preferred future occupation" (Vilhjálmsdóttir and Arnkelsson, 2013, p. 581).

The data from Statistics Iceland (2018), which indicate a serious gap between urban and rural education levels, and a gender interaction with location, suggest that rural boys in particular should be the target of specific efforts by policy makers to improve their educational achievement, and to attend to their career plans in particular. It is thus important that the values and goals professed in schools coincide with the students' identities and interests.

Further study and analysis should include a detailed examination of the question in the PISA student questionnaire where students are asked about what job they think they will have when they are 30 years old. A comparison between metropolitan areas and the rest of the country would make it possible to further examine the relationship between motivation and science achievement in those areas and whether professions in science are less attractive to students outside metropolitan areas.

The capital versus regional distinction, or the urban-rural distinction, is quite crude, and one cannot draw conclusions about individual regions or areas based on this study. While the evidence in Iceland (e.g. Iceland PISA report, Halldórsson and Ólafsson, 2016) suggests that there is a difference between the capital and rural areas, it is important to keep in mind that the explanations put forward to explain underachievement in one rural area might not necessarily apply in another. However, the need to study the cultural context in order to understand what affects learning achievement in science in different places is paramount.

7.13 References

Abrams, E. and Middleton, M. (2017). Towards multidimensional approaches to research on rural science education. Cultural Studies of Science Education, 12(1) 167–176, Special issue.

Armila, P., Käyhkö, M. and Pöysä V. 2018. On the educational edges of a learning society: the Finnish hinterland as a framework of educational choices for young people. Journal of Youth Studies. DOI: 10.1080/13676261.2018.1453128

Bernelius, V. and Kauppinen, T.M. (2011). School Outcomes and Neighbourhood Effects: A New Approach Using Data from Finland. In: Neighbourhood Effects Research: New Perspectives. In van Ham, M., Manley, D., Bailey, N., Simpson, L. & Maclennan, D. (eds.). London: Springer, 225–247.

Bernelius, V. and Vaattovaara, M. 2016. Choice and segregation in the 'most egalitarian' schools: Cumulative decline in urban schools and neighbourhoods of Helsinki, Finland. Urban studies Vol. 53(15), 3155–3171.

Bourdieu, P. (1984[1979]). Distinction. A social criteria for the judgement of taste. London: Routledge.

Bäck, U-M. 2016. Rural Location and Academic Success—Remarks on Research, Contextualisation and Methodology. Scandinavian Journal of Educational Research, 60:4, 435–448. DOI: 10.1080/00313831.2015.1024163

Curtis, E., Wikaire, E., Jiang, Y.N., McMillan, L., Loto, R., Poole, P., Barrow, M., Bagg, W. and Reid, P. (2017) BMJ OPEN, 7(8). Examining the predictors of academic outcomes for indigenous Maori, Pacific and rural students admitted into medicine via two equity pathways: a retrospective observational study at the University of Auckland, Aotearoa, New Zealand.

Finnish National Agency for Education (2017). Teachers and principals in Finland 2016. http://www.oph.fi/download/185381_teachers_and_principals_in_Finland_2016_brochure.pdf

Ganzeboom, H.B.G., De Graaf, P.M. and Treiman, D.J. (1992). A Standard International Socio-Economic Index of Occupational Status. Social Science Research, 21, 1–56.

Gilbert, A. and Yerrick, R. (2001). Same school, separate worlds: A socio-cultural study of identity, resistance, and negotiation in a rural, lower track science classroom. Journal of Research in Science Teaching, 38, 574–598.

Guðjónsdóttir, H. and Karlsdóttir, J. (2012). Skóli án aðgreiningar og kennaramenntun. Tímarit um menntarannsóknir [Journal of Educational Research (Iceland)], 9, 132–152. (In Icelandic)

Halldórsson, A.M. and Ólafsson, R.F. (2016). Helstu niðurstöður PISA 2015. Reykjavík: Menntamálastofnun. (In Icelandic).

Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N. and Bernelius , V. (2016) Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data, Education Inquiry, 7:4. DOI: 10.3402/edui.v7.29413

Harmon, H.L., Henderson, S.A. and Royster, W.C. (2003). A research agenda for improving science and mathematics education in rural schools. Journal of Research in Rural Education, 18, 52–58.

Howley, C. B., & Gunn, E. (2003). Research about mathematics achievement in the rural circumstance. Journal of Research in Rural Education, 18(2), 79–100.

Mohammadpour, E. and Ghafar, M.N.A. (2014). Mathematics Achievement as a Function of Within- and Between-School Differences. Scandinavian Journal of Educational Research, 58(2), 189–221.

Mussa, R. (2013). Rural-urban differences in parental spending on children's primary education in Malawi. Development Southern Africa, 30(6), 789–811.

OECD (2009). PISA data analysis manual. SAS® second edition. Paris: OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264056251-en

OECD (2013). PISA in focus. Paris: OECD Publishing.

Piyaman, P., Hallinger, P. and Viseshsiri, P. (2017). Addressing the achievement gap Exploring principal leadership and teacher professional learning in urban and rural primary schools in Thailand. Journal of Educational Administration, 55(6), 717–734.

Smit, R., Hyry-Beihammer, E.K., Raggl, A. (2015). Teaching and learning in small, rural schools in four European countries: Introduction and synthesis of mixed-/multi-age approaches. International Journal of Educational Research, 74, 97-103.

Statistics Iceland (2018, March 8). Educational attainment of women and men differs by region. Retrieved from http://www.statice.is/publications/news-archive/social-affairs/women-and-men-in-iceland-2018/

Vettenranta, J., Välijärvi, J., Ahonen, A., Hautamäki, J., Hiltunen, J., Leino, K., Lähteinen, S., Nissinen, K., Nissinen, V., Puhakka, E., Rautopuro, J. & Vainikainen, M.-P. (2016). PISA 15 Ensituloksia. Huipulla pudotuksesta huolimatta. Opetus- ja kulttuuriministeriön julkaisuja 2016:41. (In Finnish)

Vilhjálmsdóttir, G. (2008). Habitus íslenskra ungmenna á aldrinum 19–22 ára. In Rannsóknir í félagsvísindum IX. Gunnar Þ. Jóhannesson and Helga Björnsdóttir (eds.), p. 195–202. Reykjavík: Félagsvísindastofnun Háskóla Íslands. (In Icelandic)

Vilhjálmsdóttir, G. and Arnkelsson, G.B. (2013). Social aspects of career choice from the perspective of habitus theory. Journal of Vocational Behavior, 83(3), 581–590.

Wößmann, L., West, M. (2006). Class-Size Effects in School Systems Around the

World: Evidence from Between-Grade Variation in TIMSS. European Economic Review, 50

(3), 695–736.

Young, D.J. (2006). Rural and urban differences in student achievement in science and mathematics: A multilevel analysis. School Effectiveness and School Improvement, 9(4), 386–418. DOI:10.1080/0924345980090403

7.14 Appendix A

Table A1: List of student and home-related background variables

variable name in PISA data set	Description	Scale
ESCS[38]	PISA index of economic, social and cultural status	standardized[39] interval score
WEALTH	family wealth	standardized interval score
CULTPOSS	cultural possessions at home	standardized interval score
HEDRES	home educational resources	standardized interval score
ICTRES	ICT resources at home	standardized interval score
ST012Q09	number of musical instruments at home	ordinal, range 1–4
ST013Q01	number of books at home	ordinal, range 1–6
MISCED	mother's educational level (ISCED scale)	ordinal, range 1–6
FISCED	father's educational level (ISCED scale)	ordinal, range 1–6
HISEI	parents' highest occupational status (ISEI scale)	interval score, range 10–90
BSMJ	student's expected occupational status (ISEI scale)	interval score, range 10–90
ST111Q01	student's expected educational level (ISCED scale)	ordinal, range 1–6
MOTIVAT	student's achieving motivation	standardized interval score
JOYSCIE	enjoyment of science	standardized interval score
INSTSCIE	instrumental motivation to learn science	standardized interval score
SMINS	science learning time (minutes)	interval score
EMOSUPS	parents' emotional support	standardized interval score
ST062Q03	how often student arrived late for school	ordinal, range 1–4

Table A2: List of school-related background variables

Variable name in PISA data set	Description	Scale
STUBEHA	student behavior hindering learning	standardized interval score
TEACHBEHA	teacher behavior hindering learning	standardized interval score
PROATCE	percentage of certified teachers in school	ratio
PROSTCE	percentage of certified science teachers in school	ratio
STAFFSHORT	shortage of educational staff	standardized interval score
EDUSHORT	shortage of educational material	standardized interval score
SCIERES	science-specific resources of school	interval, range 0–7
STRATIO	number of students per teacher in school	ratio
DISCLISCI	disciplinary climate in science classes	standardized interval score

[38] ESCS is a composite score derived from parental education (MISCED, FISCED), parental occupation (HISEI) and home possessions (WEALTH, CULTPOSS, HEDRES, ICTRES, ST013Q01) (OECD 2018, Ch. 16).
[39] Standardized to have a mean 0 and a standard deviation 1 for the student population in OECD countries, with each country having equal weight (OECD 2016, 205).

7.15 Appendix B

Table B1: Correlations of student and home-related background variables / Finland. Correlations which exceed 0.30 are italic

Variable name	ESCS	WEALTH	CULTPOSS	HEDRES	ICTRES	ST012Q09	ST013Q01	MISCED	FISCED	HISEI	BSMJ	ST111Q01	MOTIVAT	JOYSCIE	INSTSCIE	SMINS	EMOSUPS	ST062Q03
ESCS	1	*.40*	*.52*	*.38*	*.40*	*.34*	*.45*	*.63*	*.55*	*.86*	*.33*	.29	.20	.18	.16	.13	.21	−.01
WEALTH	*.40*	1	.18	.26	*.83*	.20	.16	.14	.13	.16	.08	.12	.09	.03	.06	−.03	.08	.01
CULTPOSS	*.52*	.18	1	*.43*	.20	*.57*	*.52*	.23	.21	.29	.22	.22	.13	.23	.16	.10	.20	−.05
HEDRES	*.38*	.26	*.43*	1	*.35*	.19	.28	.17	.15	.19	.16	.17	.16	.21	.15	.05	.24	−.06
ICTRES	*.40*	*.83*	.20	*.35*	1	.21	.19	.15	.16	.19	.10	.12	.10	.05	.07	−.01	.09	.01
ST012Q09	*.34*	.20	*.57*	.19	.21	1	*.33*	.16	.15	.21	.14	.14	.06	.09	.07	.08	.10	−.02
ST013Q01	*.45*	.16	*.52*	.28	.19	*.33*	1	.25	.21	.29	.24	.24	.15	.20	.13	.13	.12	−.04
MISCED	*.63*	.14	.23	.17	.15	.16	.25	1	*.45*	*.39*	.21	.19	.12	.11	.11	.09	.12	−.01
FISCED	*.55*	.13	.21	.15	.16	.15	.21	*.45*	1	*.37*	.20	.17	.14	.11	.10	.07	.13	.00
HISEI	*.86*	.16	.29	.19	.19	.21	.29	*.39*	*.37*	1	.28	.24	.17	.12	.10	.13	.16	.01
BSMJ	*.33*	.08	.22	.16	.10	.14	.24	.21	.20	.28	1	*.41*	*.31*	.23	.29	.15	.17	−.07
ST111Q01	.29	.12	.22	.17	.12	.14	.24	.19	.17	.24	*.41*	1	.25	.22	.21	.16	.13	−.06
MOTIVAT	.20	.09	.13	.16	.10	.06	.15	.12	.14	.17	*.31*	.25	1	.26	.23	.11	.21	−.09
JOYSCIE	.18	.03	.23	.21	.05	.09	.20	.11	.11	.12	.23	.22	.26	1	*.41*	.13	.20	−.13
INSTSCIE	.16	.06	.16	.15	.07	.07	.13	.11	.10	.10	.29	.21	.23	*.41*	1	.11	.13	−.07
SMINS	.13	−.03	.10	.05	−.01	.08	.13	.09	.07	.13	.15	.16	.11	.13	.11	1	.03	−.01
EMOSUPS	.21	.08	.20	.24	.09	.10	.12	.12	.13	.16	.17	.13	.21	.20	.13	.03	1	−.12
ST062Q03	−.01	.01	−.05	−.06	.01	−.02	−.04	−.01	.00	.01	−.07	−.06	−.09	−.13	−.07	−.01	−.12	1

Table B2: Correlations of student and home-related background variables / Iceland. Correlations which exceed 0.30 are italic

Variable name	ESCS	WEALTH	CULTPOSS	HEDRES	ICTRES	ST012Q09	ST013Q01	MISCED	FISCED	HISEI	BSMJ	ST111Q01	MOTIVAT	JOYSCIE	INSTSCIE	SMINS	EMOSUPS	ST062Q03
ESCS	1	*.38*	*.46*	*.35*	*.36*	.28	*.40*	*.64*	*.55*	*.80*	.18	.27	.20	.15	.09	.02	.22	−.02
WEALTH	*.38*	1	.17	.20	*.72*	.16	.11	.11	.14	.10	.06	.03	.09	.00	.04	.04	.10	−.01
CULTPOSS	*.46*	.17	1	*.42*	.20	**.61**	**.49**	.22	.19	.22	.10	.19	.15	.19	.09	.03	.18	−.05
HEDRES	*.35*	.20	*.42*	1	.25	.16	**.32**	.20	.16	.13	.09	.15	.19	.16	.10	.02	.26	−.13
ICTRES	*.36*	*.72*	.20	.25	1	.18	.17	.14	.14	.11	.10	.08	.11	.05	.04	.05	.10	−.03
ST012Q09	*.28*	.16	*.61*	.16	.18	1	**.31**	.12	.13	.14	.04	.10	.07	.10	.04	.02	.09	−.01
ST013Q01	*.40*	.11	*.49*	*.32*	.17	**.31**	1	.25	.21	.22	.13	.12	.16	.19	.09	.00	.14	−.07
MISCED	*.64*	.11	.22	.20	.14	.12	.25	1	*.38*	*.33*	.11	.22	.14	.11	.07	.00	.15	−.01
FISCED	*.55*	.14	.19	.16	.14	.13	.21	*.38*	1	*.31*	.15	.21	.17	.10	.04	.03	.16	.02
HISEI	*.80*	.10	.22	.13	.11	.14	.22	*.33*	*.31*	1	.14	.17	.11	.08	.07	−.02	−.12	.02
BSMJ	*.18*	.06	.10	.09	.10	.04	.13	.11	.15	.14	1	*.37*	.23	.21	.17	.02	.11	−.05
ST111Q01	.27	.03	.19	.15	.08	.10	.12	.22	.21	.17	*.37*	1	*.33*	.19	.15	.06	.18	−.10
MOTIVAT	.20	.09	.15	.19	.11	.07	.16	.14	.17	.11	.23	*.33*	1	.23	.16	.02	.27	−.15
JOYSCIE	.15	.00	.19	.16	.05	.10	.19	.11	.10	.08	.21	.19	.23	1	.23	.05	.11	−.07
INSTSCIE	.09	.04	.09	.10	.04	.04	.09	.07	.04	.07	.17	.15	.16	.23	1	.07	.08	−.06
SMINS	.02	.04	.03	.02	.05	.02	.00	.00	.03	−.02	.02	.06	.02	.05	.07	1	.00	−.02
EMOSUPS	.22	.10	.18	.26	.10	.09	.14	.15	.16	−.12	.11	.18	.27	.11	.08	.00	1	−.14
ST062Q03	−.02	−.01	−.05	−.13	−.03	−.01	−.07	−.01	.02	.02	−.05	−.10	−.15	−.07	−.06	−.02	−.14	1

Table B3: Correlations of school-related background variables / Finland. Correlations which exceed 0.30 are italic

Variable name	STUBEHA	TEACHBEHA	PROATCE	PROSTCE	STAFFSHORT	EDUSHORT	SCIERES	STRATIO	DISCLISCI
STUBEHA	1	*.36*	.06	.01	.23	.22	–.03	.09	–.04
TEACHBEHA	*.36*	1	.01	.04	.23	.09	.06	.07	.01
PROATCE	.06	.01	1	**.30**	–.01	–.04	.08	.08	–.04
PROSTCE	.01	.04	**.30**	1	.02	–.01	.06	.10	–.06
STAFFSHORT	.23	.23	–.01	.02	1	*.39*	–.24	.06	–.03
EDUSHORT	.22	.09	–.04	–.01	*.39*	1	*–.45*	.04	–.04
SCIERES	–.03	.06	.08	.06	–.24	*–.45*	1	.00	.01
STRATIO	.09	.07	.08	.10	.06	.04	.00	1	.11
DISCLISCI	–.04	.01	–.04	–.06	–.03	–.04	.01	.11	1

Table B4: Correlations of school-related background variables / Iceland. Correlations which exceed 0.30 are italic

Variable name	STUBEHA	TEACHBEHA	PROATCE	PROSTCE	STAFFSHORT	EDUSHORT	SCIERES	STRATIO	DISCLISCI
STUBEHA	1	*.44*	–.04	.06	**.32**	.04	.10	.09	–.08
TEACHBEHA	*.44*	1	–.10	–.01	.29	.03	.14	.06	–.04
PROATCE	–.04	–.10	1	.20	.01	–.07	.14	.00	.02
PROSTCE	.06	–.01	.20	1	–.05	–.09	.16	–.06	.06
STAFFSHORT	**.32**	.29	.01	–.05	1	.05	.04	–.05	–.01
EDUSHORT	.04	.03	–.07	–.09	.05	1	*–.31*	.12	.05
SCIERES	.10	.14	.14	.16	.04	*–.31*	1	–.09	.09
STRATIO	.09	.06	.00	–.06	–.05	.12	–.09	1	.08
DISCLISCI	–.08	–.04	.02	.06	–.01	.05	.09	.08	1

7.16 Appendix C

Table C1: The mean differences of student and home-related background variables between Helsinki and other Finnish regions. Standard errors are in the parentheses

	Difference to Helsinki mean			
Variable	Southern	Western	Eastern	Northern
Economic. social and cultural status	−0.33 (0.07) ***	−0.46 (0.07) ***	−0.52 (0.07) ***	−0.46 (0.08) ***
Family wealth	0.05 (0.05)	−0.03 (0.05)	−0.18 (0.05) ***	−0.14 (0.06) *
Cultural possessions at home	−0.28 (0.10) **	−0.36 (0.09) ***	−0.47 (0.10) ***	−0.27 (0.11) *
Home educational resources	−0.21 (0.05) ***	−0.23 (0.05) ***	−0.39 (0.06) ***	−0.27 (0.07) ***
ICT resources at home	−0.10 (0.04) *	−0.19 (0.04) ***	−0.35 (0.04) ***	−0.25 (0.05) ***
Number of musical instruments at home	−0.08 (0.10)	−0.07 (0.10)	−0.19 (0.10)	0.04 (0.10)
Number of books at home	−0.39 (0.12) ***	−0.56 (0.11) ***	−0.72 (0.12) ***	−0.41 (0.13) **
Mother's educational level	−0.41 (0.08) ***	−0.58 (0.08) ***	−0.51 (0.09) ***	−0.41 (0.09) ***
Father's educational level	−0.58 (0.11) ***	−0.69 (0.10) ***	−0.79 (0.11) ***	−0.70 (0.12) ***
Parental occupational status	−9.48 (1.66) ***	−12.59 (1.64) ***	−14.12 (1.82) ***	−13.38 (2.18) ***
Student's expected occupational status	−8.17 (1.22) ***	−9.80 (1.15) ***	−11.55 (1.45) ***	−8.82 (2.15) ***
Student's expected educational level	−0.49 (0.13) ***	−0.63 (0.11) ***	−0.71 (0.11) ***	−0.60 (0.13) ***
Student's achieving motivation	−0.23 (0.05) ***	−0.32 (0.05) ***	−0.49 (0.05) ***	−0.36 (0.05) ***
Enjoyment of science	−0.16 (0.08)	−0.21 (0.07) **	−0.25 (0.10) *	−0.21 (0.08) **
Instrumental motivation to learn science	−0.15 (0.06) **	−0.19 (0.06) ***	−0.19 (0.05) ***	−0.12 (0.06) *
Science learning time	−7.20 (10.56)	−14.16 (11.23)	−14.15 (12.24)	6.23 (11.51)
Parents' emotional support	−0.29 (0.05) ***	−0.28 (0.05) ***	−0.29 (0.05) ***	−0.43 (0.07) ***
Student arriving late for school	−0.12 (0.04) **	−0.19 (0.05) ***	−0.31 (0.06) ***	−0.28 (0.04) ***

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

Table C2: The mean differences of school-related background variables between Helsinki and other Finnish regions. Standard errors are in the parentheses

	Difference to Helsinki mean			
Variable	Southern	Western	Eastern	Northern
Student behavior hindering learning	−0.05 (0.18)	−0.16 (0.18)	0.18 (0.22)	0.18 (0.24)
Teacher behavior hindering learning	−0.10 (0.21)	−0.38 (0.22)	0.08 (0.26)	0.11 (0.24)
% certified teachers in school	0.04 (0.04)	0.06 (0.05)	0.04 (0.07)	0.09 (0.04) *
% certified science teachers in school	0.02 (0.05)	−0.02 (0.05)	0.00 (0.07)	0.04 (0.05)
Shortage of educational staff	0.13 (0.19)	0.02 (0.22)	0.27 (0.22)	−0.03 (0.22)
Shortage of educational material	0.08 (0.21)	−0.09 (0.24)	−0.01 (0.28)	0.02 (0.25)
Science–specific resources of school	−0.80 (0.32) *	−0.76 (0.38) *	−0.90 (0.57)	−1.00 (0.41) *
Number of students per teacher in school	0.74 (0.38)	−0.27 (0.42)	−0.89 (0.38) *	0.30 (0.43)
Disciplinary climate in science classes	−0.08 (0.08)	−0.16 (0.09)	−0.10 (0.10)	−0.06 (0.09)

Note: * p≤0.05.
** p≤0.01.
*** p≤0.001.

Table C3: The mean differences of student and home-related background variables between Reykjavik and other Icelandic regions. Standard errors are in the parentheses

	Difference to Reykjavik mean	
Variable	towns (pop. 2,000–18,000)	rural areas (pop. < 2,000)
Economic. social and cultural status	−0.29 (0.03) ***	−0.47 (0.04) ***
Family wealth	−0.06 (0.03) *	−0.13 (0.02) ***
Cultural possessions at home	−0.21 (0.05) ***	−0.15 (0.06) *
Home educational resources	−0.14 (0.04) ***	−0.18 (0.04) ***
ICT resources at home	−0.10 (0.03) **	−0.19 (0.04) ***
Number of musical instruments at home	−0.12 (0.05) *	0.12 (0.05) *
Number of books at home	−0.36 (0.05) ***	−0.41 (0.06) ***
Mother's educational level	−0.32 (0.07) ***	−0.72 (0.08) ***
Father's educational level	−0.55 (0.06) ***	−0.84 (0.08) ***
Parental occupational status	−8.16 (0.85) ***	−11.45 (1.02) ***
Student's expected occupational status	−2.25 (0.87) **	−8.92 (0.96) ***
Student's expected educational level	−0.19 (0.07) *	−0.69 (0.08) ***
Student's achieving motivation	−0.24 (0.04) ***	−0.31 (0.05) ***
Enjoyment of science	−0.26 (0.05) ***	−0.29 (0.07) **
Instrumental motivation to learn science	−0.02 (0.04)	−0.09 (0.05)
Science learning time	−1.55 (2.43)	−2.33 (2.39)
Parents' emotional support	−0.01 (0.05)	−0.25 (0.05) ***
Student arriving late for school	−0.11 (0.04) **	−0.10 (0.04) *

Note: * p≤0.05.
** p≤0.01.
*** p≤0.001.

Table C4: The mean differences of school-related background variables between Reykjavik and other Icelandic regions. Standard errors are in the parentheses

	Difference to Reykjavik mean	
Variable	Towns (pop. 2,000–18,000)	Rural areas (pop. < 2,000)
Student behavior hindering learning	−0.56 (0.01) ***	−0.71 (0.02) ***
Teacher behavior hindering learning	−0.66 (0.02) ***	−0.41 (0.02) ***
% certified teachers in school	−0.02 (0.00) ***	−0.20 (0.01) ***
% certified science teachers in school	−0.07 (0.01) ***	−0.12 (0.01) ***
Shortage of educational staff	−0.34 (0.01) ***	−0.36 (0.02) ***
Shortage of educational material	−0.04 (0.01) ***	−0.10 (0.01) ***
Science–specific resources of school	−0.68 (0.02) ***	−0.88 (0.04) ***
Number of students per teacher in school	0.20 (0.04) ***	−1.39 (0.05) ***
Disciplinary climate in science classes	−0.10 (0.04) *	−0.09 (0.04) *

Note: * $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

Sammanfattning

Resultaten från PISA 2015 och TIMSS 2015 publicerades i november och december 2016. Alla nordiska länder deltog i PISA. Danmark, Finland, Norge och Sverige deltog i TIMSS årskurs 4 och Norge och Sverige deltog i TIMSS årskurs 8.

I denna rapport analyseras och diskuteras ett antal viktiga frågor i den utbildningspolitiska debatten:

- vikten av intresse och motivation samt feedback till elever
- hur lärare kan göra skillnad
- mätning av och effekterna av social ojämlikhet
- övergången till datorbaserade prov.

De nordiska länderna har likheter, men också skillnader, vilket gör det intressant och värdefullt att genomföra analyser i ett nordiskt perspektiv. I denna rapport har forskare från hela Norden utfört djupanalyser på viktiga teman utifrån de resultat som presenteras under 2016. Syftet med denna rapport har varit att presentera policyrelevanta analyser av TIMSS och PISA på ett sätt som är anpassat för beslutsfattare på olika nivåer i de nordiska länderna, i syfte att bidra till ytterligare utveckling inom utbildningsområdet

Forskare från alla nordiska länder har bidragit till denna rapport.

Nordic Council of Ministers
Nordens Hus
Ved Stranden 18
DK-1061 Copenhagen K
www.norden.org

Northern Lights on TIMSS and PISA 2018

The results from PISA 2015 and TIMSS 2015 were published in November and December 2016. All the Nordic countries participated in PISA. Denmark, Finland, Norway and Sweden participated in TIMSS grade 4 and Norway and Sweden participated in TIMSS grade 8.

The Nordic countries have similarities but also differences, which makes it interesting and valuable to carry out analyses in a Nordic perspective. In this report researchers from all the Nordic countries have done in-depth analyses on different policy relevant themes based on the results presented in 2016. The purpose of this report has been to present policy relevant analyses of TIMSS and PISA in a way that is accessible for policy makers on different levels in the Nordic countries, with the aim to contribute to further development in the education area.

9 789289 355650