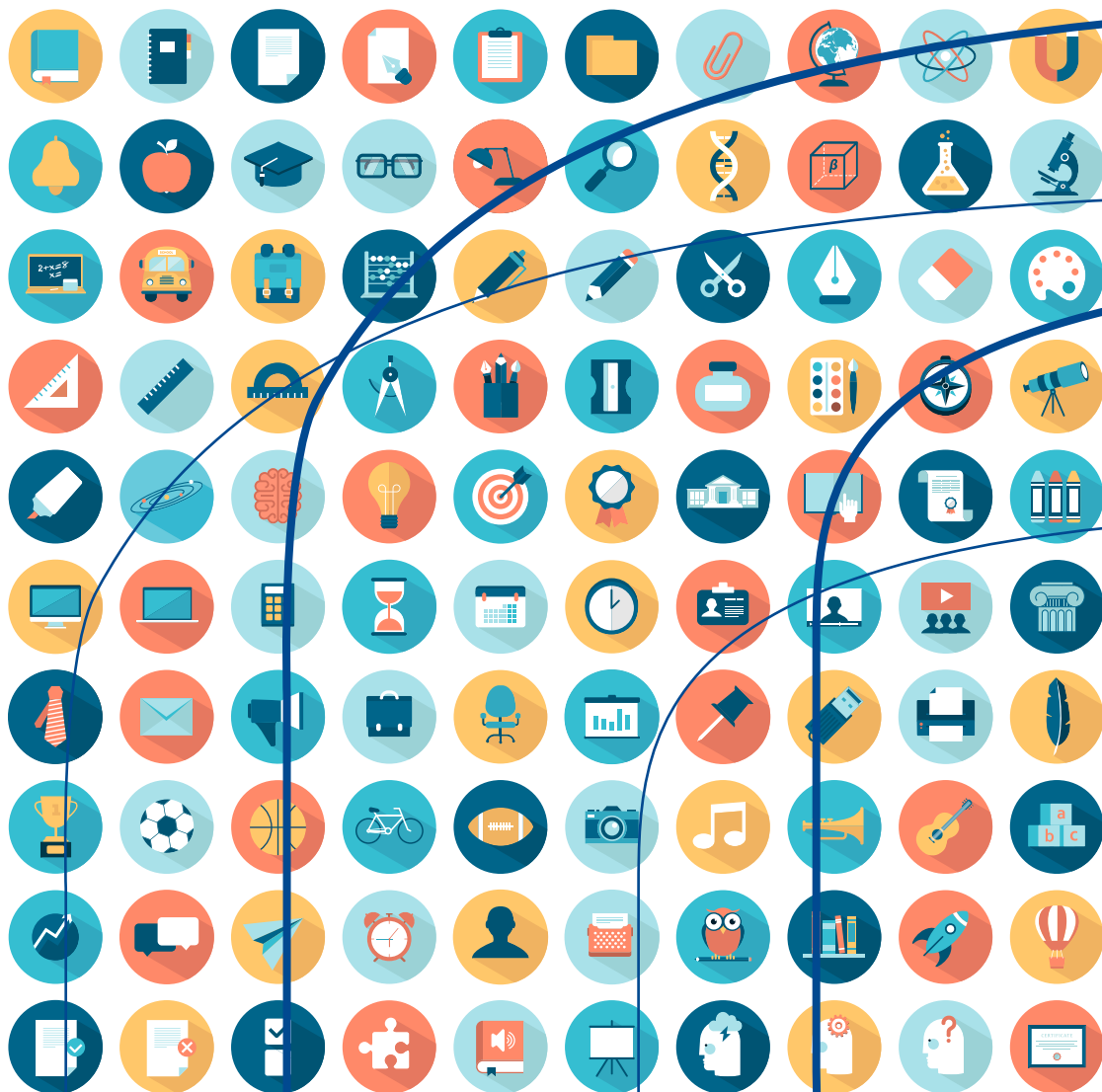


Northern Lights on PISA and TALIS





Northern Lights on PISA and TALIS

*Sten Ludvigsen, Guri A. Nortvedt, Andreas Pettersen,
Astrid Pettersson, Samuel Sollerman, Ragnar F. Ólafsson,
Matti Taajamo, Joakim Caspersen, Peter Nyström and Johan Braeken*

Northern Lights on PISA and TALIS

Sten Ludvigsen, Guri A. Nortvedt, Andreas Pettersen, Astrid Pettersson, Samuel Sollerman, Ragnar F. Ólafsson, Matti Taajamo, Joakim Caspersen, Peter Nyström and Johan Braeken

ISBN 978-92-893-4521-7 (PRINT)

ISBN 978-92-893-4523-1 (PDF)

ISBN 978-92-893-4522-4 (EPUB)

<http://dx.doi.org/10.6027/TN2016-517>

ISSN 0908-6692

© Nordic Council of Ministers 2016

Layout: Hanne Lebech

Cover photo: ImageSelect

Print: Rosendahls Schultz-Grafisk

Copies: 500

Printed in Denmark



This publication has been published with financial support by the Nordic Council of Ministers. However, the contents of this publication do not necessarily reflect the views, policies or recommendations of the Nordic Council of Ministers.

www.norden.org/nordpub

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, the economy, and culture. It plays an important role in European and international collaboration, and aims at creating a strong Nordic community in a strong Europe.

Nordic co-operation seeks to safeguard Nordic and regional interests and principles in the global community. Common Nordic values help the region solidify its position as one of the world's most innovative and competitive.

Nordic Council of Ministers

Ved Stranden 18

DK-1061 Copenhagen K

Phone (+45) 3396 0200

Contents

Foreword.....	9
1. Introduction Comparative studies of the Nordic countries: implications for educational policy	11
1.1 Introduction: The Nordic model	11
1.2 Change in educational systems.....	14
1.3 Comparisons beyond the Nordic states.....	15
1.4 PISA and TALIS studies	16
1.5 Overview of the chapters.....	18
1.6 Discussion	22
1.7 References	25
2. Is PISA 2012 relevant to mathematics education in Norway and Sweden?.....	27
2.1 Summary	27
2.2 Introduction.....	28
2.3 Methodology	30
2.4 Curriculum structures and goal definitions	32
2.5 Goals at the end level: Fundamental reasons for teaching mathematics	36
2.6 Goals at the aim level: General mathematical competence.....	39
2.7 Goals at the objective level: Mathematical content.....	43
2.8 Relevance of the PISA mathematics assessment items to mathematics education in Norway and Sweden.....	47
2.9 Conclusions	54
2.10 Acknowledgements	55
2.11 References	56
3. Principal Leadership Styles in Nordic Countries and Their Relationship with Teacher Performance and Attitudes	59
3.1 Summary	59
3.2 Introduction.....	60
3.3 Aims and research questions.....	68
3.4 Methods.....	68
3.5 Results	71
3.6 Profile descriptions of the four clusters of principals	78
3.7 Other principal characteristics associated with the four clusters.....	80
3.8 Association between the four principal clusters and teacher practices and attitudes	80
3.9 Discussion	86
3.10 References	90
3.11 Appendices	93

4. Teachers' professional development in Nordic countries.....	101
4.1 Summary.....	101
4.2 Introduction	102
4.3 Teachers' continuing professional development.....	103
4.4 Teachers' professional development in the Nordic countries.....	107
4.5 Implementations of TALIS	112
4.6 Results.....	113
4.7 Discussion	126
4.8 References.....	129
5. Can collegial work and school leader feedback improve teachers' self-efficacy in Nordic classrooms?	135
5.1 Summary.....	135
5.2 Introduction	136
5.3 Research questions and approach	139
5.4 Collegial work, feedback and appraisal in Nordic schools	141
5.5 Variation in feedback within countries.....	145
5.6 Self-efficacy in Nordic classrooms – differences and variation.....	147
5.7 The relationship between feedback and self-efficacy in the Nordic countries	148
5.8 Impact of collaboration and feedback on novice and experienced teachers.....	151
5.9 Conclusion.....	155
5.10 References.....	156
6. Characteristics of high- performing students in mathematics An exploratory analysis of PISA data from the Nordic countries.....	161
6.1 Summary.....	161
6.2 Introduction	162
6.3 Background	163
6.4 Aim.....	168
6.5 Design and method	168
6.6 Results.....	170
6.7 PISA 2003	170
6.8 PISA 2012	178
6.9 PISA 2003 and PISA 2012 comparison	184
6.10 Analysis and discussion.....	186
6.11 Further research.....	192
6.12 Conclusions.....	192
6.13 References.....	193

7. International large-scale educational assessments: Elephants at the gate?.....	195
7.1 Summary	195
7.2 Introduction.....	196
7.3 Design & Data.....	197
7.4 Statistical Analysis	203
7.5 Models as useful approximate summaries.....	204
7.6 Reproducibility and Transparency.....	206
7.7 Global Procedural Overview	207
7.8 Communication	210
7.9 Research Agenda	213
7.10 Conclusion.....	214
7.11 References	215
Sammendrag.....	217

Foreword

In this publication, scholars take a closer look at the PISA 2012 and TALIS 2013 studies. The authors represent almost all the Nordic countries and carry with them their different insights and perspectives. As the former editions in the Northern Lights series, this publication has received financial support from the Nordic Council of Ministers.

The Nordic Evaluation Network group has been responsible for editorial work. Sten Ludvigsen, Jouni Välijärvi and Jan-Eric Gustafsson have also participated to this issue and lend their expertise to the editorial group. This group has been led by Hallvard Thorsen and Marianne Nordengen from The Norwegian Directorate for Education and Training.

On behalf of the editorial group, we would like to thank all of the contributors. We would also like to thank Sten Ludvigsen for writing the introduction chapter.

We hope that Northern Lights on PISA and TALIS will be of interest to policymakers in the Nordic countries. Our aim and ambition with this publication is to give input to further policy development.

Oslo, May 2016

Marianne Nordengen and Hallvard Thorsen
The Norwegian Directorate for Education and Training

1. Introduction

Comparative studies of the Nordic countries: implications for educational policy

By Sten Ludvigsen, University of Oslo, Norway

1.1 Introduction: The Nordic model

The notion of the Nordic model of society has become popular in research and the media. Two years ago, the well-known magazine *The Economist* had a picture of a Viking on its cover, claimed that politicians from the left and the right can learn from Nordic countries (2 February, 2013). As citizens in Nordic countries, we may enjoy the idea of living in a super-model, from which the other countries can learn. The educational systems in the Nordic countries are seen as well functioning and are highly trusted by the citizens of these countries. But before examining the important dimensions of Nordic educational systems, I contextualize the recent works on these systems by presenting an overview of the Nordic model. The review distinguishes between a Nordic model of society and models that include different features.

Academics from many fields and disciplines have investigated characteristics that are common among the Nordic countries. Features often emphasised include the open, market-oriented small economies, strong institutional collaboration between key actors in society, well-developed welfare states and well-organised labour markets. Further, the gaps in income and standards of living between citizens are smaller than in most

other countries, women participate in the labour market to a higher degree and gender equality is generally well entrenched. In terms of working life, the Nordic model is often referred to as the triangle model, which refers to the comprehensive institutional coordination between key actors in the labour market, the welfare state, economic policy and the government (Dølvik *et al.*, 2015). The power of these actors is balanced when arriving at a consensus on conflicting issues, which is believed to contribute to stability and reduce economic inequality. In addition, Nordic countries have a strong tradition of investment in human resources and knowledge mobilisation of labour. These factors create the conditions for social trust and a welfare state system that is fair for all citizens.

These features of society create a large number of meeting opportunities in which key actors can meet, talk and develop a common understanding of how changes in society can be interpreted and responded to. Such coordination and collaboration also occur in the educational sector. Relationships between the governments, unions and employers' organisations are very strong and create what can be called a pragmatic stability, which often includes steps for incremental change. Historically, the relationship between these institutions and their roles has differed within the Nordic countries, and they are given different weight. However, broadly speaking, it's still reasonable to claim that the described characteristics are the common features of the Nordic national states.

Despite differing ideological stances, political parties are able to agree on solutions. To put it simply, in general, politicians in the Nordic countries seek solutions that work. It is important to emphasise that the Nordic model is dynamic and that the main elements are somewhat different between the countries. Thus far, the Nordic model has been able to ensure stability for its citizens; however, whether it can continue to do remains an open question.

A large-scale research project, called the NorMod, investigated the factors that affect the construction of Nordic societies (Dølvik *et al.*, 2015). The results showed that these societies are constructed by the cumulative effects of decisions made over long periods of time, which create the everyday social practices in which we participate. Most studies that investigate the Nordic model use a macro perspective. In such studies, the educational system is often cited as an important component of

the comprehensive system of the welfare state, but seldom analysed closely. Education and healthcare are seen as the two main systems that create the conditions for citizens' social, cognitive and emotional development and well-being. Thus, in the Nordic countries, a comprehensive welfare state ensures a secure and well-functioning society. Or, as was described in *The Economist*, Nordic citizens seem happier to pay higher taxes than anyone else in the world.

An interesting finding of the NorMod project pertains to the level of satisfaction with and the conception of the healthcare and educational systems – a comparison that was based on data from the European Social Survey ESS6 and FAFO's own analysis. The NorMod project compared Nordic countries with countries such as Germany and the United Kingdom (UK) and asked Nordic citizens to rate the overall state of education in their countries on a 10-point scale, ranging from extremely bad to extremely good. While all the Nordic countries had high scores, trust in education was the highest in Finland. More than 50% of those who participated believed that their educational systems are very good. In fact, if the scores for medium to very good were combined, almost 90% of the population thought that their educational systems work well. The scores for the educational systems were higher than those for the healthcare systems. Nordic citizens had higher trust in their educational systems than the citizens of Germany and the UK, despite some major differences between the countries (Dølvik *et al.*, 2015). Interestingly, international comparative studies such as Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) show that there is variation in performance among the Nordic countries and internally within each country (Kavli & Thorsen, 2014).

A reasonable explanation for the high levels of trust is the general level of functioning in Nordic institutions. The trust has been built through long-term development and is historically anchored. As social institutions, schools guarantee that the next generation will develop the competences needed to pursue higher education and transition into the labour market. Schools in the Nordic countries have a history of being well functioning, and the Nordic citizens appreciate the norms, values and knowledge that a functioning educational system imparts.

1.2 Change in educational systems

Nordic countries have gradually built up the capacity required for further development and change. Capacities, here, refers to how school authorities are organised, how school principals work, how the curriculum is developed, teachers' competencies, how parents are involved and how the unions and other organisations work together to develop the educational system. Changes to the system cannot too radical or they will break these well-developed, long-term models. The relationship between the actors involved explains why educational systems change slowly. The historical anchoring of the institutions and the stakeholders involved must agree on the changes that are desirable. Further, large-scale system reforms (often connected to governance and/or curriculum) and small-scale and local innovations connected to individual schools and teaching and learning need to be distinguished. Large-scale systems, such as the educational system, cannot change too quickly since many national and institutional mechanisms are involved in keeping them stabilized. Developing and adjusting new educational policies is a complex and long-term effort because it often involves measures that target a part of the system, the function of the principals or the function of teachers. In addition, the effects of new complex measures are often difficult to identify in the short term. Despite robust findings from research, scaling up the results and systemic changes takes time and long-term effort.

Today, we have a multitude of system information, research and evaluations that provide insights into how educational systems work. While some studies offer an overview, others present a more detailed view of certain practices. How such studies are connected raises the question of how to recognize different forms of evidence. Complex systems must always be understood from multiple perspectives. By this, I mean that we must seek to understand and explain how educational systems work by combining multiple sources of evidence (Gough *et al.*, 2012). In other words, as strongly recommended in recent Norwegian green papers, we need different types of studies to understand and change educational systems (NOU, 2014:7; NOU, 2015:8).

In the last 15 years, students and citizens of the Nordic countries have participated in surveys/trend studies organised by the Organization for Economic Cooperation and Development (OECD) and the International Association for the Evaluation of Educational Achievement (IEA). Such studies give the participating countries important insights into their educational systems. Most of the surveys are time series studies that make it possible to compare achievements in each country over periods of time as well those of other relevant countries. Such studies are important in monitoring the development of individual countries (Kavli & Thorsen, 2014). In the international context, open economies must know how their educational systems work and whether changes in direction are needed.

1.3 Comparisons beyond the Nordic states

Comparisons of students' performance between Nordic countries are useful as these societies have common characteristics, strong institutional coordination and many shared curriculum features. In social and educational sciences the value of comparing similar cases with extreme or deviant cases is generally well established. Beyond the Nordic region, we can compare student performance with countries in the South-East region of Asia. Schools in South-East Asia have consistently fared well in PISA and TIMSS (OECD 2014a). These trends have been coherent over time. In the everyday discourse the results are often explain by the learning culture (authoritarian) and more specifically to rote learning and the use surface learning strategies.

However, given that more than 20% and up to 50% of the South-East Asian students achieve top scores in mathematics, reading and sciences (OECD 2014a), it's difficult to believe that the results only can be explained by rote learning. A curriculum feature that is emphasised in South East Asia is the formal aspects of mathematics, and it is now being given more importance in the Nordic countries as well. Countries that perform well on problem solving in the PISA 2012 study (OECD 2014a) and Shanghai-China have very high scores in the OECD study on financial literacy (OECD 2014b). Problem solving and financial literacy can both

be seen as areas of application in which principles in mathematics are used. The use of mathematics principles in new areas is conceptualized as transfer of cognitive skills and is often connected to in-depth learning. Transfer between subjects is not possible if the students have only relied on rote learning and surface learning strategies, but cognitive mechanisms cannot fully explain the high scores emphasised here. To understand the complexities of students' performance, social and cultural explanations, related to status, social norms and expectations from parents and the school system, are needed. Other important factors are high effort, low frequency of sick leave and participation in after-school teaching programs. The teachers are very professional and work systematically, offering student-centred supervision, feedback and evaluation. Teachers' professional development is seen as part of a systematic effort to continuously improve the students' performance. These high expectations of the teachers' performance start with strategies for recruitment and the structure and content of teacher education programmes.

Thus, a multitude of factors contribute to student performance in the high-performing countries than is commonly acknowledged in the everyday discourse about students' performance. The Nordic countries (except Finland) have rather few top performers, and too many low achievers. Therefore, it may be worthwhile to explore the curriculum and the teaching and learning practices of high-performing countries. If the Nordic countries want to renew their school subjects, develop a new curriculum and enable the use of new modes of instruction that leads to more in depth learning, they may have to look for inspiration not only at countries that are similar to them but also to those that conduct schooling differently and achieve very good results.

1.4 PISA and TALIS studies

The international studies covered in this book are the PISA programme and the OECD Teaching and Learning International Survey (TALIS). Below are short descriptions of both these studies.

What is PISA?

The Programme for International Student Assessment (PISA) is an international assessment of the reading, science and mathematical literacy of 15-year-old students. The tests are designed to assess the extent to which students at the end of compulsory education can apply their knowledge to real-life situations and be equipped for full participation in society. The information collected through background questionnaires also provides a context for the application of that knowledge, which can help analysts interpret the results.

In most OECD countries, students at this age are approaching the end of compulsory schooling. PISA is conducted in 3-year cycles. Three main areas or domains are examined in every cycle, but the major domain changes with each cycle.

Around 510,000 students from 65 economies took part in the PISA 2012 assessment, representing about 28 million 15-year-olds globally. PISA provides information about education systems and allows scholars to compare students across a large number of countries.

In PISA 2000, the major domain was reading literacy, in PISA 2003, it was mathematical literacy, in PISA 2006, it was science and in PISA 2009, reading literacy was once again the main domain. Mathematics was the major domain for a second time in PISA 2012, and science will again be the major domain in 2015.

What is TALIS?

The OECD Teaching and Learning International Survey (TALIS) is a large-scale international survey that focuses on the working conditions of teachers and the learning environment in schools. TALIS, a collaboration among participating countries and economies, the OECD, an international research consortium, social partners and the European Commission, aims to provide valid, timely and comparable information to help countries review and define policies for developing a high-quality teaching workforce.

TALIS examines the ways in which teachers' work is recognised, appraised and rewarded. It also assesses teachers' participation in professional development activities. The study provides insights into teachers' beliefs about and attitudes

towards teaching, the pedagogical practices that they adopt and the factors related to teachers' sense of self-efficacy and job satisfaction.

TALIS also examines the roles of school leaders and the support they provide to teachers. The first cycle of TALIS was conducted in 2008 and surveyed teachers and school leaders of lower secondary education in 24 countries. In 2013, 34 countries and economies participated in TALIS. The OECD is now planning TALIS 2018.

1.5 Overview of the chapters

The chapters in this book contain secondary analyses of data from the Nordic countries. The comparison between these countries is interesting because of their cultural similarities and the differences in the organisation of their educational systems, which influence the levels of achievement. Each chapter of the book is briefly introduced below, and interesting results from are emphasised. Survey studies are based on how individuals respond and perform on certain instruments, while the curriculum analysis is conducted at the institution level. The results are seen as evidence of how the educational system or parts of the system perform.

The chapter by G. Nortvedt, A. Pettersen, A. Pettersson and S. Sollerman analyses the relevance of the 2012 PISA results to mathematics education in Sweden and Norway. Performing well in mathematics is seen as one of the most important factors for developing students' motivation for further studies in mathematics, science and engineering. Understanding of principles in different areas of mathematics seems to be gaining more importance. This is because mathematics has been integrated into new methods and many knowledge domains in almost all sectors of society. Although such technical-economic arguments are used to give priority to mathematics in the national curriculum, the OECD downplays this line of argument, probably because the OECD PISA study claims to measure problem-solving capacities more generally and not only the mathematical content.

Nortvedt *et al.* analyse the structure and content of the PISA framework for mathematics and compare it with the curriculum in the two countries. The authors give a detailed account of different parts of the curriculum and items used in the PISA study. The results clearly show a high degree of overlap between the PISA items used in the study and the content of the intended mathematics curricula in Norway and Sweden. From a policy point of view, this is important because public discussion sometimes questions the relevance of PISA studies in providing valuable insights into the relation between curriculum and mathematical performance at the national level.

R. F. Olafsson's chapter is based on the TALIS study. The aims of this chapter are to identify the clusters of leadership styles among principals in the Nordic countries, how leadership styles are connected to teachers' attitudes and behaviours and how the clusters are related to student achievement. The study analyses data at the Nordic level and suggests that other units of analysis are more interesting than finding the national average. The results show clear differences between the clusters, which are (1) collaborative, instructional and administrative leadership; (2) reactive "under siege" leadership; (3) moderate instructional leadership, with an emphasis on mentoring and little reaction to teacher appraisals; and (4) reactive leadership, with financial incentives and consequences of teacher appraisals.

The first cluster is most strongly associated with high student achievement. This result is based on a literature review, TALIS data, and the characteristics of the principals in schools that promote high student achievement. The variation between the clusters and how they relate to student achievement poses a number of questions. One is whether we can expect school principals to perform well on all the indicators measured in TALIS. It could be that groups of leaders at schools are a more adequate unit for measuring leadership effects on high-school students' performance. At the policy level, these results can be integrated into educational programmes for principals to reflect on and to transform their practices.

The chapter by M. Taajamo addresses the association between teachers' professional development, job satisfaction and self-efficacy in the

Nordic countries. The analysis of teachers' work in the classroom is described as complex, and teachers need to continuously learn and develop as adaptive experts. A basic assumption in this chapter is that when differences in the student population increase, more varied repertoires of teaching methods are needed. Teacher education can be seen as the foundation for the profession. However, through participation in the social practices in schools, adaptive expertise can be enhanced with a focus on students' learning.

The chapter presents two interesting results: (1) self-efficacy reflects teachers' perception of their goal attainment in working with students. All the Nordic countries scored above the mid-point range in the TALIS study, (2) high self-efficacy and job satisfaction seems to be strongly related to mentoring activities. However, it is interesting and contradictory that Finnish teachers are the least involved in mentoring and yet have scores higher than other Nordic countries on student achievement tests. This could imply that the teacher education programme in Finland provides a better foundation for teaching and learning since they strongly emphasize knowledge about students' learning. Whether this foundation is sufficient for further development is an open question. A general finding of this chapter is that induction to the teaching profession, in-services and continuous training are fragmented in all countries (with some variations).

The chapter by J. Caspersen based on TALIS data asks: Can feedback from colleagues and school leaders improve teachers' self-efficacy in Nordic classrooms? This chapter addresses questions related to those analyzed and discussed in the chapters by Olafsson and Taajamo. However, this chapter builds on other data and analyzes the phenomena of appraisals and feedback in relation to self-efficacy in a different manner. Results show that feedback from colleagues and school leaders varies between schools and between countries and is dependent on what teachers consider important. Not surprisingly, teachers with many years of experience appreciate different types of feedback more than newcomers to the teaching profession. While new teachers seem to appreciate deepening their own teaching practices, more experienced teachers appreciate moving horizontally, which means coordination and collaboration with peer teachers, with a less intense focus on their own teaching. From the

perspective of introducing changes to practices, this seems like a dilemma. If experienced teachers do not produce the expected results, how can change be achieved? Another finding is related to how school leaders choose to talk about students' test scores and the teachers' experienced self-efficacy. It is not the test scores itself that create self-efficacy; feedback becomes a tool for talking about work in the classroom.

A broader result discussed in this chapter is that professional collaboration seems to be positively related to self-efficacy, and this is valid for expert and novice teachers. One can argue that it is through collaboration that a teacher's standards, methods and ways of working become transparent. Through such practices, a teacher develops into a professional who integrates the collective knowledge of the profession and uses varied methods when working with students. If a teacher's work is too individualized, professional development will be hampered.

The chapter by P. Nyström focuses on high-performing students in mathematics based on a comparison of results from PISA 2003 and 2012. In the chapter, the aim is to find the characteristics of high-performing students and whether the characteristics had changed during the time period analyzed here. In addition to factors related to the students' socio-economic background, the literature about high-performing students often emphasizes that students are self-confident and have a high degree of mastery and ease with learning in mathematics.

The study confirms findings from extant literature in that high performance in mathematics is strongly related to students' cultural and educational background and socio-economic status. The survey data contain information on achievement scores and self-reported scores on motivation, self-confidence and self-efficacy. Simply put, high performance means to be a part of positive learning cycles, in which a higher degree of mastery in the domain is expected. The findings show that high-performing students think that they spend more time on mathematics in class, they are more positive towards their teachers, they have more advanced cognitive strategies and are better able to employ their existing knowledge when working with new problems than students who are median performers.

The chapter by J. Braeken contributes to public debate among researchers, policy makers and the media about studies such as PISA. A

starting point for this chapter is the occasional heated media debate about the PISA study. The media is often interested in and attempts to create a controversy by asking experts for opposing views, and the researchers themselves lose control of what is debated. The chapter is an important contribution to the methodological context of large-scale comparative assessment studies and presents a nuanced discussion on how we can interpret large-scale educational assessment studies. This contribution opens the black box and looks at the strengths and weaknesses of studies such as PISA as a prototypical example.

This chapter describes the problems associated with PISA study's design, data and statistical analysis, which extend beyond everyday knowledge and are rooted in the technical modelling and sampling of data. For instance, PISA results are valid for countries but not for schools. The statistical model used to construct the results is technically complex, almost incomprehensible to individuals lacking the expertise. However, the reliability and validity of the PISA results is based on this technical model, while the communication of the results becomes open to interpretation, irrespective of the technical statistical instruments used. This chapter explains the strengths of large-scale assessment studies as well as their limitations. The implication is that we must consider what different types of research design can produce regarding evidence of school practices. We need different types of studies to develop robust policy recommendations.

1.6 Discussion

Most studies based on large-scale surveys highlight the variations that exist between countries, between schools and school districts and between teachers and students. As R. Olafsson *touches* upon, the unit of analysis serves as a filter to address the relevant research questions. The unit of analysis is a technical notion used in many types of research. On the basis of previous research, theory and assumptions a part of a phenomenon is selected and certain methods are used to create the needed boundaries. This type of adequate reduction is part of conducting survey studies and other types of research.

The findings about the leadership styles of school principals call for further research and critical reflection on the education of principals. The other main findings related to school leadership and professional development can be synthesized towards the importance of creating cultures in which student learning must be more visible and transparent, in order for practices to change. This can be done by making teachers' work transparent and encouraging discussion and critique of professional values, norms and standards by other teaching professionals. Through such processes, novices and experts can improve their work.

In-service training and continuous training of teachers are often seen as the most important factors for sparking change and improvement. Taajamos's study asks for a clear strategic direction for professional development and learning. Without such a strategy, a country's capacity to change and improve school practices may be hampered. This finding seems to be relevant for all the Nordic countries.

At a broad level, most of the findings emphasise the interdependence between different factors. Local autonomy for schools, principals or teachers is often used as a concept to describe how schools are organised in relation to government and local authorities. However, the argument against local autonomy is that it hides more than it clarifies. When using interdependency as an analytic concept, we can see how teachers' performances are dependent on a number of factors. Novice teachers depend on involvement from more experienced colleagues and principals, while experienced teachers work with another set of dependencies. Thus, the question that educators should ask is which set of interdependencies creates the best conditions for improving students' in-depth learning. Looking for autonomy does not give us analytic lenses for how to improve our schools.

As mentioned in the introduction, this book is based on survey studies (except for one chapter). They provide valuable and needed knowledge about especially the "what" question. These pertain, for example, to student performance in a subject such as mathematics or how different actors perceive themselves in their professional work. However, other research designs provide a better view of how we can improve schools' practices. We need more detailed observations of daily

practices or targeted interventions based on well-researched phenomena. Such measures will help us understand and explain how and why students or participants choose to participate the way that they do.

Some of the chapters shed light on teachers' knowledge foundation. They argue that the programmes for professional development and learning are hampered by a weak strategic direction, related to the quantity of time used and the quality of the knowledge developed. The main goal of professional development should be students' learning. The educational sector in the Nordic countries seems to struggle to develop and establish a knowledge system and mechanism for the use of scientific knowledge relevant to the actors in the sector.

From a policy perspective, the variations (e.g. student achievements, how educational practices are carried out, what school principals prioritise etc.) described in the chapters give rise to dilemmas. Showing variations is an important step towards bettering policies and results. Given this context, should researchers search for interventions that can support changes in practice more generally or develop specific measures for the populations that need improvement, or both? According to the chapters in this book, the answers are not very obvious.

Some of the results in the chapters are counterintuitive, while others confirm findings from previous studies. This is why new policies for educational systems must consider normative expectations and new empirical evidence about school practices, both of which are challenging to understand but required for initiating systematic and effective change. When initiating systemic change, one should always look for multiple sources of evidence (NOU, 2015:8).

Lastly, what about the Nordic model? My interpretation is that Nordic citizens appreciate and trust the public school system as one of the most important institutions in the comprehensive welfare state system. The institutions and the actors that produce educational services and its supporting structures deliver knowledge, skills and competences for continuous development of its citizens and for the society at large. If the variation within each Nordic country or between them increases radically, the common features that are emphasised in this introduction and the trust relations they rely upon become at stake.

1.7 References

- Dølvik, J.E., Fløtten, T., Hippe, J.M. & Jordfald, B. (2015). *The Nordic model towards 2030. A new chapter?* NordMod2030. Final report. Fafo-report 2015:07
- Gough, D., Oliver, S. & Thomas, J. (2012). *An introduction to systematic reviews*. London: Sage Publications.
- Kavli, A. B. & Thorsen, H. (2014). *Northern lights on TIMSS and PIRLS 2011. Differences and similarities in the Nordic countries*. The Norwegian Directorate for Education, Oslo. Nordic Council of Ministers. <http://dx.doi.org/10.6027/TN2014-528>
- OECD (2014a). *PISA 2012. Results in Focus. What 15-year-olds know and what they can do with what they know*. Paris: OECD Publishing.
- OECD (2014b). *PISA 2012 Results: Students and money. Financial literacy skills for the 21st century*. Paris: OECD Publishing.
- NOU (2014:7). *Elevenes læring i fremtidens skole. Et kunnskapsgrunnlag*. Oslo: Kunnskapsdepartementet.
- NOU (2015:8). *The future of schooling. Renewal of subjects and competences*. Oslo. Ministry of Education and Research.
- The Economist (2013). The Nordic countries. The next supermodel. *The Economist*. 2 February.

2. Is PISA 2012 relevant to mathematics education in Norway and Sweden?

By Guri A. Nortvedt,¹ Andreas Pettersen¹, Astrid Pettersson² and Samuel Sollerman²

2.1 Summary

Our aim is to describe and discuss the relevance of PISA 2012 to mathematics education in Norway and Sweden. In both countries, PISA is used to provide trend data on educational progress and to inform policy making. It is therefore imperative to gain better insights into how and to what degree PISA is relevant.

We first compare the structure of the PISA mathematics framework and the national mathematics curricula documents for Norway and Sweden. All the documents contain goals that explain the rationale underlying mathematics education and define the mathematical activity and content to be learned. Strong similarities are found in the stated purpose of mathematics education, which address the needs of both the individual and society and focus on the mathematical knowledge and abilities needed to be a constructive, engaged and reflective citizen. Surprisingly, the PISA framework downplays the technical-economical reasons for mathematics education, unlike the two curricula documents.

¹ University of Oslo, Norway.

² Stockholm University, Sweden.

We find that mathematical activity is connected to mathematical modeling and problem solving processes in all documents. PISA does not aim to provide participating countries with the mathematics content to be taught. Even so, considerable alignment between the outlined PISA mathematical content and curriculum content areas is evident. The strong overlap between the PISA mathematics framework and the two curricula indicates that PISA 2012 is relevant to mathematics education in the two countries. This observation is supported by our analysis of PISA assessment items. We used the mathematics content strands in the national curricula as categories to evaluate the PISA assessment items and found that all items assess content belonging to the national curricula. This analysis, however, does not indicate to what extent the mathematical content covered in the national curricula is tested by PISA items. Many aspects of both the PISA assessment and the assessment items were not part of our investigation, such as the test and item format, the test situation and the language used in the mathematics items. These aspects are also important to consider when discussing the relevance of PISA.

Nonetheless, mindful of the limitations of this study, we conclude that PISA is relevant to mathematics education in Norway and Sweden.

2.2 Introduction

The goals of mathematics education are similar in many countries, with a strong focus on mathematical literacy (Burkhardt, 2014; Niss & Jablonka, 2014). Burkhardt (2014, p. 14) claims:

Around the world people seem to have much the same goals for the outcomes of a mathematics education. Students should emerge with a reliable command of a wide range of mathematical skills, a deep understanding of the concepts that underlie them, and an ability to use them, flexibly and effectively, to tackle problems that arise – within mathematics and in life and work beyond the classroom.

A common view is that compulsory education should provide students with the knowledge and skills they need, both for further education and life outside the educational system. Although the above quote does not

specifically address why the society needs mathematically competent citizens, their role has long been recognised (Clements, 2013; Niss, 1996). The need for mathematically competent citizens stems from the crucial role mathematics plays in the development of society both from a technological and sociological perspective (Niss, 1994).

According to Dindyal (2014), although there has been a long tradition of comparative studies aimed at determining how mathematics is taught elsewhere, the use of international comparative studies on mathematical achievement has increased significantly in the last few decades. The mathematical competence and general educational level of students leaving compulsory education concerns society to a large extent, and the recent trend is that international comparative studies are used to monitor the effectiveness of educational systems. For instance, the Nordic countries – Denmark, Finland, Iceland, Norway and Sweden – have participated in all cycles of the Program for International Student Assessment study (PISA) (OECD, 2013a). Both in Norway and Sweden, PISA is used to inform policy makers. In Norway, insights from international studies are frequently cited in government white papers (Elstad, Nortvedt, & Turmo, 2009) to provide information to the national educational system. Indeed, national reports from the Norwegian directorate for education and training (NDET) identify international studies as provider of trend data on educational progress to the national quality assessment system (NKVS) (Allerup, Kovac, Kvåle, Langfeldt, & Skov, 2009; Elstad *et al.*, 2009) (see, for instance, NDET, 2014). As in Norway, international studies also contribute to the Swedish quality assessment system and provide trend data on educational progress for Skolverket (the Swedish National Agency for Education, Skolverket, 2013). International studies, such as PISA, were included in the material used to develop the Swedish curriculum of 2011 (Skolverket, 2011a). Political interest in the first few PISA surveys was not particularly high, but after Sweden's average performance dropped significantly in 2012, governmental interest in the survey grew strongly (OECD, 2015). When the results of the 2012 PISA study were published, the government decided that a group of experts from the OECD should undertake an in-depth analysis of the results to provide advice on how to change and improve the educational system (OECD, 2015).

When outcomes of international studies are used by policy makers to inform educational policy, as in Norway and Sweden, it is vital to discuss their relevance to the intended policy (Dindyal, 2014; Leung, 2014). PISA measures students' preparedness for life after compulsory school, with a focus on students as active problem solvers engaging in the core processes of mathematical modeling (OECD, 2013a, 2013b). Critics claim that since the PISA framework is not necessarily aligned with every participating country's curriculum, the PISA survey cannot provide relevant information about the national educational system (e.g. Sjøberg, 2014). However, Leung (2014) argues that the competence students use to answer the PISA assessments are mainly acquired in school. In addition, problem solving is recognised as an important part of the national curriculum in most countries, and modeling is viewed as an "activity at the core of the *utility* of mathematics" (Burkhardt, 2014, p. 24). To add to the discussion about the relevance of PISA, we aim to investigate the following research questions:

- How is the definition of mathematical literacy in the PISA 2012 mathematics framework aligned with the goal definitions of the national curricula in Norway and Sweden?
- To what extent is the mathematical content assessed by the PISA 2012 mathematics assessment items covered by the mathematical content contained in the national curricula of Norway and Sweden?

2.3 Methodology

We aim to analyse and compare the PISA 2012 mathematics framework and the Norwegian and Swedish mathematics curricula documents. We will discuss the degree of alignment between the PISA 2012 mathematics framework and the national curricula in Norway and Sweden, and subsequently, how relevant PISA 2012 is to mathematics education in the two countries. As the assessment items operationalise the assessment framework, we also categorise the PISA mathematics items according to

the mathematical content strands in the two national curricula – to discuss the relevance of what is measured by PISA.

The following three documents have been analysed:

- The Norwegian curriculum for the common core subject of mathematics (NDET, 2015a), including the framework for basic skills (NDET, 2012).
- The Swedish Lgr 11 curriculum for the compulsory school, preschool class and the recreation centre 2011 (Skolverket, 2011b).
- The OECD PISA 2012 assessment and analytical framework for mathematics, reading, science, problem solving and financial literacy (OECD, 2013b).

In all three documents, only the pages describing the mathematics curriculum were selected for analysis. PISA aims at assessing the mathematical competence of 15-year-old students (OECD, 2013b). Consequently, only the text describing the goals for students aged 13–15 (grades 8–10 in Norway and 7–9 in Sweden) was selected from curricula documents, which describe learning goals for students at various stages of compulsory education.

To allow comparison of the content of the documents, the first analysis considered the structure of the selected texts. The content was categorised into three levels according to Niss (1996): goal definitions at the *end*, *aim* and *objective* level. Identified sections were analysed and compared pairwise at each of the three levels, to investigate

1. if a similar purpose of mathematics education could be identified (end level)
2. how different aspects of mathematical competence were described in the documents (aim level)
3. if corresponding mathematical content was included in all the three documents (objective level).

Content analysis (Robson, 2002), applying categories that are developed by drawing on key texts, was used to analyse and compare content at each of the three levels (categories are described in the beginning of each of the three sections). For example, the categories used to analyse and compare the PISA framework and the national curricula at the end level (purpose of mathematics education) were the *fundamental reasons* for mathematics education developed from Niss (1996). Although published 20 years ago, this handbook chapter is still frequently quoted and provides an analytical lens to investigate the relevance of PISA to mathematics education in Norway and Sweden.

Items in the PISA 2012 paper-based mathematics assessment were categorised on the basis of the content strands in Norwegian and Swedish curricula. This was done to investigate how much of the mathematical content of PISA assessment items was covered in the national curricula. For each country, a national team of authors and a national external rater performed the categorisation. Differences were discussed and a common category was agreed upon.

2.4 Curriculum structures and goal definitions

There are many different uses and understandings of the word *curriculum* across the world, and in searching for a definition of curriculum, Cai and Howson (2013) noted that “it is almost impossible to give a universally acceptable definition” (p. 951). In the US, it might refer to a textbook series and in the UK to a set of classroom experiences (Burkhardt, 2014). In this article we understand *curriculum* as the aims, content and goals described in official documents regulating mathematics education on a national level. This is referred to as *læreplan* in Norwegian or *läroplan* (including *kursplan*) in Swedish (NDET, 2015b; Skolverket, 2011b).

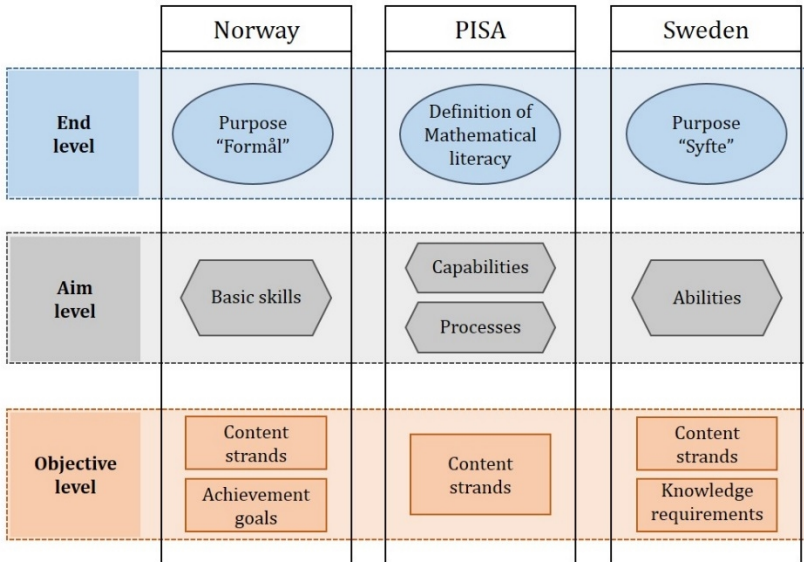
In this section, we will look at the structure of the mathematics curricula only, at the levels that Niss (1996) refers to as the goals of mathematics education. Educational goals might be described at several levels. At one extreme we find end-level goals, stating the overall goal of teaching mathematics in schools. End level goals are often vague and difficult

to assess. At the other extreme we find objective level goals, stating specific content or strategies to be learned (Niss, 1996). For this analysis, we have used three levels. In addition to the two extremes, the end and objective level, we have defined an intermediate level, the aim level that comprises more general mathematical competences.

Using the three levels of goal definitions as categories and drawing on Niss' (1996) work, content in the curricula documents can be categorised into each of the levels (see Figure 1). As in Niss (1996), parts of the documents pertaining to the end level comprise overall goals for mathematics education. Some might describe these as the final outcome of mathematics education in compulsory education. The content categorised as belonging to the aim level typically comprises goals describing general mathematical competences that do not belong to specific mathematical content, such as communicating mathematically (Niss & Højgaard, 2011). The third level, the objective level, covers goals describing the mathematical content to be learned. This level comprises content strands or topic lists that give information about, for instance, what kind of theorems, concepts and procedures students should acquire. Content strands typically found in mathematics curricula include, for instance, algebra and geometry.

Figure 1 displays the outcome of this first analysis: the structure of the Norwegian and Swedish curricula and at what levels goals are defined. In addition, what might be termed the goal definitions in the PISA mathematics framework are included in the goal structure to allow comparison with each of the two national curricula.

Figure 1: Structures of Norwegian and Swedish mathematics curricula and PISA mathematics assessment framework



The Norwegian mathematics curriculum (LK06) consists of a purpose, i.e. an overarching aim of teaching mathematics located at the end level. Descriptions of five basic skills (oral, writing, reading, numeracy and ICT) and how they develop during the teaching and learning of mathematics are goals allocated to the aim level since these describe the competences and activities that are not tied to specific mathematical content. For this analysis, we use the framework for basic skills in numeracy (NDET, 2012) as this framework underlies the mathematics curricula document (Ministry of Education, 2010). Finally, aims at the objective level consist of the content strands that describe the main mathematical domains the students should encounter and topic lists that comprise detailed descriptions (achievement goals) of what students should be able to do at different levels within compulsory school (NDET, 2015a).

The Swedish curriculum includes syllabuses for each subject. The mathematics syllabus contains purposes and overarching aims of teaching mathematics (Skolverket, 2011b) allocated to the end level. In addition, this curriculum document includes a description of five abilities the students should develop within the compulsory mathematics education. These abilities are defined across mathematical content, e.g. communicate mathematically, and describe general mathematical competences. Consequently, these abilities are allocated to the aim level. Finally, content strands describe the mathematical content that students should encounter through classroom activities and knowledge requirements for different grades. These strands and knowledge requirements provide fairly concrete and well defined content to be learned, and as such are allocated to the objective level.

Comparing the structure of the national curricula to the PISA 2012 mathematics framework, both curricula have goals formulated at the end level which provide a contextual description of mathematics, in addition to describing the outcome of the teaching to the society and to the individual. The definition of mathematical literacy in the PISA assessment framework (OECD, 2013b) can also be placed at the end level as it describes the importance of mathematics to the individual as a participant in the society. We will address this level later in this chapter and discuss to what degree the three documents comprise similar purposes for mathematics education.

At the aim level, the two curricula define goals describing general mathematical competences; mathematical problem solving and communicating mathematically are, for instance, included both in the Norwegian basic skills and in the Swedish abilities. The PISA framework covers general mathematical competence in the form of processes and capabilities that resemble those in the national curricula documents. We will discuss these competences in the section on goals at the aim level.

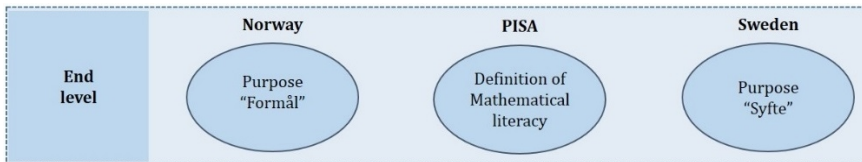
Both curricula define goals at the objective level. Goals at this level provide a more detailed and specific description of the mathematical content that should be learned. The PISA framework includes a description of concrete mathematical content, which comprises four content categories (OECD, 2013b). For each category, the mathematical content students should be able to engage in is described and some examples are provided.

This content is allocated to the objective level. The alignment between the PISA content strands and the national curricula content strands will be discussed in the section on objective level goals.

2.5 Goals at the end level: Fundamental reasons for teaching mathematics

According to Niss, “in many democratic countries, today, it [mathematics] is further intended to empower pupils to enter society as competent, independent, active and critical individuals and citizens” (Niss, 1996, p. 12). This intention is still the main goal of mathematics curriculum development in many countries (Burkhardt, 2014; Clements, 2013). Such an articulation of the overall national purpose for teaching mathematics can be seen as a national policy, giving direction to compulsory mathematics education. At this level, the needs of both society and the individual are addressed (Niss, 1994).

Figure 2: Goals at the end level of Norwegian and Swedish mathematics curricula and PISA mathematics assessment framework



According to Niss (1996), three fundamental reasons are often cited to justify mathematics education. These reasons are not the same as the goals of mathematics education, however, “the demarcation line between the two is not always so easily drawn in practice” (Niss, 1996, p. 15). This is evident at the end level, where global goals of mathematics education are stated. These goals are directed towards fulfilling society’s need for mathematically educated persons to fill the many roles in society, as well as the individual’s need for mathematical competence (Niss, 1994). Consequently, in this article, at the end level we do not distinguish between justifications

and goals and apply the three fundamental reasons for mathematics education based on Niss (1996, all quotations below from p. 13) to categorise the national curricula text at the end level (Formål and Syfte) and the definition of mathematical literacy in the PISA framework:

- *Technical-economical*: “contributing to the ‘technological and socio-economic development’ of society at large, either as such or in competition with other societies/countries”.
- *Societal*: “contributing to ‘society’s political, ideological and cultural maintenance and development’, again either as such or in competition with other societies/countries”.
- *Individual*: “providing ‘individuals with prerequisites which might help them to cope with life’ in the various spheres in which they live: education or occupation; private life; social life; life as a citizen”.

In the PISA 2012 framework, mathematical literacy (see below) is mainly tied to citizenship and activity, to modeling and problem solving and to the use of different aspects of one’s mathematical competence. In addition, the role of mathematics is included. However, no specific reference to the technical-economical perspective is mentioned:

Mathematical literacy is an individual’s capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals in recognising the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.

(OECD, 2013b, p. 25)

The attention given to the role of mathematics in society, problem solving and modeling in the Norwegian curriculum document resembles the PISA definition to a large extent. The document specifically addresses the reasons for mathematics education that belong to the societal and the individual categories, (e.g. “[t]he subject of Mathematics contributes to de-

veloping the mathematical competence needed by society and each individual”) (NDET, 2015a, p. 2). Reasons belonging to the technical-economical category are less visible, although some reference is made (e.g. “active democracy requires citizens who are able to study, understand and critically assess quantitative information, statistical analyses and economic prognoses” (p. 2) and “[t]he subject is part of many vital societal areas, including medicine, economy, technology, communication, energy management and construction” (p. 2)). This category is mainly paired with citizenship and development of society. It might be inferred that the Norwegian curriculum links the three fundamental reasons for teaching mathematics to each other.

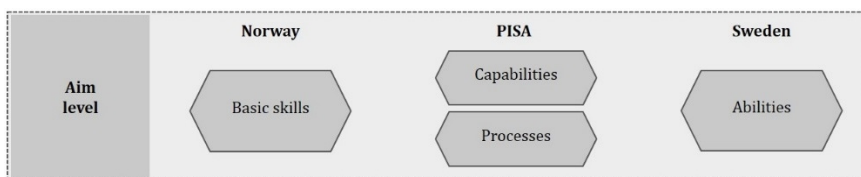
In the introduction to the Swedish curricula, mathematical activity is linked to the development of society and falls into both the technical-economical and societal categories. In addition the importance of mathematics to the individual is stressed, also in relation to society (e.g. “Mathematics is [...] closely linked to societal, social and technological development. Knowledge of mathematics gives people the preconditions to make informed decisions in the many choices faced in everyday life and increases opportunities to participate in decision-making processes in society”) (Skolverket, 2011b, p. 59). The text goes on to describe in detail what the individual should achieve from participating in mathematics education. This section is more specific and detailed than what is often observed at the end level.

To summarise, the PISA framework mainly addresses the fundamental reasons at the individual level by referring to the mathematical abilities and knowledge needed to be a constructive, engaged and reflective citizen – objectives that fall into the societal category. Similar reasons, belonging to the individual and societal categories, are present in both the Norwegian and Swedish curricula. Surprisingly, despite OECD’s mandate, fundamental reasons that fall into the technical-economical category are less prominent in the PISA framework than in the Swedish and Norwegian curricula.

2.6 Goals at the aim level: General mathematical competence

At the aim level, the Norwegian and Swedish curricula define and describe five basic skills and five abilities respectively. Goals at the aim level define more general mathematical competences.

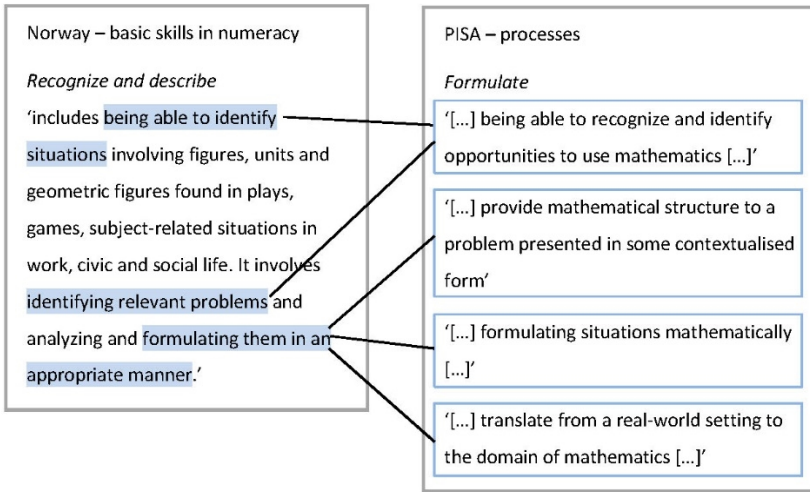
Figure 3: Goals at the aim level of Norwegian and Swedish mathematics curricula and PISA mathematics assessment framework



In the PISA framework (OECD, 2013b), the three processes of mathematical problem solving and modeling are described. This model is well known from the research literature (see, for instance, Lesh & Caylor, 2009; Lesh & Doerr, 2003; Lesh & Zawojewski, 2007) and describes the processes of modeling as formulating real-world problems and situations mathematically (*formulate*); employing mathematical concepts, facts, procedures and applying problem solving heuristics (*employ*); and interpreting, applying and evaluating mathematical outcomes (*interpret*). The many models include the same processes and the stages between them; however, in some models the *interpret* process is split into two processes with an intermediate stage. Because the PISA framework draws on this research (e.g. Niss, Blum, & Galbraith, 2007), the PISA processes described above are used as categories (names given in italics).

Drawing on the description of the PISA processes (categories), Figure 4 illustrates our categorisation process. The national documents are categorised by using descriptions of activities from the PISA processes. For instance, formulating comprises the activity of recognising situations and formulating the content of the situation mathematically. Such activities are also found in the Norwegian definitions of the basic skill numeracy.

Figure 4: Categorisation of the basic skill aspect Recognize and describe



In addition to these processes, the PISA framework describes mathematical capabilities that underpin both the processes and mathematical literacy in practice, for instance, to be able to reason and argue mathematically and to mathematise real-life situations (OECD, 2013b). These capabilities build on the work of Niss and Højgaard (2011); students need to activate one or more of the capabilities when engaging in the processes of solving problems. The PISA capabilities are used as categories.

Table 1: Categorisation of the Norwegian basic skill numeracy

Norway, basic skills in numeracy	PISA, processes
<p><i>Recognize and describe</i> “includes being able to identify situations involving figures, units and geometric figures found in plays, games, subject-related situations in work, civic and social life. It involves identifying relevant problems and analyzing and formulating them in an appropriate manner.”</p>	Formulate
<p><i>Apply and process</i> “involves being able to choose strategies for problem solving. It involves using appropriate units of measurement and levels of precision, carrying out calculations, retrieving information from tables and diagrams, drawing and describing geometric figures, processing and comparing information from different sources.”</p>	Employ
<p><i>Communicate</i> “means being able to express numerical processes and results in a variety of ways. Communicate also means being able to substantiate choices, communicate work processes and present results involving numbers.”</p>	Interpret, Employ
<p><i>Reflect and assess</i> “means interpreting results, evaluating validity and reflecting on effects. It involves using results as basis for a conclusion or an action.”</p>	Interpret

Note: Quotations from the Norwegian framework for basic skills (NDET, 2012, p. 14).

The four sub-categories defined within in the Norwegian basic skills in numeracy (NDET, 2012) shown on the left column of Table 1 are analysed using the PISA categories of formulate, apply and interpret. The outcome of our analysis is shown in Table 1 where the four sub-categories of the Norwegian basic skill numeracy and the three PISA processes are matched. Recognize and describe; apply and process; and reflect and assess are very much aligned with formulate, employ, and interpret, respectively. The fourth aspect, communicate, comprises elements from both the interpret and employ processes.

The capabilities described in the PISA framework, seemingly underpin the Norwegian basic skills much in the same way as the PISA processes. For instance, the description of the communication capability (OECD, 2013b, p. 30) includes “recognise and understand a problem situation” (resembling recognize and describe), “[r]eading, decoding and interpreting statements, questions, tasks or objects” (resembling apply and process) and “present the solution, and perhaps an explanation or

justification, to others” (resembling the Norwegian definition of communicate). The other capabilities relate to one or more of the Norwegian basic skills in numeracy, with one notable exception: the *using mathematical tools* capability, which is described as having knowledge about and being able to use mathematical tools (e.g. calculators, computer-based tools and measuring instruments), is not recognised in any of the basic skills. Rather, this capability is to some degree captured in the basic digital skills (NDET, 2012).

The Swedish abilities (Skolverket, 2011b), which describe mathematical activities students should be able to do, resemble both the three processes and the seven capabilities defined in the PISA framework. Table 2 shows the outcome of the categorisation of the Swedish abilities using the PISA processes and capabilities as categories.

Table 2: Categorisation of Swedish abilities

Sweden Abilities	PISA Capabilities – processes
<p><i>Problem solving ability</i> “Formulate and solve problems using mathematics and also assess selected strategies and methods”</p>	<p>Capabilities: <i>Communication, Mathematising, Devising strategies for solving problems</i> Process: <i>Formulating, Interpreting</i></p>
<p><i>Concept ability</i> “Use and analyse mathematical concepts and their interrelationships”</p>	<p>Capabilities: <i>Mathematising</i> Process: <i>Employing</i></p>
<p><i>Method ability</i> “Choice and use appropriate mathematical methods to perform calculations and solve routine tasks”</p>	<p>Capabilities: <i>Using symbolic, formal and technical language and operation, Devising strategies for solving problems</i> Process: <i>Employing</i></p>
<p><i>Reasoning ability</i> “Apply and follow mathematical reasoning”</p>	<p>Capabilities: <i>Reasoning and argument</i> Process: <i>Employing, Interpreting</i></p>
<p><i>Communication ability</i> “Use mathematical forms of expression to discuss, reason and give an account of questions, calculations and conclusions”</p>	<p>Capabilities: <i>Representation, Reasoning and argument, Using symbolic, formal and technical language and operation</i> Process: <i>Formulating, Employing, Interpreting</i></p>

Note: Quotations from the Swedish curriculum (Skolverket, 2011b, pp. 59–60).

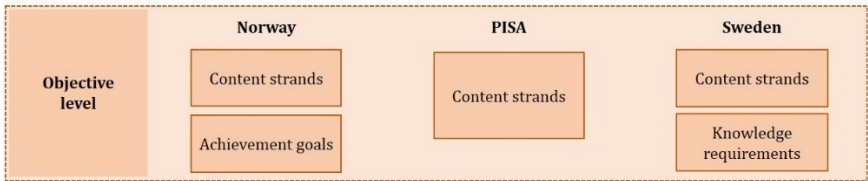
In the PISA framework the communication capability is directed more towards interpretation than accounting and has more similarities with the Swedish ability of problem solving than communication. The PISA mathematising capability is close to modeling and is also more likely to be associated with the Swedish problem solving ability. The capability reasoning and argument in PISA has a much stronger emphasis on arguments than the Swedish reasoning ability. Similarly, devising strategies for solving problems under PISA has a strong emphasis on heuristic strategies and is more similar to the Swedish problem solving ability than for instance the method ability. However, using symbolic, formal and technical language and operations in PISA involves in-depth communication in addition to mastery of different concepts and methods; therefore, it is related to several of the Swedish abilities: method ability, concept ability and communication ability. PISA's focus on using mathematical tools is covered only to a small degree in the Swedish abilities; it is somewhat captured by the method ability.

Problem solving and modeling processes are highlighted in both the Norwegian and Swedish curricula and in the PISA framework. Further, the capabilities described in the PISA framework resemble the abilities in the Swedish curriculum, while they are covered at a more general level, underpinning the basic skills, in the Norwegian curriculum. In conclusion, strong similarities are found between each curriculum document and the PISA framework at the aim level.

2.7 Goals at the objective level: Mathematical content

Mathematics curricula have traditionally been organised around mathematical topics and content strands to be covered (Niss & Højgaard, 2011). However, which strands and the number of strands differ between documents, as the mathematical content might be organised according to different principles.

Figure 5: Goals at the objective level of Norwegian and Swedish mathematics curricula and PISA mathematics assessment framework



Goals at the objective level in the form of identified mathematical content are more explicitly described in the Norwegian and Swedish curricula than in the PISA 2012 framework. Still, PISA contains four content strands that “characterise the range of mathematical content that is central to the discipline and illustrate the broad areas of content [...] *change and relationships, space and shape, quantity, and uncertainty and data*” (OECD, 2013b, p. 33). The mathematical content knowledge characterising each of these strands briefly describes what students should be able to do and what kind of mathematical activities each strand comprises.

The organisation of the objective level differs somewhat between the Norwegian and Swedish curricula. The Norwegian curriculum (grade 8–10) is structured into five content strands labelled “main subject areas” (NDET, 2015a, p. 2): *Numbers and algebra in practice, Geometry, Measurement, Statistics, probability and combinatorics* and *Functions*. Each strand is briefly described, outlining the mathematical content belonging to this strand (Ibid., pp. 3–4). In addition to each content strand, achievement goals, stating what students should be able to do, are formulated. For instance, the first achievement goal for Geometry in grades 8–10 is as follows: “investigate and describe the characteristics of two- and three-dimensional figures and use them for constructions and calculations” (NDET, 2015a, p. 9).

The Swedish curriculum (grade 7–9) contains six content strands labelled “core content” (Skolverket, 2011b, pp. 62–64): *Understanding and use of numbers, Algebra, Geometry, Probability and statistics, Relationships and change*, and *Problem solving*. The Swedish content strands differ from the Norwegian strands in that each Swedish strand is described using a list of descriptive statements, which state the content that the students should experience in their mathematics education (Skolverket, 2015, p. 12). For instance, one of the content strands for grade 7–9 reads “Geometrical objects and their relationships. Geometrical properties of these objects” (Skolverket, 2011b, p. 63).

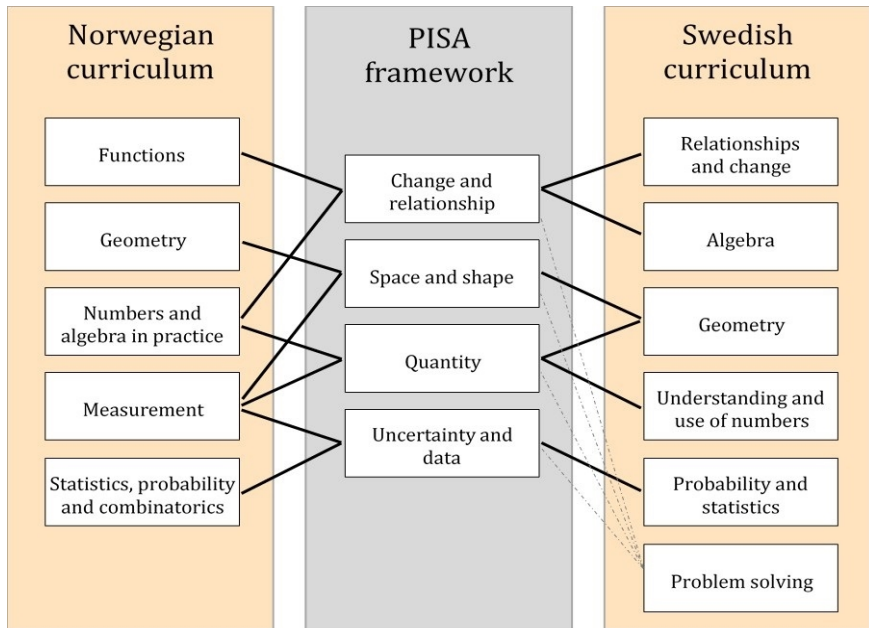
Table 3 exemplifies the multiple linkages between the documents from the analysis of the objective level goal definitions. For instance, the change and relationship strand in the PISA framework is described as “modelling the change and relationships with appropriate functions and equations, as well as creating, interpreting, and translating among symbolic and graphical representations of relationships” (OECD, 2013b, p. 33), including “the traditional mathematical content of functions and algebra” (Ibid., p. 34). In the Norwegian curriculum, this description fits the content strand of functions, which is defined as “change or development of an amount that depends on another” (NDET, 2015a, p. 3). Such transformations can be expressed in a number of ways (e.g. formulas, tables and graphs). Further, the change and relationship strand also comprises aspects of the content found in the numbers and algebra in practice strand under the Norwegian curriculum (i.e. calculations performed with numbers and other symbols). A similar comparison with the Swedish curriculum shows that the change and relationship strand in PISA resembles the Swedish strand of algebra, which comprises knowledge about the concept of variables, the use of variables in algebraic expressions, formulas and equations, and methods for solving equations (Skolverket, 2011b). Further, content related to functions and the use of functions to examine change and rate of change, covered under the relationship and change strand in Swedish curriculum, covers the content in the change and relationships strand in the PISA framework.

Table 3: Examples of content strand descriptors found in the three curricula documents

Norwegian curriculum	PISA framework	Swedish curriculum
<p><i>Functions</i> "A function unambiguously describes change or development of an amount that depends on another" "Analysis of functions [which] may be expressed in a number of ways, for example using formulas, tables and graphs" <i>Numbers and algebra in practice</i> "Algebra in school generalises calculation with numbers by representing numbers with letters or other symbols." "describe and analyse patterns and relationships."</p>	<p><i>Change and relationship</i> "modelling the change and relationships with appropriate functions and equations" "creating, interpreting, and translating among symbolic and graphical representations of relationships" "functions and algebra [...] are central in describing, modelling, and interpreting change phenomena"</p>	<p><i>Algebra</i> "Meaning of the concept of variable and its use in algebraic expressions, formulae and equations." <i>Relationships and change</i> "How functions can be used to examine change, rate of change and other relationships."</p>

Figure 6 shows the overall relationships between the content strands in the PISA framework and the national curricula emerging from the objective level analysis. Both curricula are compared to the PISA framework, although not to each other.

Figure 6: Relationships between the PISA content categories and the Norwegian and Swedish curriculum content strands



It should be noted that Figure 6 displays the main trends. Clearly, there is no one-to-one correspondence between the PISA content strand categories and the strands in the Norwegian and Swedish curricula. Still, as shown in Figure 6, the content categories are broadly similar in terms of subject areas and knowledge areas.

2.8 Relevance of the PISA mathematics assessment items to mathematics education in Norway and Sweden

The analyses in the previous sections show that the national curricula and the PISA 2012 framework are similar, with regard to the fundamental reasons for mathematics education and the more specific mathematical competences that students should achieve through their mathematics

education. In this section, the relevance question is addressed through an analysis of the PISA mathematics assessment items that operationalise the PISA framework.

By categorising the PISA mathematics items into the content strands in the Norwegian and Swedish curricula, we examine how the mathematical content comprised in the PISA items is covered by the national curricula. All the items in the regular paper based mathematics assessment in PISA 2012 were categorised into one content strand only for each curriculum. If the mathematical content of an item fitted multiple categories, the item was categorised into what was regarded as the main content strand. As PISA items rarely assess content from a single strand, content strands that often represent minor content areas, could appear underrepresented. For instance, algebraic language is often used when solving geometry problems. Such items were categorised as geometry.

The results of the categorisation are displayed in Tables 4 and 5, for Norway and Sweden respectively. As seen in the previous section, the organisation of mathematics content into content strands from the Norwegian and Swedish curricula produces quite different patterns. For instance, in Norway number and algebra is thought of as one single strand, while in Sweden these two topics are placed in separate strands. A similar observation can be made about the content strands of geometry and measurement in Norway and geometry in Sweden. Consequently, the categorisation of PISA items into the content strands will not be the same for the two curricula.

Table 4: Distribution of PISA 2012 mathematics items across Norwegian mathematics curriculum content strands

	Numbers and algebra in practice	Geometry	Measurement	Statistics, probability and combinatorics	Functions
PISA mathematics assessment items	33%	15%	9%	34%	8%

Table 5: Distribution of PISA 2012 mathematics items across Swedish mathematics curriculum content strands

	Understanding and use of numbers	Algebra	Geometry	Probability and statistics	Relationships and change	Problem solving
PISA mathematics assessment items	24%	5%	25%	29%	18%	0%

A main finding of the analysis is that all PISA items fit at least one content strand category in each curriculum document, indicating that the PISA items address mathematical content included in both the Norwegian and Swedish curricula.

Table 4 shows that for the Norwegian curriculum, the PISA items are distributed across all five content strands. As we can see from Table 5, the PISA items are spread across five of the six Swedish content strands. No PISA items have been categorized to have problem solving as the main content, although many of the items contain problem solving as a minor content area. This is because problem solving is described as both an ability (“Formulate and solve problems using mathematics and also assess selected strategies and methods”) and a content strand (“for example, simple mathematical models and how they can be used in different situations”) in the Swedish curriculum.

By now, we can conclude that PISA is relevant to mathematics education in Norway and Sweden. In the next section, we take a closer look at a sample of released items from the PISA 2012 mathematics assessment and discuss these through the lens of the PISA framework and the national curricula, taking into consideration the three levels of goal definitions. Subsequently, we will discuss the purpose of the item; the mathematical process and capabilities required for solving the item; and the content strand it fits.

The PISA mathematics assessment is designed to cover all content areas and processes to ensure a balanced assessment (OECD, 2013b). Each item is categorised by an expert mathematics group to indicate the content area and process that it mainly assesses. Items are organised in units. Each unit comprises a stimulus text, which provides a context, and

one or more questions (items) that are linked to this context (OECD, 2013b). Often the stimulus contains essential information; however, sometimes, the stimulus simply provides a relevant setting.

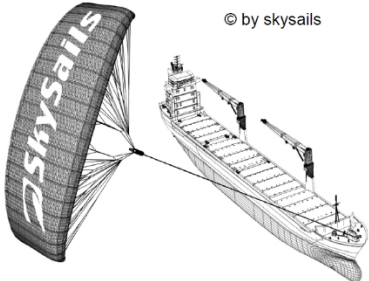
Figure 7 displays the stimulus for the PISA unit “Sailing ships” (OECD, 2013c, p. 12). This stimulus text provides information on the situation in international goods transport. However, this specific information is not necessary for solving the items that belong to the unit (see Figures 8 and 9). Rather, the stimulus refers to the technical-economical situation in the world as well as the societal, addressing the need to reduce both costs and impact on the environment. By relating the mathematical situation to the real world, a rationale is provided. The rationale in this stimulus is in accordance with the fundamental reasons for teaching mathematics stated in the Norwegian and Swedish curricula.

Figure 7: Stimulus PISA 2012 mathematics assessment unit “Sailing ships”

SAILING SHIPS

Ninety-five percent of world trade is moved by sea, by roughly 50 000 tankers, bulk carriers and container ships. Most of these ships use diesel fuel.

Engineers are planning to develop wind power support for ships. Their proposal is to attach kite sails to ships and use the wind's power to help reduce diesel consumption and the fuel's impact on the environment.



© by skysails

Source: OECD, 2013c, p. 12.

Figure 8 shows the first question developed for the above stimulus (Sailing ships). The short introduction text, expanding on the stimulus, provides students with both necessary and redundant information. However, the non-essential information provides more detail on the context and as such might make it easier for the student to relate to the item. It might be argued that the additional information provides a more concrete context to the fundamental reasons addressed in the stimulus.

At the aim level, according to the PISA categorisation, this item assesses students' ability to employ mathematical procedures (OECD,

2013c) as they have to calculate the correct speed by employing a suitable procedure. In the Norwegian curriculum, this corresponds to the basic numeracy skill, labelled apply and process. In the Swedish curriculum, this activity would mainly be linked to the method ability, and to some extent, the concept ability. This item mainly assesses the ability to calculate a percentage within a real life situation. At the objective level, the concept of percentages belongs to different content strands within the two national curricula. While it is placed within the relationships and change strand in the Swedish curricula, it belongs to the number and algebra strand in the Norwegian curriculum.

Figure 8: Question 1, PISA 2012 mathematics assessment item unit “Sailing ships”

Question 1: SAILING SHIPS

PM923Q01

One advantage of using a kite sail is that it flies at a height of 150 m. There, the wind speed is approximately 25% higher than down on the deck of the ship.

At what approximate speed does the wind blow into a kite sail when a wind speed of 24 km/h is measured on the deck of the ship?

- A 6 km/h
- B 18 km/h
- C 25 km/h
- D 30 km/h
- E 49 km/h

Source: OECD, 2013c, p. 12.

Figure 9 displays the third question developed for the stimulus situation of “Sailing ships”. In this item, the context is further extended and thus more concrete details about the described real life situation are added. At the aim level, this item can be assigned to the employ process in the PISA framework and similarly to the apply and process category in the Norwegian basic skill framework, as students need to apply the Pythagoras theorem to solve the item. Items that fall into the employ category in the PISA framework can be tied to either the method ability, in case of stereotypical items, or the problem solving ability, in case of novel or non-stereotypical problems, in the Swedish curriculum. Question 3 asks students to employ a mathematical theorem in a novel situation, and consequently has been allocated to the problem solving ability.

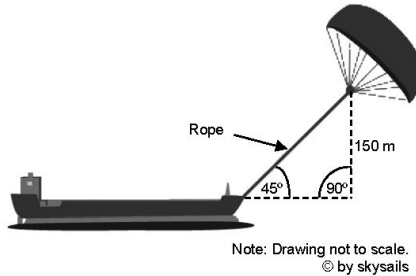
Figure 9: Question 3, PISA 2012 mathematics assessment item unit “Sailing ships”

Question 3: SAILING SHIPS

P11823Q03

Approximately what is the length of the rope for the kite sail, in order to pull the ship at an angle of 45° and be at a vertical height of 150 m, as shown in the diagram opposite?

- A 173 m
- B 212 m
- C 285 m
- D 300 m



Source: OECD, 2013c, p. 13.

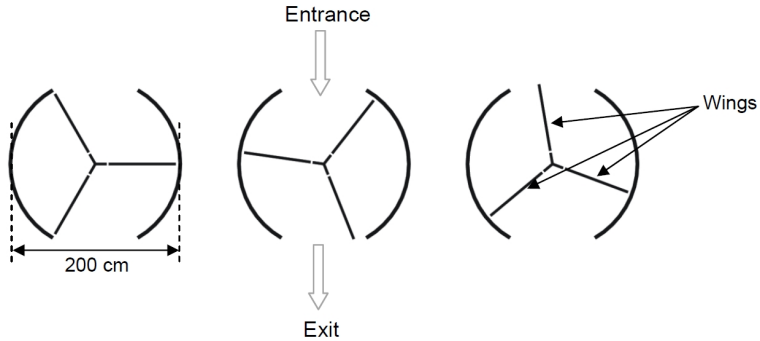
At the objective level, the mathematical content knowledge required to solve question 3 is listed in all three documents. Both the PISA framework and the Norwegian curriculum specifically mention the Pythagoras theorem under the content strands shape and space and geometry respectively, while in the Swedish curriculum, applying theorems is covered in the geometry strand.

Figure 10 displays the stimulus and the first question for the PISA unit “Revolving door” (OECD, 2013c, p. 33). This stimulus explains a technical situation. This item might be seen as a typical mathematical task (de Lange, 1995; Nortvedt, 2012) in that no redundant or irrelevant information is provided. Although the context is technical, no direct link to the goal definitions at the end level is expressed in the stimulus.

Figure 10: Stimulus, PISA 2012 mathematics assessment item unit “Revolving doors”

REVOLVING DOOR

A revolving door includes three wings which rotate within a circular-shaped space. The inside diameter of this space is 2 metres (200 centimetres). The three door wings divide the space into three equal sectors. The plan below shows the door wings in three different positions viewed from the top.



Source: OECD, 2013c, p. 33.

Figure 11: Question 1, PISA 2012 mathematics assessment item unit “Revolving doors”

Question 1: REVOLVING DOOR

PM995Q01 – 0 1 9

What is the size in degrees of the angle formed by two door wings?

Size of the angle:°

Source: OECD, 2013c, p. 33.

Solving the item requires students to understand a 2D representation of a 3D body and connecting the “door wings” mentioned in the text to the corresponding lines in the graphic. To find the answer, students also need to employ their knowledge of the circle geometry. This places Question 1 in the employ process in the PISA framework, and in apply and process category in the Norwegian basic skill numeracy. The calculation itself might be simple, once the student has grasped the situation. Consequently, within the Swedish curriculum, the most relevant ability might

be the concept ability. This is one of the instances where an item is categorised differently at the aim level. However, at the objective level, properties of geometrical objects are included in the geometry strand in the two curricula documents and in the space and shape strand in PISA.

2.9 Conclusions

The results of our analysis show that PISA is indeed relevant to mathematics education in Norway and Sweden. Our first research question pertains to how the definition of mathematical literacy in the PISA mathematics framework is aligned with the goal definitions of the national curricula in Norway and Sweden. In our analysis we identified three levels of goal definitions, representing different levels of generality. At all three levels strong similarities were found. At the most global level, the end level, all the three documents address the fundamental reasons for teaching mathematics at an individual, societal and technical-economical level (as described by Niss, 1996), with the PISA framework mainly focusing on the individual level. Surprisingly, PISA focuses the least on the technical-economical aspects.

Comparisons at the aim level show that both the PISA framework and the two national curricula address many of the same aspects of mathematical activity. For instance, in all the documents the processes involved in problem solving are given much priority. At the objective level where the two national curricula documents provide detailed lists of content to be learned, we can see that although the definition and labelling of the strands differ, mainly the same mathematical content is included under each of the curriculum documents and the PISA framework.

Our second research question examines to what extent the content assessed by the items of the PISA 2012 mathematics assessment is covered by the national curriculum of Norway and Sweden. When allocating the PISA assessment items to the content strands from the two national curricula, we found that all items fitted into at least one of the content strands for both countries. This demonstrates that the mathematical content assessed by the PISA items is included in the national curricula of Norway and Sweden. However, as only the main category was identified

for each country, this analysis does not reveal to what degree different mathematical areas contained in each curriculum is covered by the PISA mathematics assessment. Nonetheless, this analysis confirms PISA's relevance to mathematics education in Norway and Sweden.

While making claims on the relevance of PISA, it is important to keep in mind that our analysis is limited to the intentions of the OECD and the national educational governments. Our study only covers what might be termed an intended curriculum (Cai & Howson, 2013): the stated goals for mathematics education in the PISA framework and the national curricula documents. The realised curriculum, what students learn from mathematics education, has not been addressed.

In addition, many aspects of both the PISA assessment and the assessment items were not part of our investigation. Aspects such as the test and item format, the test situation, the language used in the mathematics items and what constitutes an acceptable answer are also important to consider when discussing the relevance of PISA to the mathematics education in Norway and Sweden. It should also be noted, that both the PISA framework and the national curricula, in particular, address skills and abilities (e.g. oral and collaborative skills) that are considered just as fundamental as the goals discussed in this article.

Thus, the mathematical literacy defined in the PISA framework is reflected in both the Norwegian and Swedish curricula. When solving the PISA assessment items, students have to use mathematical knowledge, abilities and skills described in the national curricula. Consequently, and mindful of the limitations of this study, we conclude that PISA is relevant to mathematics education in Norway and Sweden.

2.10 Acknowledgements

We would like to acknowledge Margareta Enoksson at Stockholm University and Hege Kaarstein at the University of Oslo for their contributions on the categorisations of PISA assessment items.

2.11 References

- Allerup, P. A., Kovac, V., Kvåle, G., Langfeldt, G., & Skov, P. (2009). *FOU-rapport 8/2009. Evaluering av det nasjonale kvalitetsvurderingsystemet for grunnsopplæringen*. Kristiansand/Århus: Agderforskning/The Danish School of Education.
- Burkhardt, H. (2014). Curriculum design and curriculum change. In: Y. Li & G. Lapan (Eds.), *Mathematics curriculum in school education* (pp. 13–33). Dordrecht: Springer.
- Cai, J., & Howson, G. (2013). Toward an international mathematics curriculum. In: M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (Vol. 27, pp. 949–974). New York: Springer.
- Clements, M. A. (2013). From the few to the many: Historical perspectives on who should learn mathematics. In: M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 7–40). New York: Springer. <http://dx.doi.org/10.1007/978-1-4614-4684-2>
- de Lange, J. (1995). Assessment: No change without problems. In: A. Romberg (Ed.), *Reform in school mathematics and authentic assessment* (pp. 87–172). Albany: State University of New York Press.
- Dindyal, J. (2014). International comparative studies in mathematics: An overview. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 320–325). Amsterdam: Springer Reference.
- Elstad, E., Nortvedt, G. A., & Turmo, A. (2009). The Norwegian assessment system: An accountability perspective. *Cadmo*, 17(2), 89–103.
- Lesh, R. A., & Caylor, B. (2009). Differing conceptions of problem solving in mathematics education, science education, and professional schools. In: L. Verschaffel, B. Greer, W. Van Dooren & S. Mukhopadhyay (Eds.), *Words and worlds. Modeling verbal descriptions of situations* (pp. 333–350). Rotterdam: Sense Publishers.
- Lesh, R. A., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning, and problem solving. In: R. A. Lesh & H. M. Doerr (Eds.), *Beyond constructivism. Models and modeling perspectives on mathematics problem solving, teaching, and learning* (pp. 3–34). Mahwah, NJ: Laurence Erlbaum Associate Publishers.
- Lesh, R. A., & Zawojewski, J. (2007). Problem solving and modeling. In: F. K. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 763–804). Charlotte, NC: Information Age.
- Leung, F. K. S. (2014). What can and should we learn from international studies in mathematics achievement? *Mathematics Education Research Journal*, 26(3), 579–605. <http://dx.doi.org/10.1007/s13394-013-0109-0>
- Ministry of Education (2010). *Oppdragsbrev 42-10* [Assignment 42.10]. http://www.udir.no/Upload/larerplaner/forsok/Oppdragsbrev_42-10.pdf
- NDET. (2012). *Framework for basic skills*. Oslo: NDET.
- NDET. (2014). *The education mirror 2014*. Oslo: NDET.

- NDET. (2015a). *Curriculum for the common core subject of mathematics*. Retrieved from <http://www.udir.no/kl06/MAT1-04/Hele/?lplang=eng>
- NDET. (2015b). *Læreplan i matematikk fellesfag* [Mathematics curriculum]. Retrieved from <http://www.udir.no/Lareplaner/Finn-lareplan/#matematikk>
- Niss, M. (1994). Mathematics in society. In: R. Biehler, R. W. Scholz, R. Strässer, & B. Winkelmann (Eds.), *Didactics of mathematics as a scientific discipline* (Vol. 13, pp. 367–378). Dordrecht: Kluwer.
- Niss, M. (1996). Goals of mathematics teaching. In: A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook on mathematics education* (pp. 11–47). Amsterdam: Springer. http://dx.doi.org/10.1007/978-94-009-1465-0_3
- Niss, M., Blum, W., & Galbraith, P. L. (2007). Introduction. In: W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 3–32). Boston, MA: Springer Science + Business Media, LCC.
- Niss, M., & Højgaard, T. (2011). Competencies and mathematical learning. Ideas and inspiration for the development of mathematics teaching and learning in Denmark. English edition. Roskilde: IMFUMFA. This is an English translation of Niss, M., & Jensen, T. H. (2002). *Kompetencer og matematiklæring. Ideer og inspiration til udvikling af matematikundervisning i Danmark*. København: Undervisningsministeriet.
- Niss, M., & Jablonka, E. (2014). Mathematical literacy. In: S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 391–396). Amsterdam: Springer Reference. http://dx.doi.org/10.1007/978-94-007-4978-8_100
- Nortvedt, G. A. (2012). Bruk av tekstoppgaver på matematikktester og -prøver: et kort review. In: T. H. Hopfenbeck, M. Kjærnsli & R. V. Olsen (Eds.), *Kvalitet i norsk skole. Internasjonale og nasjonale undersøkelser av læringsutbytte og undervisning* (pp. 212–222). Oslo: Universitetsforlaget.
- OECD. (2013a). *PISA 2012 Results. Student performance in mathematics, reading, science*. Volume I. PISA: OECD Publishing.
- OECD. (2013b). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. PISA: OECD Publishing.
- OECD. (2013c). *PISA 2012 released mathematics items*. PISA: OECD Publishing.
- OECD. (2015). *Improving schools in Sweden: An OECD Perspective*. OECD: OECD Publishing.
- Robson, C. (2002). *Real world research*. (2nd ed.). Malden, MA: Blackwell Publishing.
- Sjøberg, S. (2014). PISA-syndromet. Hvordan norsk skolepolitikk blir styrt av OECD. *Nytt norsk tidsskrift*, 31(1), 30–43.
- Skolverket. (2011a). *Kommentarmaterial till kursplanen i matematik* [Commentary material to the mathematics curriculum]. Stockholm: Skolverket.

- Skolverket. (2011b). *Lgr11, Curriculum for the compulsory school, preschool class and the recreation centre 2011*. Stockholm: Skolverket.
- Skolverket. (2013). *PISA 2012: 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap*. [PISA212: 15-year-olds knowledges in mathematics, reading and science]. Stockholm: Skolverket.
- Skolverket. (2015). *Med fokus på matematik. Analys av samstämmighet mellan svenska styrdokumentet och den internationella studien PISA* [Focusing on mathematics. An analysis of the alignment between the Swedish steering documents and the international PISA study]. Stockholm: Skolverket.

3. Principal Leadership Styles in Nordic Countries and Their Relationship with Teacher Performance and Attitudes

By Ragnar F. Ólafsson, Directorate of Education, Iceland³

3.1 Summary

The attitudes and practices of school principals in the Nordic countries participating in TALIS 2013 were analysed using hierarchical cluster analysis. Four leadership styles were identified and described, with reference to previous typologies and international reviews on good school leadership practice, and their incidence was compared across the Nordic countries. Each leadership style differs from the other in terms of collaboration, instructional leadership provided, emphasis on mentoring and teacher appraisal. Principals in Group 1 displayed practices and attitudes consistent with instructional leadership, power sharing and collaboration, but had lower scores than Group 3 on mentoring practices. Groups 1 and 3 were similar in many respects, although the latter group's scores were slightly lower on most items. Group 1 scored lower than Group 4 on items assessing consequences of teacher appraisal on career advancement and financial outcomes. Principals in Group 4 most clearly felt that teacher appraisal has tangible consequences or affects teachers' career

³ The findings and conclusions of the chapter are those of the author and do not necessarily reflect the views of the Directorate.

prospects. Group 2 scored low on instructional practices and relatively low on power sharing. Gender differences were also observed in Group 2. The group had twice the number of male principals as female principals, indicating that women employ more instructional leadership and power sharing.

The characteristics of teachers working with each type of principal are explored by comparing a few scales across the four groups, including self-efficacy and job satisfaction. While causal relationships cannot be inferred, this exploratory study suggest that neglect of instructional practices among principals is associated with lower self-reported teacher efficacy in instruction. In fact, teachers in Group 2 are low on most variables except job satisfaction. Teachers in Group 3 report high efficacy in instruction and a greater need for professional development. Teachers in Group 4 have a relatively low job satisfaction, but report high efficacy in student engagement. The importance of exploring the link between principal's characteristics and teaching practices and the eventual practical implications of such findings are discussed.

3.2 Introduction

Nordic schooling systems have undergone radical changes in the recent years. These include devolution of power from state to local municipalities and greater emphasis on accountability and effectiveness (Ärlestig & Johansson, 2011; Hansen, Jóhannsson, & Lárusdóttir, 2008; Moos, Møller, & Johansson, 2004). The depth and scope of these developments vary between the Nordic countries. For example, Sweden has surpassed Norway in introducing incentive tools (Helgøy & Homme, 2004).

These changes in the schooling systems have affected the practices of school principals and eventually led to role conflicts. Moos (2011) explains, "the contemporary neo-liberal governance with new demands for strong leaders [are] difficult to reconcile with the (traditional) urge for trust", especially among Danish and Norwegian principals (Moos, 2011, p. 148). The principal needs to strike a balance between demands for accountability (e.g. PISA competition) and those for "the comprehensive Democratic Bildung and room for collaboration and creativity" (Moos,

2011, p. 157). Further, while the state provides detailed aims and standards, some of the tasks and decisions are decentralised to schools.

The school principal has to respond to these conflicting demands (Moos, 2011). However, despite an increased focus on managerial accountability and effectiveness, principals still seem to follow the “norm of non-interference in the teachers’ classroom activities” (Møller, 2011, p. 129). Møller (2011) argues that classroom practice in Norway has undergone little change, possibly because schools (at least outside Oslo) are not under “threat of sanctions, if exam scores are low” (p. 129).

The aim of the present chapter is to explore principals’ leadership styles in Nordic countries on the basis of data from TALIS 2013 principals’ questionnaires. Cluster analysis is employed to group principals on the basis of self-report data on their practices and attitudes as leaders.

The Nordic countries are, for the purposes of this study, treated as one geographic and cultural entity. Comparisons are drawn between groups of school principals, regardless of the Nordic country they work in. Further, teaching practices and characteristics of the teaching staff are compared across groups of principals with a view to explore a link between particular leadership styles and teaching practices and attitudes.

The following section presents an overview of some of the previous and traditional strategies of assessing and classifying leadership styles of principals. Further, international and Nordic research on the relationship between leadership style and students’ academic achievement is discussed. We then examine how the present leadership styles of principals align with indicators of good practice discussed in previous studies on leadership practice and student achievement.

3.2.1 *Typology of principals’ leadership style*

Transactional, transformational, pedagogical and distributed leadership styles have been discussed in previous studies. Research also suggests that these styles are not mutually exclusive (Urick & Bowers, 2014) and that other ways of classification are also possible.

Burns (1978, as cited in Hansen, 2013) distinguished between transactional and transformational leadership. The *transactional* style is business-like and assumes that both the follower and leader have a direct

benefit from the relationship (Hansen, 2013, p. 80). Examples of such benefits would be a promotion or pay rise. Robinson, Hohepa and Lloyd (2009) point out that a transactional leader “specifies what is expected and provides consequences for meeting or not meeting those expectations” (Robinson *et al.*, 2009, p. 85).

Transformational leadership is more interactive in that both the leader and follower engage in constructive criticism, encouragement and vision sharing (Hansen, 2013, p. 80). Transformational leaders provide inspiration and encourage creativity. They convey a sense of purpose and make sure that the teachers trust the leadership and feel valued (Robinson *et al.*, 2009, p. 85).

According to Robinson *et al.* (2009, p. 88), *pedagogical leadership* consists of establishing an academic mission, monitoring and providing feedback on teaching and learning and “promoting professional development”. The items used for the assessment of instructional leadership typically measure direct involvement of the principal in teaching and learning (e.g. conducting observations in class, discussing progress with staff, reviewing student results) (Camburn, Huff, Goldring, & May, 2010; Robinson *et al.*, 2009, p. 89).

Shared instructional leadership differs from the traditional instructional leadership in the sense that teachers assume the leadership responsibility as they “encourage others to improve their professional practice, or learn together with their school colleagues” (Moller & Katzenmeyer, 1996, as cited in Marks & Printy, 2003, p. 374).

A principal can be characterised by two or more of these leadership styles. Such combinations can have added benefits, as revealed in Marks and Printy’s (2003) study conducted in 24 schools from 16 states in the United States undergoing reform efforts. An interesting finding of the study is that transformational leadership is a necessary but insufficient condition for instructional leadership. When the two occur together (*integrated leadership*), both teachers and students display high performance (Marks & Printy, 2003).

Urick and Bowers (2014) believe that there is little evidence on the extent to which principals choose to exercise any of three “idealized” leadership styles (i.e. transformational, instructional and shared instructional). They maintain that these styles have substantial conceptual overlap and may not represent adequately the differences between principals. In their study on practiced leadership styles involving 8,000 principals in the United States, Urick and Bowers (2014) identified three types of principals: *controlling* (frequently exercised principal leadership), *balkanizing* (leadership shared with teachers) and *integrating* (consisting of frequent principal leadership as well as sharing leadership with teachers). All three types of principals reported practicing managerial tasks or transactional leadership. Controlling and integrating principals also engaged in transformational and instructional leadership; however, the former were more reluctant to assign managerial or instructional leadership tasks to the teachers (p. 121).

Goldring, Huft, May and Camburn (2008) used cluster analysis to identify different types of principals on the basis of their time allocation practices. Three different leadership styles were identified: *eclectic* (whose activities were distributed across different types of practices), *instructional* and *student* (centred on student affairs).

Many researchers believe that leadership studies in schools should not focus exclusively on the principal as leader. Hansen (2013) explains that *distributed leadership* can occur at various levels in an institution where an individual or group defines the discourse on the topic under consideration. Such leadership has been termed *collaborative*, *team*, *shared* or *participative* (Hansen, 2013, p. 80).

Spillane (2005) criticises approaches that equate “leadership with the actions of those in leadership positions” because leadership is more distributed and not always formalised (p. 145). Leadership can be a social force that comes into being within an institution, transcending formal roles, as a practical problem is tackled (Hansen, 2013, p. 81).

Ludvigsson (2009) echoes this sentiment by stating that leadership is produced and formed by school heads and teachers together – *co-produced leadership*. She questions the popular image of an independent leader who is expected to govern her/his organisation with a firm hand

and argues that school leadership is the product of involvement from both heads and teachers (p. 9).

Abrahamsen (2008, p. 33) also sees leadership as something that is created in the relationship between different actors instead of an outcome of a set of leadership qualities of a given “heroic” leader. Similarly, Nordzell (2007, p. 79) adopts a discursive approach to understanding how leadership emerges through team-level interactions. Nordzell uses interesting categories for describing school leaders: *innovative*, *creative*, *progressive* and *improving* (p. 81).

Three types of leadership have been identified by analysing data from 67 Danish gymnasia: (1) planning collegial leadership which focuses on “central and formalized steering via planning, management of details and clear demands on teams and teachers”, (2) dialogue-based leadership, which is based on “teacher involvement, collaboration and dialogue in teams in order to produce shared collective meaning” (Pedersen & Ryberg, 2009; Pedersen, 2010, as cited in Moos, 2011) and (3) strategic leadership, which revolves around a hierarchical structure geared towards the market place and external expectations (Moos, 2011, p. 142–143).

Another interesting approach to examining principals’ leadership typology is how they present their schools in meetings with parents (Larsson, 2006, as cited in Årlestig & Johansson, 2011, p. 102). While one set of principals talks about school tradition, the other tends to highlight school efficiency or emphasise new and different learning strategies.

3.2.2 Leadership styles and student academic achievement

Why is it important to examine the leadership styles of school principals? One of the main reasons is evidence to the effect that these styles may influence student scholastic achievement directly or indirectly. Interestingly, Louis, Leithwood, Wahlstrom and Andersen (2010, as cited in Hansen, 2013) indicated that only up to 20% of pupils’ academic achievement can be traced to schools. Marzano (2000, cited in Hansen, 2013) assigns 7% to school factors (e.g. time spent learning, supervision, expectations, parent participation, leadership and cooperation of teachers) and 13% to factors inside the classroom (Hansen, 2013, p. 83). Klitgaard

and Hall (1974, as cited in Marzano, 2000, p. 48), however, point out some schools have very large or small effect sizes and adopting this approach tends to disregard such schools.

On the basis of a meta-analysis, Robinson *et al.* (2009) conclude that pedagogical leadership is a better predictor of student outcomes than practices associated with transformational leadership. The mean effect size estimate for pedagogical leadership “is about three times that of the transformational leadership” (Robinson *et al.*, 2009, p. 90; Robinson, Lloyd & Rowe, 2008).

Mattar (2012) compared principals in five high-performing and five low-performing schools in Lebanon. He concluded that principals of the high-performing schools adopted a stronger instructional leadership style than those of the low-performing schools.

Shatzer, Caldarella, Hallam and Brown (2014) examined the impact that school leaders have on student achievement to identify the leadership practices associated with increased student achievement. They found that *instructional* leadership explained more achievement variance than *transformational* leadership.

In successful schools studied by Ärlestig (2008, cited in Ärlestig & Johansson, 2011, p. 86) the principals are much involved in learning and instruction; they visit classrooms more often and provide feedback.

Kreiner and Mehlbye’s (2000) study of 89 Danish schools indicates that three aspects of school leadership are important for developing a good school: pedagogic leadership (i.e. leaders contribute to the teaching and learning environment), close relations between teachers and leaders and active creation of a good working environment (Moos, 2011, 145).

Ringmose and Mehlbye (2004) studied leadership in 15 high-performing schools (*d. gode eksempler*) in Denmark and described effective leadership as including feedback (i.e. follow up on decisions or plans) and involving staff members in decisions (e.g. on strategy plans) (Moos, 2011, p. 148).

A Norwegian study on the characteristics of successful leadership practice in primary and secondary schools indicated that successful school leadership requires team work (Møller, Eggen, Fuglestad, Langfeldt, Presthus, Skrøvet, Stjernstrøm and Vedøy, 2005, as cited in Møller, 2011, p. 121) and “democratic principles and values” (Møller *et*

al., 2005, p. 584). Hallinger and Heck (2010) also found that “collaborative school leadership can positively impact student learning” (Hallinger & Heck, 2010, p. 95).

3.2.3 *Teacher preferences: Openness and participation*

Teachers’ preferences for leadership styles among principals may also provide clues on desirable aspects of school leadership. Olofsson (2011, as cited in Årlestig & Johansson, 2011, p. 97) showed that teachers favoured a transparent and democratic leadership style.

Teachers welcomed feedback from school leaders and believed that too few conversations between teachers and principals led to poor transmission of the school’s vision and objectives (Årlestig, 2011; Årlestig & Johansson, 2011, p. 98). Interestingly, Hjartardóttir (2002) indicates that principals, too, want to spend more time on collaborating with teachers than administrative tasks.

Hansen (2013, p. 97) reports that teachers in Iceland seek to participate in the decision-making process within the school (Björnsdóttir, Hansen, & Jóhannsson, 2006), which is an integral part of the decentralisation of education.

3.2.4 *Extensive Reviews*

Hallinger (2011) reviewed studies from the past three decades on leadership and learning and concluded that principals must articulate explicitly the values of the school, as they guide decision making (p. 137). He points out that while shared, collaborative and distributive leadership styles are all valuable, the underlying role of a principal is to support the leadership of other members (p. 138).

Hallinger (2011) also notes that to improve the school’s capacity, the principal needs to develop human capacity, i.e. get involved in the professional learning of teachers. Although instructional and transformational leadership influence learning, it is difficult to identify the optimum leadership style for promoting learning. Instead, the most effective leadership style is one that is adapted to the school context (Hallinger, 2011, p. 137; Hallinger & Heck, 2010).

On the basis of their literature reviews and meta-analyses, Robinson *et al.* (2009, p. 84) conclude that leaders can make a difference by “engaging with instruction, setting direction and goals, increasing teacher capacity, creating systems that support teaching, and building relationships in which people feel valued and supported”. They add that an assessment of *indirect* effects of principal leadership on student outcomes is more likely to show an effect for leadership style than measuring only direct effects. Thus, to examine the influence of leadership, they call for an evaluation of “the impact of leadership on school conditions *and* the impact of those conditions – such as teacher culture – on student outcomes” (Robinson *et al.*, 2009, p. 83).

3.2.5 *Effect of leadership styles on job satisfaction of teachers*

Job satisfaction of teachers is another factor that could be influenced by the leadership style of principals. Bogler (2002) examined different variables of job satisfaction among 745 Israeli teachers in elementary, junior high and high schools, including teachers’ evaluations of transactional and transformational leadership styles. Teachers with high job satisfaction perceived their principals as more transformational and less transactional, while teachers with low job satisfaction perceived their principals to be more transactional.

Overall, the studies and reviews reported above seem to place emphasis on instructional or pedagogical leadership style, but also accord importance to transformational leadership. The items in the TALIS 2013 principal questionnaire do not cover fully each of the leadership styles mentioned above but approximate them to different degrees, which allows for comparisons between the literature and the clusters identified in the present study. Of special interest is evaluating the present groups against the evidence on what constitutes desirable or good practice.

3.3 Aims and research questions

Firstly, this study aims to identify leadership styles or practices among principals in the Nordic countries that participated in TALIS 2013, using response data from items related to attitudes and practices of principals. Items related to personal background (e.g. age or education) and those related to the school (e.g. size, location or other characteristics of the institution) are not included. Of course, a principal's views and practices may be determined to a great extent by certain characteristics of the school (e.g. demographic composition, economic status); however, this study aims to construct a portrait of the principals that is not influenced by such external characteristics. Leadership styles of the principals are identified and their distribution within and between the Nordic countries is explored.

As a second step, the identified leadership styles are linked to a number of attitudinal and behaviour scales from the TALIS teacher questionnaire in an effort to link principal's characteristics with teacher practices and attitudes. Special attention is given to measures of teacher efficacy and job satisfaction, as these are known to be associated with student outcomes. It is important to mention that the analysis design does not support the drawing of any causal inferences.

Thirdly, the identified clusters of leadership styles are examined vis-à-vis characteristics identified in previous research as conducive to greater student achievement. This comparison has some limitations in that the characteristics of different leadership styles identified in the literature are not fully captured by the TALIS items used in the present analysis.

3.4 Methods

Participants consisted of school principals in lower secondary schools (ISCED 2) of Nordic countries (Denmark, Finland, Iceland, Norway and Sweden) who responded to the TALIS Principal Questionnaire 2013. Only participants (principals) who completed all the 30 items were included, as replacing missing values with means or other derivations of available data may bias the results of the cluster analysis. A total of 540

principals were included, and the number of teachers working in the schools of these principals and providing data for comparisons ranged from 6,827 to 8,808.

A total of 30 items (Table 6) were selected from the TALIS Principal Questionnaire as they pertained to principals' attitudes and/or practices. They were rated on 4-point Likert scales (1 = strongly disagree to 4 = strongly agree; 1 = never to 4 = very often or always; 1 = not at all to 4 = a lot; 1 = not important at all to 4 = of high importance). These items are taken from the TALIS 2013 Principal Questionnaire (questions 21 (a,b,c,d,e,f,g,h,i), 22 (a,b,c,d,e), 26 (h,i), 29 (a,b,c,d,e,f,g,h) and 38 (a,b,c,d,e,f)), and they describe leadership practices, attitudes to power sharing, views on certain potentially limiting factors to the principal's effectiveness, actions following teacher appraisal and attitudes to mentoring.

Table 6: Items from the TALIS 2013 questionnaire included in the cluster analysis

Question number	Item wording
Q21	Please indicate how frequently you engaged in the following in this school during the last 12 months
a	I collaborated with teachers to solve classroom discipline problems
b	I observed instruction in the classroom
c	I took actions to support co-operation among teachers to develop new teaching practices
d	I took actions to ensure that teachers take responsibility for improving their teaching skills
e	I took actions to ensure that teachers feel responsible for their students' learning outcomes
f	I provided parents or guardians with information on the school and student performance
g	I checked for mistakes and errors in school administrative procedures and reports
h	I resolved problems with the lesson timetable in this school
i	I collaborated with principals from other schools
	1 = never or rarely; 4 = very often
Q22	How strongly do you agree or disagree with these statements as applied to this school?
a	This school provides staff with opportunities to actively participate in school decisions
b	This school provides parents or guardians with opportunities to actively participate in school decisions
c	This school provides students with opportunities to actively participate in school decisions
d	I make the important decisions on my own
e	There is a collaborative school culture which is characterized by mutual support
	1 = strongly disagree; 4 = strongly agree
Q26	To what extent do the following limit your effectiveness as a principal in this school?
h	High workload and level of responsibilities in my job
i	Lack of shared leadership with other school staff members
	1 = not at all; 4 = a lot

Question number	Item wording
Q29	Please indicate the frequency that each of the following occurs in this school following a teacher appraisal
a	Measures to remedy any weaknesses in teaching are discussed with the teacher
b	A development or training plan is developed for each teacher
c	If a teacher is found to be a poor performer, material sanctions such as reduced annual increases in pay are imposed on the teacher
d	A mentor is appointed to help the teacher improve his/her teaching
e	A change in a teacher's work responsibilities (e.g. increase or decrease in his/her teaching load or administrative/managerial responsibilities)
f	A change in a teacher's salary or a payment of a financial bonus
g	A change in the likelihood of a teacher's career advancement
h	Dismissal or non-renewal of contract 1 = never; 4 = always
Q38	How would you generally rate the importance of mentoring for teachers and schools?
a	To improve teachers' pedagogical competence
b	To strengthen teachers' professional identity
c	To improve teachers' collaboration with colleagues
d	To support less experienced teachers in their teaching
e	To expand teachers' main subject(s) knowledge
f	To improve students' general performance 1 = not important at all; 4 = of high importance

Most of the participants in the Nordic countries completed the survey online. Details about the data collection, sampling, administration and other issues related to the methodology are available in TALIS international reports (e.g. OECD, 2014) and national reports.

The responses were subjected to hierarchical cluster analysis (HCA), using SPSS 19. Principals were grouped on the basis of the similarities in their attitudes and practices. Ward's method was used with Squared Euclidean distance. Cluster analysis is an exploratory method for classifying observations or cases into groups without any prior theory about the number of groups (Papi & Teimouri, 2014). After exploring a number of solutions, a four cluster solution was adopted. The analysis indicated differences between countries, yet the clusters overlapped across countries and no group contained very few members.

ANOVAs (with Scheffé post hoc tests) were performed on a number of scales or single items to determine whether certain teacher practices or attitudes were associated with any of identified leader groups. The scale names and items are displayed in appendix 1 (see also OECD, 2014).

In the analysis presented here, mean scores on the scales were computed after conducting confirmatory factor analyses using the Nordic data set, without separating the countries, and considering RMSEA, CFI, TLI and SRMR (Brown, 2015; West, Taylor and Wu, 2012).

3.5 Results

Cluster membership was examined across countries (Table 7) by means of HCA.

Table 7: Number of principals (n=540) per country who fall into each of the clusters⁴

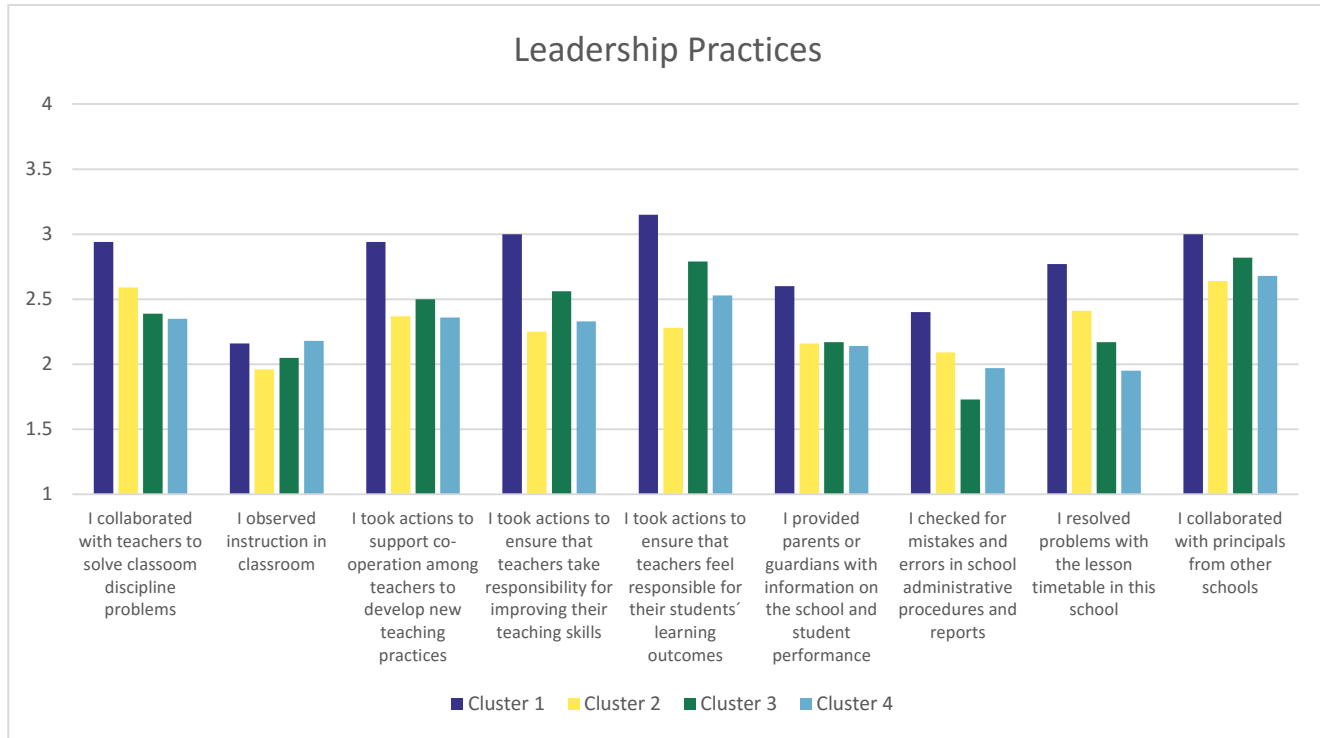
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Denmark	36	57	12	2	107
Finland	36	59	3	4	102
Iceland	30	20	27	2	79
Norway	36	47	10	2	95
Sweden	34	53	14	56	157
Total	172	236	66	66	540

Biases in the sampled or completed questionnaires are not corrected for by weights and therefore comparisons between countries should be made with caution. The principals are grouped into clusters on the basis of their responses. Interestingly, Cluster 4 consists mostly of Swedish principals. Clusters 1 and 2 also have a high concentration of Swedish principals. Cluster 3 has a relatively high proportion of Icelandic principals. The distribution of Norwegian and Danish principals is similar; most are in Clusters 2 and 1. The largest portion of Finnish principals are also distributed between these two clusters.

To examine the differences between the clusters of principals, ANOVAs were performed on the 30 items. Results of this analysis are displayed and discussed below (see figures 12–16 and appendix 2).

⁴ “Cluster” and “Group” are used interchangeably.

Figure 12: Comparison of mean scores of the four clusters on engagement in certain leadership practices (Q21)



Note: 1 = Never or rarely; 2 = Sometimes; 3 = Often; 4 = Very often.

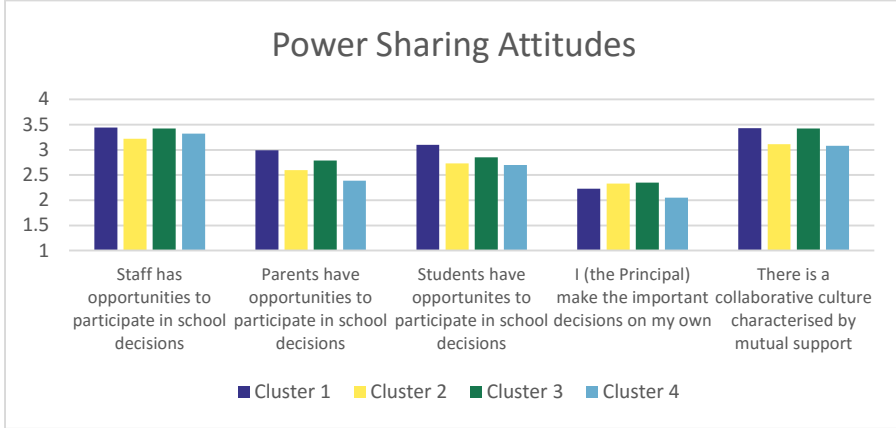
Principals in Cluster 1 “often” perform various instructional tasks such as supporting cooperation among teachers to develop new teaching practices, ensuring that teachers take responsibility for improving teaching skills as well as for students’ learning outcomes. They also “often” collaborate with principals from other schools and with teachers to solve classroom discipline problems. They routinely provide information about school and student performance to parents or guardians, more often than any other group of principals. They are quite involved in resolving problems related to lesson timetables or administrative tasks such as checking for mistakes. However, they are not very regular with observing instructions in the classroom.

Cluster 2 scores low on instructional tasks (e.g. ensuring that teachers take responsibility for improving teaching skills or helping them to develop new teaching practices). Principals in this cluster are relatively high on solving classroom discipline problems or on resolving problems with the timetable. They have the second highest scores on checking for mistakes and errors in administration and reports. Collaborating with principals from other schools is the topmost priority for this cluster, as indicated by the scores.

The profile of Cluster 3 is similar to the first group, but with lower frequencies. It has the second highest score on instructional tasks, but scores lower than Clusters 1 and 2 on dealing with problems of various sorts (i.e. checking for errors in administrative procedures/reports and solving problems with timetable and discipline problems).

Cluster 4 falls between Clusters 2 and 3 in most respects and well below Cluster 1. This group scores lower than Cluster 3 on the instructional tasks. Principals in this cluster are concerned with ensuring that the teachers feel responsible for their students’ learning outcomes. Compared with the other groups, it scores relatively high on observing instruction in the classroom.

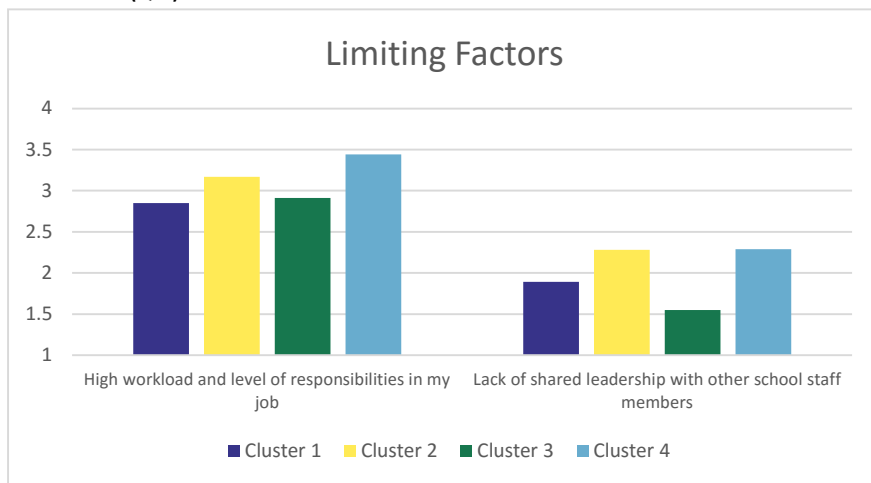
Figure 13: Comparison of mean scores of the four clusters on statements about power sharing (Q22)



Note: 1 = Strongly disagree; 2 = Disagree; 3 = Agree; 4 = Strongly agree.

Overall, the involvement of parents in school decisions is the item with the most varying means. Principals in Cluster 1 give parents and students ample opportunities to participate in school decisions and seldom make important decisions on their own. Clusters 1 (and 3) prioritise collaboration within the school (i.e. giving staff opportunities to participate in decision making and having a collaborative culture in the school). Principals in Cluster 4 seek the least parental involvement in school decisions.

Figure 14: Comparison of mean scores of the four clusters on factors limiting principals' effectiveness (Q26)

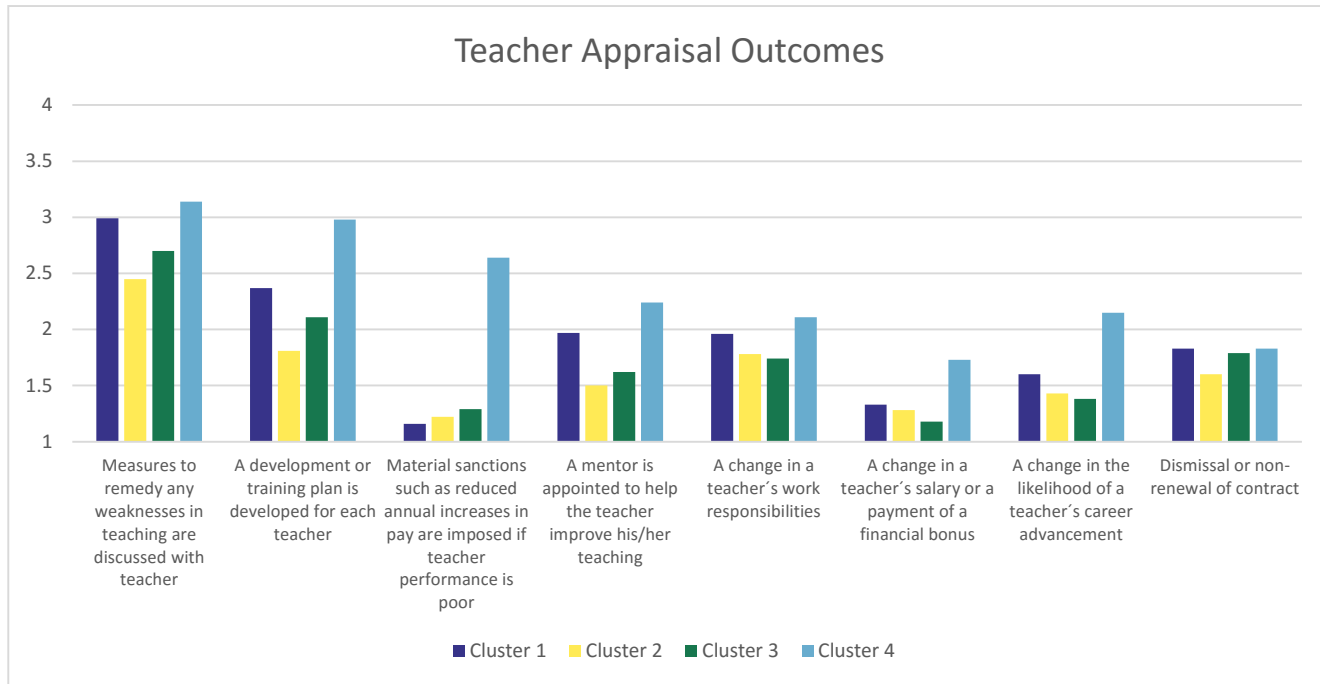


Note: 1 = Not at all; 2 = Very Little; 3 = To some extent; 4 = A lot.

Principals in Clusters 4 and 2 find work load to be highest deterrent to their effectiveness. The lack of shared leadership is the least important limiting factor for Clusters 3 and 1.

Overall, the trends in Figures 13 and 14 suggests that shared leadership may be associated with lower workload since Clusters 1 and 3 have high scores on items related to power sharing and complain less about high workload.

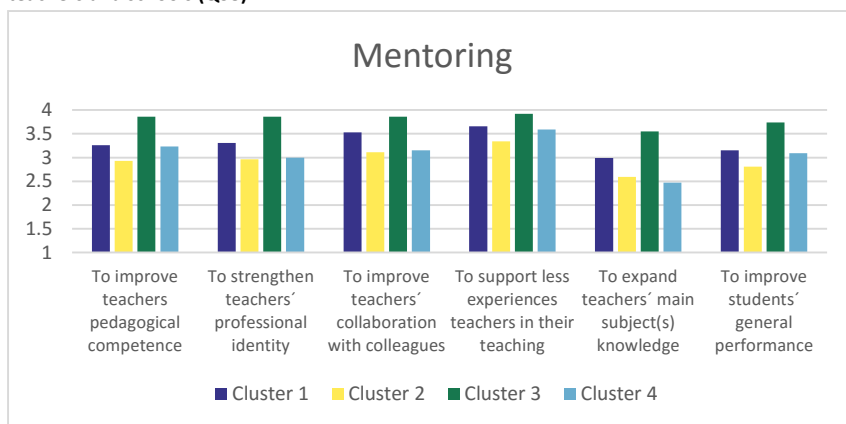
Figure 15: Comparison of mean scores of the four clusters on outcomes of teacher appraisal (Q29)



Note: 1 = Never; 2 = Sometimes; 3 = Most of the time; 4 = Always.

Cluster 4 scores highest on the enforcement of various teacher appraisal outcomes, most notably material sanctions such as reduced pay increases if the teacher’s performance is poor. Principals of this group, more than any other group, claim that changes in teacher’s career advancement are more likely to follow appraisal. They also discuss remedial measures with teachers to address weaknesses in teaching and help develop a training plan. The other three groups are similar in most respects and different from Group 4 in terms of the tangible appraisal outcomes for the teacher. The other groups discuss weaknesses of the teacher, but a more formal training plan does not necessarily follow.

Figure 16: Comparison of mean scores of the four clusters on the importance of mentoring for teachers and schools (Q38)



Note: 1 = Not important at all; 2 = Of low importance; 3 = Of moderate importance; 4 = Of high importance.

Cluster 3 accords the highest importance to all measured aspects of mentoring, followed by Cluster 1. Cluster 2 assigns the lowest importance (“moderate” importance) to mentoring. It is interesting to note that Cluster 4 does not prioritise mentoring, despite favouring tangible appraisal outcomes for teachers and correction measures (with development plans and discussions) for overcoming teaching weaknesses. These results suggest that for Cluster 4 teacher development is mostly reactive (i.e. when problems occur) rather than continuous, as steady mentoring would imply.

3.6 Profile descriptions of the four clusters of principals

Cluster 1: *Collaboration, instructional and administrative leadership*: Principals in Cluster 1 consider it their role, more so than other principals, to help teachers develop new teaching practices, improve their teaching skills and assume responsibility for their students' learning outcomes. They are willing to collaborate with teachers to solve discipline problems in the classroom and readily share information with parents about the performance of the school and its students. They check for mistakes and errors, resolve timetable problems and collaborate with principals from other schools. Overall, these principals consider motivating and guiding teachers a part of their role. Their willingness to engage in collaboration, both inside and outside the school, is reflected in all measures (e.g. items pertaining to mentoring).

They are open to sharing power, are in favour of involving parents and/or students in school decisions and do not make important decisions on their own. They mostly discuss weaknesses with the teachers following an appraisal, yet this does not often result in the creation of a training plan for the teacher. The appraisal is not likely to affect the salary, pay or career advancement of teachers.

Cluster 2: *Reactive – under siege*: This group of principals forms a sort of antithesis to the principals in Cluster 1. Their scores are low on majority of the items, except those related to *limiting factors*, which indicates that they experience high workload and responsibility as well as a lack of shared leadership with other staff members. Principals in this group are at opposing ends with those in Cluster 1 in terms of their participation in developing teacher instructional practices or skills. These principals seem to be beset by problems of various sorts, as they score relatively high on “checking for mistakes” and solving discipline and/or timetable problems. They are quite low on power sharing attitudes and tend to (relatively) make the important decisions themselves. Teacher appraisal by such principals rarely leads to any tangible outcome. The same applies for mentoring, to which they do not accord much importance, compared to the other groups.

Cluster 3: *Moderate instructional leadership, emphasis on mentoring, little reaction following teacher appraisal.* Principals in Cluster 3 are in many respects similar to those in Cluster 1, yet slightly less supportive in developing teaching practices, increasing teachers' sense of responsibility for student learning outcomes and improving teaching skills. They differ more markedly from Cluster 1 as they are seldom involved in solving discipline problems, checking mistakes in reports or solving problems with the timetable. These principals do not feel burdened by work load or by lack of shared leadership with other school staff members.

They outdo Cluster 1 in terms of the importance accorded to mentoring. More than any other group, principals in this group consider mentoring as important for improving pedagogical knowledge and subject knowledge, supporting less experienced teachers, strengthening teachers' professional identity and improving teacher collaboration and student performance.

While obviously keen on improving teacher skills through mentoring, these principals do not make great use of teacher appraisal and certainly do not favour pay-based changes following appraisal outcomes. While they assign high importance to mentoring, they are only in second place (after cluster 4) when it comes to appointing mentors to help teachers improve following an appraisal. Their work on improving teachers seems to take place via mentoring and through (more regular?) instructional practices, not at periodic teacher appraisals.

Cluster 4: *Reactive, with financial incentives. Importance of teacher appraisal with consequences.* The key characteristic of this group of principals is that they favour tangible outcomes following appraisals and expect teachers to face the consequences, positive or negative, of the outcome. These principals tend ("sometimes" or "most of the time") to take action following an appraisal by discussing remedies to teaching weaknesses with the teacher and developing a training plan for him/her. It is noteworthy that principals in this cluster are more willing than others to impose material sanctions such as reduced pay increases if a teacher's performance is poor. Principals in this group complain the most about high workload and level of responsibility. They share little power with parents or students and accord relatively little importance to mentoring, which is interesting given the relatively severe penalties they are willing

to impose for negative teacher appraisals. They may be labelled as “reactive” more than “proactive” in the area of teacher improvement and do not regularly engage in instructional leadership practices.

3.7 Other principal characteristics associated with the four clusters

It is interesting to note that there are hardly any gender differences in the membership of principals in each of the four clusters except Cluster 2, which has almost twice as many males as females (Table 8).

Table 8: Number of male and female principals in each of the four leadership groups

	1	2	3	4	Total
Female	88	84	31	38	241
Male	84	152	35	28	299
Total	172	236	66	66	540

3.8 Association between the four principal clusters and teacher practices and attitudes

The second aim of the study is to link teacher practices and attitudes to the leadership practices of their principals, as identified from the cluster analysis. The following figures and tables compare (with ANOVAs, see appendix 2) the four groups of principals on the basis of their teachers’ mean self-efficacy, job satisfaction, teacher-student relations and disciplinary climate, teacher collaboration, perceived need for professional development and adherence to constructivist beliefs. Scales for measuring these constructs are developed from the TALIS Teacher Questionnaire dataset. These comparisons help establish whether the principals’ practices are associated with particular teacher practices and attitudes.

Figure 17: Self-efficacy of teachers working with principals from each of the four leadership groups



Note: 1 = Not at all; 4 = A lot.

A significant difference is observed between the scores of efficacy in instruction across the groups. Teachers working with Cluster 2 principals have lower efficacy in instruction than those working principals from Clusters 3. Efficacy for student engagement is higher in cluster 4 than in clusters 2 and 3. Efficacy in classroom management does not show a significant difference between the groups.

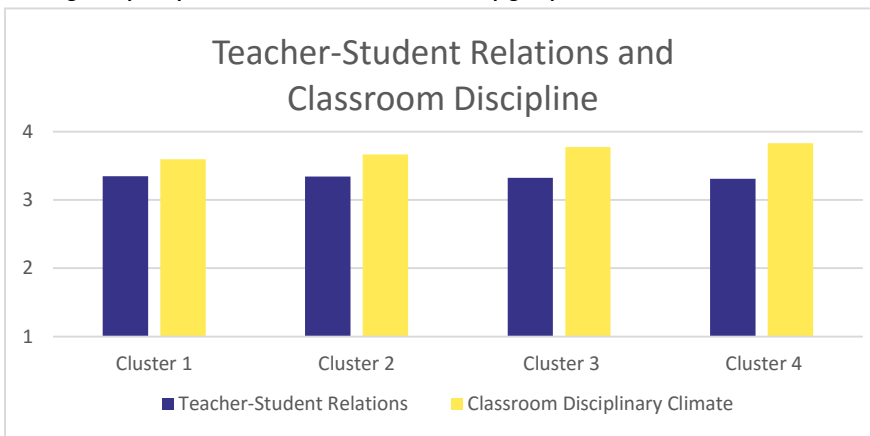
Figure 18: Job satisfaction of teachers working with principals from each of the four leadership groups



Note: 1 = Strongly disagree; 4 = Strongly agree.

Teacher job satisfaction is measured with two scales. Results show that satisfaction with the current work environment and the profession is similar across three of the four groups, but lower among teachers working with a principal in Cluster 4. Groups 1 and 2 believe more than the others that the teaching profession is valued in society. Interestingly, teachers in cluster 4 are satisfied with their *own* performance but are low on other satisfaction measures.

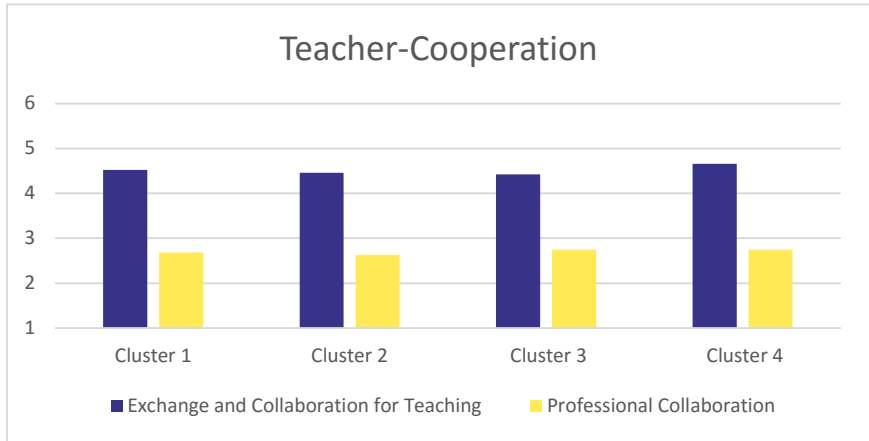
Figure 19: Teacher-student relations and classroom disciplinary climate according to teachers working with principals from each of the four leadership groups



Note: 1 = Strongly disagree; 4 = Strongly agree.

No significant differences are observed in teacher-student relations but classroom discipline is perceived to be better by teachers working under principals in clusters 3 and 4 compared to cluster 1.

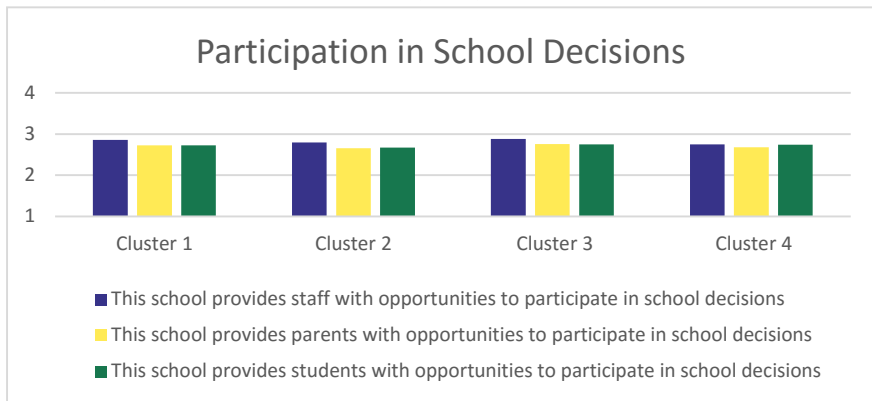
Figure 20: Collaboration among teachers working with principals from each of the four leadership groups



Note: 1 = Never; 6 = Once a week or more.

Teacher cooperation is assessed with two scales: (1) exchange and coordination and (2) professional collaboration. Principals of Cluster 4 ensure the highest cooperation among teachers on the first dimension. Cluster 1 is also higher than cluster 3 (but lower than 4) on that dimension. Clusters 4 and 3 are higher than cluster 2 on professional collaboration.

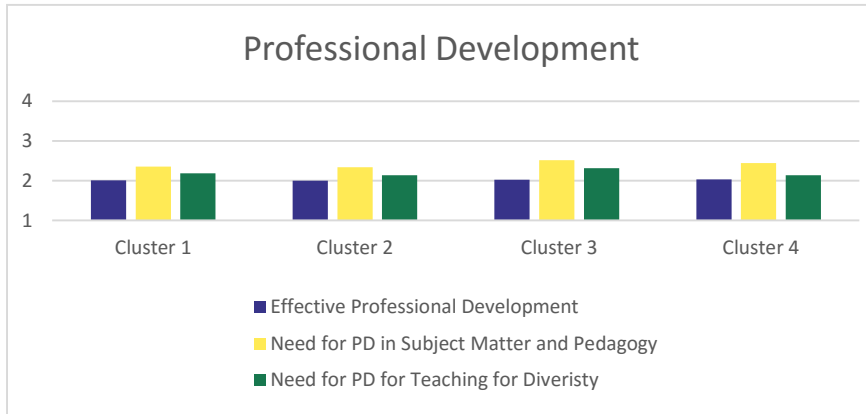
Figure 21: Participation of staff, parents and students in school decisions among teachers working with principals from each of the four leadership groups



Note: 1 = Strongly disagree; 4 = Strongly agree.

The level of cooperation among teachers does not seem to be related to the level of participation in school decisions among other stakeholders such as parents, since Clusters 1 and 3 perform slightly better in that respect than Clusters 2 and 4.

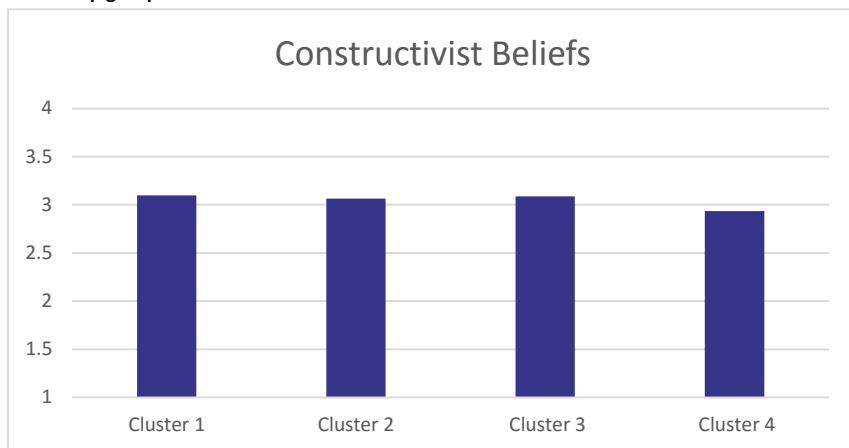
Figure 22: Professional development among teachers working with principals from each of the four leadership groups



Note: 1 = Not in any activities; 4 = Yes, in all activities; 1 = No need at present; 4 = High level of need.

Teachers working with principals from Cluster 3 express more need for professional development than teachers in the other groups. Cluster 4 is also significantly higher than clusters 1 and 2 (but lower than cluster 3) on the need for professional development in subject matter and pedagogy.

Figure 23: Constructivist beliefs of teachers working with principals from each of the four leadership groups



Note: 1 = Strongly disagree; 4 = Strongly agree.

Teachers working with principals in Cluster 4 scored the lowest on the constructivist beliefs scale.

To summarise, teachers working with principals in Cluster 2 have the lowest efficacy in instruction and teachers in cluster 4 have the highest efficacy in student engagement. Job satisfaction is lowest among teachers working under Cluster 4, especially satisfaction with the profession, but these teachers are however satisfied with their own performance. Disciplinary climate is high in Clusters 3 and 4. The needs for professional development is higher among teachers working with principals from Clusters 3, and their instructional efficacy is also high. Teachers associated with Cluster 4 are low on the constructivist beliefs scale but report greater collaboration within the school – among themselves. Teachers working with Cluster 3, however, provide stakeholders with more opportunities to participate in school decisions than those working with principals in Cluster 2.

3.9 Discussion

On the basis of questionnaire items from TALIS 2013, which assess principal practices and attitudes, four clusters of principals have been identified. As with the results of any HCA, the selection of a four-cluster solution is to a certain extent arbitrary. Care has been taken to balance the number of clusters and items in order to avoid a “grasp all – lose all” situation. The clusters identified here are meaningful and manageable and significantly associated with important teacher practices.

Cluster 1 consists of principals who display collaborative, instructional and administrative leadership; Cluster 2 consists of principals who neglect much of these practices (relatively) and complain of high workloads; Cluster 3 shows moderate instructional leadership, placing high emphasis on mentoring; and Cluster 4 comprises principals favouring high-stakes teacher appraisals and accountability but offering relatively little instructional input.

Each Nordic country is represented in all four leadership groups, indicating a certain overlap in practices across Nordic countries. Nonetheless, Cluster 4 consists mostly of Swedish principals. This is not surprising as in Sweden, unlike in other Nordic countries, decisions related to teacher pay have to a great extent been delegated to the school principal. Nordhaug (2014, p. 65) also draws attention to OECD reports that explain that Swedish principals have greater autonomy than, say, Norwegian principals.

Principals in Cluster 1 display many qualities that are akin to pedagogical leadership (Robinson *et al.*, 2009) such as supporting cooperation among teachers to develop new teaching practices. Their power-sharing attitudes are similar to facets of shared instructional leadership (Marks & Printy, 2003) or balkanizing leadership (Urlick & Bowers, 2014). They also display characteristics of dialogue-based leadership (Pedersen & Ryberg, 2009), especially with regard to high collaboration. These traits of Cluster 1 principals are likely to be conducive to higher student achievement (e.g. Kreiner & Mehlbye, 2000). Ringmose and Mehlbye (2004) show that shared leadership between principals and teachers has a positive effect on student outcome. Similarly, cooperation is also important for student success (Hallinger, 2011). A high

degree of “engagement with instruction”, as displayed by Cluster 1 principals is also known to contribute to better student results (Robinson *et al.*, 2009).

Cluster 3 is similar to Cluster 1, except that its principals score lower on some of the positive characteristics discussed above. However, the emphasis they place on mentoring practices is likely to affect their students’ outcomes positively, as these practices increase teacher capacity and ensure steady support for the teachers (Robinson *et al.*, 2009).

The relative lack of shared leadership in Cluster 2, coupled with a less favourable attitude to appraisal and instructional practices, is unlikely to lead to improved student achievement (Hallinger, 2011; Robinson *et al.*, 2009). Principals in this group have a tendency to make important decisions on their own, which resembles the controlling leadership style (Urlick & Bowers, 2014). Given the relatively negative profile of principals in this group, it is slightly alarming that nearly half of sampled participants fall in this category. Summing up a number of doctoral theses in Finland, Risku and Kanervio (2010) conclude that principals “seem overburdened by their work” and the “working time of the principal is not directed at pedagogical leadership as it should be.” (p. 181). This description concurs well with cluster 2.

Cluster 4 principals seem to have a “transactional” style (Burns, 1978) as they believe in extending rewards and sanctions depending on how well goals have been achieved. This attitude is similar to the strategic leadership style (Pedersen & Ryberg, 2009). Limited power-sharing and high work load may be negatively related to student achievement; however, their attitude towards teacher appraisal may offset the negative effects on student outcomes (Ringmose & Mehlbye, 2004).

Gender differences in group membership are observed only in Cluster 2, which has twice as many male members as female members. In a comparison of female and male principals, a study found that “female principals more often emphasize soft values and negotiation, and want to be easily accessed” (Juusenaho, 2004, as cited in Risku & Kanervio, 2011, p. 180) whereas among male principals “administration and having authority seems to be more important”. The relatively low power sharing in cluster 2 is broadly consistent with this description.

A number of teacher characteristics are associated with leadership styles, and exploring these contributes towards a fuller understanding of a principal's actions and attitudes. Of special interest is the finding that the relatively apathetic principals in Cluster 2, with their seeming neglect of instructional leadership, have teachers reporting slightly lower efficacy in instruction. If corroborated, this relationship has implications for policy makers in the Nordic countries that should encourage principals to take a more active or visible role in instructional matters. Teachers working for principals in Cluster 2 are low on most other variables, with the exception of job satisfaction.

It is also interesting to note that the teachers working for principals who emphasise mentoring (Cluster 3) should express more need for professional development, perhaps in response to the principal's requirement.

It is noteworthy that low job satisfaction and rejection of constructivist beliefs co-occur among teachers working with Cluster 4 principals, whose leadership practices tend to be reactive and who have unfavourable attitudes to sharing power. A change in their leadership style may be warranted. Aydin, Sarier and Uysal (2013) found that moving from a transactional leadership style to a transformational improved job satisfaction and organisational commitment among teachers. Moreover, transactional leaders as identified by Aydin *et al.* (2013) share a trait with Cluster 4 principals in this study: they favour the use of rewards or punishments for teachers to pursue organisational goals.

Teachers working for principals in cluster 4 however have a relatively high efficacy in student engagement, low levels of class disruption and greater satisfaction with their own performance. It would be worth exploring further whether these characteristics are related to the high stakes appraisal environment adhered to by their principal. Similarly, teachers working with Cluster 4 principals, who strongly favour results and rewards, are least receptive to the constructivist approach to teaching and learning. It is likely that a school environment focused on achievement and eventual sanctions does not allow for the uncertainty inherent to the constructivist philosophy.

Overall, the findings reveal a broad range of leadership attitudes and practices and associations between these leadership styles and

constructs such as efficacy in instruction, job satisfaction, professional collaboration and expressed need for professional development. Similarities can be found between the clusters identified here and dimensions of leadership identified in previous research. However, the lack of shared items between this study and a number of others cited in the introduction limits the possibility for direct comparisons. The finding that teachers' efficacy in instruction is lower among those working with Cluster 2 principals is interesting as this cluster does not display most of the characteristics that have been previously linked to student success (Hallinger, 2011; Robinson *et al.*, 2009). Further, none of the four groups of principals score highly on *all* the characteristics identified as useful for student achievement: this finding has important policy implications, as it points to room for improvement in principal attitudes and practices, across Nordic countries.

A principal's awareness of where s/he stands vis-à-vis these four clusters would be useful for self-awareness and improvement. This is especially true for principals who find that they do not exercise much instructional leadership or promote collaboration among teachers. In future, the analysis presented here could form the basis of a feedback/ appraisal tool for principals.

The attitudes and practices of the principals are probably more closely tied to background characteristics than is apparent. A more comprehensive overview of correlates with membership in the four clusters is warranted to obtain a deeper understanding of the four leadership groups, the environments they operate in and the teachers that work for them.

Since all four leadership styles are present in all the Nordic countries, *countries* may not be the most interesting unit of analysis for comparisons between leadership practices. Using country averages may confound interesting differences in leadership styles existing within a country. The interaction between leadership style, country and other background variables needs to be explored.

The clusters identified here provide a relatively coherent picture of different sets of principals' practices and attitudes, albeit with gaps. Efforts should be made to bridge these gaps by exploring at scale and item level other TALIS data on principals and teachers and teacher or principal data from other large-scale educational assessments such as PISA.

3.10 References

- Abrahamsen, M. (2008). *Ledelse til en forandring: En undersøgelse af ledelse og ledelsesmuligheder på to gymnasier*. [Leadership for change: A study of management and leadership opportunities at two high schools]. Syddansk Universitet. Det Humanistiske Fakultet.
- Årlestig, H., & Johansson, O. (2011). Research on principals in Sweden. In: O. Johansson (ed.) *Rektor – en Forskningsöversikt 2000-2010*. Stockholm: Vetenskapsrådet.
- Aydin, A., Sarier, Y., & Uysal, S. (2013). The effect of school principals' leadership styles on teachers' organizational commitment and job satisfaction. *Educational Sciences: Theory and Practice*, 13(2), 806–811.
- Björnsdóttir, A., Hansen, B., & Jóhannsson, Ó. H. (2006). Mótun skólustarfs – Hver er hlutur kennara? [The role of teachers in the management of schools]. *Tímarit um menntarannsóknir*, 3, 12–26.
- Bogler, R. (2002). Two profiles of school teachers: A discriminant analysis of job satisfaction. *Teaching and Teacher Education*, 18(6), 665–673.
[http://dx.doi.org/10.1016/S0742-051X\(02\)00026-4](http://dx.doi.org/10.1016/S0742-051X(02)00026-4)
- Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. London: The Guilford Press.
- Burns, J. M. (1978). *Leadership*. New York, NY: Harper and Row.
- Camburn, E. M., Huff, J. T., Goldring, E. B., & May, H. (2010). Assessing the validity of an annual survey for measuring principal leadership practice. *The Elementary School Journal*, 111(2), 314–335. <http://dx.doi.org/10.1086/656302>
- Goldring, E., Huff, J., May, H., & Camburn, E. (2008). School context and individual characteristics: What influences principal practice? *Journal of Educational Administration*, 46(3), 332–350. <http://dx.doi.org/10.1108/09578230810869275>
- Hallinger, P. (2011). Leadership for learning: lessons from 40 years of empirical research. *Journal of Educational Administration*, 49(2), 125–142.
<http://dx.doi.org/10.1108/09578231111116699>
- Hallinger, P., & Heck, R. H. (2010). Collaborative leadership and school improvement: Understanding the impact on school capacity and student learning. *School Leadership and Management*, 30(2), 95–110. <http://dx.doi.org/10.1080/13632431003663214>
- Hansen, B. (2013). Forysta og skólustarf [Educational Leadership]. In: Rúnar Sigþórsson, Rósa Eggertsdóttir and Guðmundur Heiðar Frímansson (eds.), *Fagmennska í skólustarfi: Skrifað til heiðurs Trausta Þorsteinssyni*. [Professionalism in education: Written in honor of Trausti Þorsteinsson]. Reykjavík: Háskólinn á Akureyri and Háskólaútgáfan.
- Hansen, B. Jóhannsson, Ó. H., & Lárusdóttir, S. H. (2008). Breytingar á hlutverki skólustjóra í grunnskólum: Kröfur, mótsagnir og togstreita. [Changes in the role of principals in compulsory schools in Iceland: Constraints, paradoxes and tensions]. *Uppeldi og menntun*, 17(2), 87–104.

- Helgøy, I., & Homme, A. (2004). *Governance in primary and lower secondary education. Comparing Norway, Sweden and England*. Working Paper no. 16–2004. Stein Rokkan Centre for Social Studies. Unifob AS.
- Hjartardóttir, H. (2002). *Hlutverk skólustjóra: Fagleg forysta – falinn fjársjóður?* [The role of the principal. Professional leadership – a hidden treasure?]. Unpublished M.Ed.-thesis. Reykjavík: Kennaraháskóli Íslands.
- Kreiner, S., & Mehlbye, J. (2000). *Arbejds miljøet i folkeskolen*. [The work environment in compulsory schools]. Copenhagen: AKF
- Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). *Investigating the links to improved student learning. Final report of research findings*. St. Paul: The University of Minnesota.
- Ludvigsson, A. (2009). *Samproducerat ledarskap: Hur rektorer och lärare formar ledarskap i skolans vardagsarbete*. [Co-produced leadership: The formation of leadership between school heads and teachers in everyday educational practice]. Doctoral thesis. Jönköping: Jönköping University.
- Marks, H. M., & Printy, S. M. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly*, 39(3), 370–397. <http://dx.doi.org/10.1177/0013161X03253412>
- Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora: Mid-continent Research for Education and Learning.
- Mattar, D. (2012). Instructional leadership in Lebanese public schools. *Educational Management Administration and Leadership*, 40(4), 509–531. <http://dx.doi.org/10.1177/1741143212438222>
- Moos, L. (2011). Research on principals in Denmark. In: O. Johansson (ed.) *Rektor – en Forskningsöversikt 2000-2010*. Stockholm: Vetenskapsrådet.
- Moos, L., Møller, J., & Johansson, O. (2004). A Scandinavian perspective on educational leadership. *Educational Forum*, 68(3), 200–210. <http://dx.doi.org/10.1080/00131720408984632>
- Møller, J. (2011). Research on Principals in Norway. In: O. Johansson (ed.) *Rektor – en Forskningsöversikt 2000-2010*. Stockholm: Vetenskapsrådet.
- Møller, J., Eggen, A., Fuglestad, O. L., Langfeldt, G., Presthus, A. M., Skrøvset, S., Stjernstrøm, E., & Vedøy, G. (2005). Successful school leadership: The Norwegian case. *Journal of Educational Administration*, 43(6), 584–594. <http://dx.doi.org/10.1108/09578230510625683>
- Nordhaug, O. A. (2014). *PISA 2012 og jakten på den effektive rektoren – En analyse av ledelseskonstrukturer og deres sammenheng med elevenes resultater i Norge og Sve- rige*. [PISA 2012 and the quest for the effective headmaster – An analysis of leadership constructs and their relationship to student performances in Norway and Sweden]. Master thesis, Oslo: University of Oslo.
- Nordzell, A. (2007). *Samtalat skolledarskap: Kategoriserings – och identitetsarbete i interaktion*. [Doing school leadership: Categorization and Identity Work in Interaction]. Linköping: Linköping University.

- OECD (2014). *TALIS 2013 technical report*. Retrieved from www.oecd.org/edu/school/TALIS-technical-report-2013.pdf.
- Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning*, 64(3), 493–525. <http://dx.doi.org/10.1111/lang.12065>
- Pedersen, D., & Ryberg, M. (2009). *Fagplanlægning, organisering og ledelse efter gymnasiereformer*. [Subject planning, organization and leadership following gymnasium school reforms] Copenhagen: Ministry of Education. Retrieved from http://www.gymnasieforskning.dk/wp-content/uploads/2013/07/091203_rapport_faglig_planlaegning_organisering_ledelse_pdf_fra_UVM.pdf
- Ringmose, C., & Mehlbye, J. (2004). *Elementer i god skolepraksis: De gode eksempler* [Elements in good school practice: The good examples] Copenhagen: AKF Forlag.
- Risku, M., & Kanervio, P. (2011). Research on principals in Finland. In: O. Johansson (ed.) *Rektor – en Forskningsöversikt 2000-2010*. Stockholm: Vetenskapsrådet.
- Robinson, V., Hohepa, M., & Lloyd, C. (2009). *School leadership and student outcomes: Identifying what works and why*. Best Evidence Synthesis Iteration. New Zealand: The Ministry of Education.
- Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly*, 44, 635–674. <http://dx.doi.org/10.1177/0013161X08321509>
- Shatzer, R. P., Caldarella, P., Hallam, P. R., & Brown, B. L. (2014). Comparing the effects of instructional and transformational leadership on student achievement: Implications for practice. *Educational Management Administration and Leadership*, 42(4), 445–459. <http://dx.doi.org/10.1177/1741143213502192>
- Spillane, J. P. (2005). Distributed leadership. *The Educational Forum*, 69(2), 143–150. <http://dx.doi.org/10.1080/00131720508984678>
- Urick, A., & Bowers, A. J. (2014). What are the different types of principals across the United States? A latent class analysis of principal perception of leadership. *Educational Administration Quarterly*, 50(1), 96–134. <http://dx.doi.org/10.1177/0013161X13489019>
- West, S.G., Taylor, A.B. and Wu, W. (2012). Model Fit and Model Selection in Structural Equation Modeling. In Rick H. Hoyle (ed.), *Handbook of Structural Equation Modeling*, 209–231. London: The Guilford Press.

3.11 Appendices

3.11.1 Appendix 1

Scales and items employed in comparisons between teachers working for principals from each of the four leadership groups.

Table 9: Self-efficacy

Scale	Question number	Item wording
	Q34	In your teaching, to what extent can you do the following?
Efficacy in classroom management	d	Control disruptive behaviour in the classroom
	f	Make my expectations about student behaviour clear
	h	Get students to follow classroom rules
	i	Calm a student who is disruptive or noisy
Efficacy in instruction	c	Craft good questions for my students
	j	Use a variety of assessment strategies
	k	Provide an alternative explanation for example when students are confused
	l	Implement alternative instructional strategies in my classroom
Efficacy in student engagement	a	Get students to believe they can do well in school work
	b	Help my students value learning
	e	Motivate students who show low interest in school work
	g	Help students think critically

Note: 1 = not at all; 4 = a lot.

Table 10: Job satisfaction

Scale	Question number	Item wording
	Q46	[Finally,] we would like to know how you generally feel about your job. How strongly do you agree or disagree with the following statements?
Satisfaction with current work environment	c†	I would like to change to another school if that were possible
	e	I enjoy working at this school
	g	I would recommend my school as a good place to work
	j	All in all, I am satisfied with my job
	i*	I am satisfied with my performance in this school
Satisfaction with profession	a	The advantages of being a teacher clearly outweigh the disadvantages
	b	If I could decide again, I would still choose to work as a teacher
	d†	I regret that I decided to become a teacher
	ff	I wonder whether it would have been better to choose another profession
	h*	I think that the teaching profession is valued in society

Note: † item was reverse coded.

* item was not included in scale but used as single item.

1 = strongly disagree; 4 = strongly agree.

Table 11: Teacher-student relations

Scale	Question number	Item wording
	Q45	How strongly do you agree or disagree with the following statements about what happens in this school?
Teacher-student relations	a	In this school, teachers and students usually get on well with each other
	b	Most teachers in this school believe that the students' well-being is important
	c	Most teachers in this school are interested in what students have to say
	d	If a student from this school needs extra assistance, the school provides it

Note: 1 = strongly disagree; 4 = strongly agree.

Table 12: Classroom disciplinary climate

Scale	Question number	Item wording
	Q41	How strongly do you agree or disagree with the following statements about this <target class>?
Classroom disciplinary climate	a†	When the lesson begins, I have to wait quite a long time for students to quiet down
	b	Students in this class take care to create a pleasant learning atmosphere
	c†	I lose quite a lot of time because of students interrupting the lesson
	d†	There is much disruptive noise in the classroom

Note: † item was reverse coded.
1 = strongly disagree; 4 = strongly agree.

Table 13: Teacher cooperation

Scale	Question number	Item wording
	Q33	On average, how often do you do the following in this school?
Exchange and coordination for teaching	d	Exchange teaching materials with colleagues
	e	Engage in discussions about the learning development of specific students
	f	Work with other teachers in my school to ensure common standards in evaluations for assessing student progress
	g	Attend team conferences
Professional Collaboration	a	Teach jointly as a team in the same class
	b	Observe other teachers' classes and provide feedback
	c	Engage in joint activities across different classes and age groups (e.g. projects).
	h	Take part in collaborative professional learning

Note: 1 = never; 6 = once a week or more.

Table 14: Stakeholder participation

Scale	Question number	Item wording
	Q44	How strongly do you agree or disagree with these statements as applied to this school?
Single items from the Participation among stakeholders scale	a*	This school provides staff with opportunities to actively participate in school decisions
	b*	This school provides parents or guardians with opportunities to actively participate in school decisions
	c*	This school provides students with opportunities to actively participate in school decisions

Note: * item was not included in scale but used as single item.
1 = strongly disagree; 4 = strongly agree.

Table 15: Effective professional development

Scale	Question number	Item wording
	Q25	Considering the professional development activities you took part in during the last 12 months, to what extent have they included the following?
Effective professional development	a	A group of colleagues from my school or subject group
	b	Opportunities for active learning methods (not only listening to a lecturer)
	c	Collaborative learning activities or research with other teachers
	d	An extended time-period (several occasions spread out over several weeks or months)

Note: 1 = not in any activities; 4 = yes, in all activities.

Table 16: Need for professional development

Scale	Question number	Item wording
	Q26	For each of the areas listed below, please indicate the degree to which you currently need professional development
Need for professional development in subject matter and pedagogy	a	Knowledge and understanding of my subject field(s)
	b	Pedagogical competencies in teaching my subject field(s)
	c	Knowledge of the curriculum
	d	Student evaluation and assessment practice
	f	Student behaviour and classroom management
Need for professional development for teaching for diversity	h	Approaches to individualised learning
	i	Teaching students with special needs (see Question [9] for definition)
	j	Teaching in a multicultural or multilingual setting
	k	Teaching cross-curricular skills (e.g. problem solving, learning-to-learn)
	l	Approaches to developing cross-occupational competencies for future work or future studies
	n	Student career guidance and counselling

Note: 1 = no need at present; 4 = high level of need.

Table 17: Constructivism

Scale	Question number	Item wording
	Q32	We would like to ask about your personal beliefs on teaching and learning. Please indicate how strongly you agree or disagree with each of the following statements
Constructivist beliefs	a	My role as a teacher is to facilitate students' own inquiry
	b	Students learn best by finding solutions to problems on their own
	c	Students should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved
	d	Thinking and reasoning processes are more important than specific curriculum content

Note: 1= strongly disagree; 4 = strongly agree.

3.11.2 Appendix 2

Table 18: ANOVAs: Scale comparisons of teachers working for principals from each of the four leadership groups

Scale	df	F	p	Differences between groups
Efficacy in classroom management	(3,8346)	1.595;	.188	2,4,1,3
Efficacy in instruction	(3,8346)	6.139;	< 0.001	2 < 3
Efficacy in student engagement	(3,8348)	3.774;	0.01	3,2 < 4
Satisfaction with current work environment	(3,8323)	8.958	< 0.001	4 < 2,3
Satisfaction with profession	(3,8323)	39.923	< 0.001	4 < 3,1,2
I think that the teaching profession is valued in society (Q46h)	(3,8303)	123.910	< 0.001	4 < 3 < 1,2
I am satisfied with my performance in this school (Q46i)	(3,8312)	4.501	0.004	2 < 4
Teacher-student relations	(3,8323)	2.664	0.046	4,3,2,1
Classroom disciplinary climate	(3,8805)	6.346	< 0.001	1 < 3,4
Exchange and coordination for teaching	(3,8357)	12.861	< 0.001	3 < 1 < 4
Professional Collaboration	(3,8355)	5.573	< 0.001	2 < 3,4
This school provides staff with opportunities to actively participate in school decisions (Q44A)	(3,8259)	9.658	< 0.001	4,2 < 3; 4 < 1
This school provides parents or guardians with opportunities to actively participate in school decisions (Q44B)	(3,8174)	10.345	< 0.001	2,4 < 3; 2 < 1
This school provides students with opportunities to actively participate in school decisions (Q44C)	(3,8208)	8.076	< 0.001	2 < 4,3
Effective professional development	(3,6824)	0.745	0.525	2,1,3,4

Scale	df	F	p	Differences between groups
Need for professional development in subject matter and pedagogy	(3,8433)	22.670	< 0.001	2,1 < 4 < 3
Need for professional development for teaching for diversity	(3,8412)	18.812	< 0.001	4,2,1 < 3
Constructivist beliefs	(3,8383)	33.959	< 0.001	4 < 2,3,1

4. Teachers' professional development in Nordic countries

By Matti Taajamo, University of Jyväskylä, Finland

4.1 Summary

This article explores the current state of teachers' professional development in Nordic countries by examining how it is prioritised in these countries and the associations between professional development, job satisfaction and self-efficacy. Here, professional development is seen as a wide spectrum of activities aimed at developing teachers' skills, knowledge, working methods and competencies. Job satisfaction refers to satisfaction with the teaching profession, and self-efficacy refers to teachers' beliefs in their ability to attain their work goals. The article investigates these constructs from a Nordic perspective and on the basis of data obtained from TALIS 2013.

TALIS data showed that while the Nordic countries offer plenty of opportunities to participate in professional development activities such as courses, workshops, seminars and mentoring programmes, the time spent by teachers on these activities is minimal. Teachers' average self-efficacy is high in Denmark and Sweden, yet the job satisfaction of Swedish teachers is lower than that of their Nordic colleagues. The present analysis showed that engaging in professional development activities is associated with both job satisfaction and self-efficacy. Serving as a mentor predicts teachers' self-efficacy in Denmark, Finland and Norway, whereas participation

in a formal induction programme predicts self-efficacy in Iceland and Norway. In Sweden, having a mentor and serving as a mentor are positively associated with job satisfaction. A formal induction seems to be more related to self-efficacy, and an informal induction more to job satisfaction among the teachers. The results showed that teachers find participation in mentoring particularly useful.

In-service and continuing training are rather fragmented in the Nordic countries and offered by many different organisations. There is inadequate integration of the basic and in-service training of teachers and a lack of a career-long trajectory in professional development. A particular concern is that the level of participation in formal, informal and general induction activities is very low. On average, less than half the schools in Nordic countries offer an induction programme for new teachers. Mentoring is well organised in Iceland, moderately well in Denmark, Norway and Sweden, but under development in Finland.

According to TALIS data, the most critical professional development needs in the Nordic countries are a better understanding of how to use ICT in teaching, how to teach students with special needs and how to adopt new technologies in schools. The main obstacles to participating in professional development activities are work schedules and lack of incentives. The needs and focus of in-service training and other professional development initiatives vary according to the stages of a teacher's career. It is, therefore, important to map educational needs and ensure that teachers' different needs and challenges are factored into professional development. It is also important to re-think the forms of professional development, perhaps, by replacing or augmenting traditional training sessions outside the school with work-related development programmes.

4.2 Introduction

Studies on the teaching profession show that teaching is a complex activity and that teachers use a broad repertoire of methods to improve student learning (e.g. Danielson, 2007; Hegarty, 2000; Timperley, 2011). Policy-level literature explains that teachers face continuous changes in

their work. New teaching paradigms, conceptions of learning and societal changes have transformed the teacher's role from that of a knowledge provider to a guider of students' learning. Today, flexibility, innovativeness, cooperation and planning skills are often listed as qualities to be possessed by a good teacher (e.g. Ferrari, Cachia, & Punie, 2009). Teachers are expected to reflect on and assess their own actions. They are also required to examine and develop their work with their colleagues, students and collaborators outside school. These challenges have birthed a new global movement aimed at improving teacher education and teachers' professional learning and development (e.g. Timperley, 2011; Wei, Darling-Hammon, Andree, Richardson, & Orphanos, 2009).

The Nordic countries are currently witnessing reforms in the domains of teacher education and in-service training. This article offers an insight into how professional development is organised in the Nordic countries. Using data from TALIS 2013, this article describes the current status of continuing professional development (CPD) among teachers, especially focusing on induction, mentoring and career-long professional development. The findings are relevant to lower secondary education [ISCED level 2] as all the Nordic countries participated at this level. The article concludes by discussing some policy implications related to CPD.

4.3 Teachers' continuing professional development

Teacher education as such cannot be expected to prepare teachers for all the challenges they are likely to face during their employment as a teacher. In research literature, challenges encountered by new teachers are often termed "praxis shocks" (see Anagnostopoulos, Smith, & Basmadjian, 2007; Goddard & Foster, 2001). The European Commission Handbook for Policymakers (Snoeck, Eisenschmidt, Holdsworth, Michaelidou, Dahl & Pachler, 2010, p. 13–16) makes the following observation with regard to preparing teachers: "Effective induction programs can avoid some of these problems ('praxis-shock' by newly-qualified teachers and consequent early drop-out from the profession) by providing all new teachers with systematic personal, social and professional support in the early years of their career. They can therefore also help improve

school and teacher performance. Induction provides a vital link in the continuum of teacher education that runs from Initial Teacher Education through induction to career-long continuing professional development.” Recent studies suggest that induction programmes such as peer-group mentoring can successfully serve these purposes (Geeraerts *et al.*, 2015).

The TALIS 2013 report defines teachers’ professional development as “activities that aim to advance teachers’ skills and knowledge, with the ultimate aim of improving their teaching practice” (OECD, 2014, p. 85). In each Nordic country, the professional development of teachers is organised slightly differently both in terms of the organisers of education and resources. However, the aims are identical: continuing professional development of teachers is expected to be reflected in students’ academic outcomes and also in the teachers’ job satisfaction (e.g. Darling-Hammond, 2000).

Different definitions can be found for professional development of teachers in the literature (e.g. De Vries, Van de Grift, & Jansen, 2014; Harland & Kinder, 2014; Hökkä & Vähäsantanen, 2014). At the core of these definitions is the view that professional development is about teachers’ learning processes: learning how to learn and transforming their knowledge into practices that benefit their students’ growth. Professional development is broadly reviewed in this article. It has been defined as all the activities that aim at the development of teachers’ skills, knowledge, expertise, working methods, networks and other elements of teacher competence. These development activities can be offered in many modes, ranging from formal courses, workshops, seminars, etc. to more informal approaches such as collaboration with other teachers and participation in extracurricular activities (Avalos, 2011; TALIS Framework, 2013; OECD, 2014a.).

During teacher education, prospective teachers develop their practical and theoretical knowledge base in the subject matter and at the same acquire both general pedagogical knowledge and pedagogical content knowledge. De Vries *et al.* (2014) explain that after the initial teacher education, while practical knowledge expands through experience gained during teaching practice, the theoretical knowledge base requires constant upgrades to respond to continuing societal and educational development. Other experts have observed that theory and practice are not

separate entities; instead, the development of high-level professional expertise requires the integration of theoretical and practical knowledge and reflection on practice with the help of conceptual knowledge (e.g. Tynjälä, 2008; Tynjälä & Gijbels, 2012). Little empirical research exists on the theory-practice relationship, although some recent studies in the field of vocational education and training lend support to the view emphasising the integration of theory and practice. For example, Virtanen, Tynjälä and Eteläpelto (2014) found that integration between school learning and workplace learning is the second most important factor explaining students' learning outcomes in the workplace. Similarly, in teacher education and teacher development, integrative pedagogy has produced promising results (e.g. Heikkinen, Jokinen, & Tynjälä, 2012).

Teaching and learning quality are the primary factors affecting student performance (OECD, 2012). Given the alarmingly high attrition rates in the teaching profession (Tynjälä & Heikkinen, 2011), strategies with a strong resource focus are required to retain and train teachers (see European Commission, 2012a). It is globally recognised that teachers' continuous professional development should start with their initial teacher education, continue with an induction phase and thereafter as in-service training throughout the teaching career (e.g. Geeraerts *et al.*, 2015). A well-organised continuum creates the foundation for well-planned career development. It is built on the mutual interests of the teachers and the organisers of education. It has been emphasised that continuous professional development is an integral part of teachers' work. Thus, professional development and learning at work are inseparable from ongoing participation and everyday work practices. Work should provide opportunities to engage in multiple communities, both within and beyond organisational boundaries (Hökkä & Vähäsantanen, 2014; Timperley, 2008).

Working in an information-driven society and a fast changing world presents many learning challenges for teachers. For example, they have to learn how to use the latest technology, how to develop their own teaching skills, how to account for student-specific differences and how to cater to the needs of the weakest students (European Commission, 2012a). Despite these challenges, teachers' engagement in CPD differs greatly. Reasons for

this range from personal beliefs about teaching and learning to the resources available for organising CPD (De Vries *et al.*, 2014).

Policy directives and vast amounts of funds are being directed towards professional development for teachers in the hope that it will improve student learning. According to Timperley (2011), much of this investment has failed to meet its goals because professional development has held little meaning for teachers. She says, “Gains for students are the result of some fundamental shifts in thinking about professional development, leadership and classroom practice” (Timperley, 2011, p. 3). These shifts refer to moving from professional development to professional learning, and it includes engaging in systematic inquiry into the effectiveness of practice. To this end, the present article examines empirical findings on the relationships between the concepts of professional development, job satisfaction and self-efficacy. These concepts are broad but in TALIS 2013, they are understood as follows.

Professional development is examined from the perspectives of induction and mentoring programmes, as well as from the perspective of career-long professional development. Teachers’ job satisfaction consists of satisfaction with the profession and satisfaction with the current work environment (Crossman & Harris, 2006). A major drawback of the earlier research on teacher job satisfaction is the lack of consensus on measuring the concept – is it to be facet-specific or does it refer to an overall sense of satisfaction with the job (Skaalvik & Skaalvik, 2011)? Consequently, both aspects have been included in the TALIS. Self-efficacy in this context refers to teachers’ beliefs in their own ability to attain their work goals. Previous studies have suggested that teachers’ sense of self-efficacy is an important factor influencing students’ academic outcomes as well as teachers’ job satisfaction (Klassen & Chiu, 2010). Unfortunately, the concept of teacher efficacy has also suffered from poor construct validity (Henson, 2001).

Relatively little is known about the linkages between teachers’ professional development, job satisfaction and self-efficacy (Caprara, Barbaranelli, Steca, & Malone, 2006). Here, teachers’ professional development, job satisfaction and self-efficacy are defined from the Nordic perspective and explored on the basis of data from TALIS 2013, mainly at

the ISCED 2 level. More specifically, answers are sought to the following questions:

- What kinds of national systems for professional development exist in the Nordic countries?
- How is professional development organised and prioritised in the Nordic countries?
- How does professional development impact job satisfaction and self-efficacy?

4.4 Teachers' professional development in the Nordic countries

The sections below present an overview of professional development of teachers as it has been organised and secured in the Nordic countries. Information has been mainly obtained from the Eurydice network (http://eacea.ec.europa.eu/education/eurydice/index_en.php) and updated with the help of an expert from each country. It is important to note that the descriptions are not wholly comparable because different issues are emphasised for each country. However, they do provide an insight into the challenges that the Nordic countries face regarding professional development of teachers.

In most Nordic countries, with some exceptions, the minimum qualification for a teacher is a master's degree. However, the countries differ in terms of their educational systems and concepts, while also sharing certain basic values such as equality and flexibility.

4.4.1 *Teacher development in Denmark*

Several diploma degrees and master's programmes are offered in Denmark. These are targeted at teachers in the municipal primary and lower secondary, upper-secondary and adult education sectors. The programmes are paid, part time and conducted at university colleges of education or universities proper. The head of the school and the teacher

jointly develop a continuing professional development plan. The plan is a tool for prioritising and maintaining the focus on education, with its ultimate aim as improving the quality of teaching. Danish teachers are also free to participate in in-service training activities offered in the forms of courses, diploma degrees and other types of competence-rewarding programmes. Many municipalities, and sometimes schools, also have their own courses for teachers. Generally speaking, there is no legislation governing the in-service training of teachers in Denmark (European Commission, 2012b.).

In-service training activities are provided both at the regional and national levels. For teachers in municipal primary and lower secondary schools, it is primarily organised by the Danish School of Education and by university colleges of education. Specialised state training institutions, county resource centres, teachers' associations and the Ministry of Education also offer in-service training. The Danish Union of Teachers contributes to the in-service training of municipal primary and lower secondary school teachers. Recently, traditional teaching is being replaced by new pedagogical methods such as interactive teaching, study visits abroad and projects carried out in cooperation with local businesses. In-service training now constitutes an integral part of a strategy to enhance both the competencies of the individual teacher and the general profile of the schools concerned (European Commission, 2012b.).

4.4.2 *Teacher development in Finland*

In Finland, continuing teacher education and training is not subject to any specific legislation. The obligation to participate in in-service training is partly defined in various statutes and partly in collective agreements. Teachers are required to participate in in-service training for any period from one to five days a year, although employers decide which training programmes and forms of education conform to the collective agreement, and thus are acceptable as in-service training. Teachers themselves are also given more autonomy for developing their professional skills and expertise. More attention is being paid to self-motivated continuing education (European Commission, 2013.).

Continuing teacher education is organised by different types of training centres such as university continuing education units, vocational teacher education colleges, university departments of teacher education, teacher training schools, summer universities and various private organisations. Continuing education and in-service training have been divided into (1) self-motivated continuing teacher education and (2) employer-driven participation in education during working hours (European Commission, 2013.).

A national, fixed-term programme (OSAAVA) was set up in 2010 by the Ministry of Education and Culture and is due to end in 2016. The objective of OSAAVA is to enable educational institutions to develop the competencies of their staff (Hämäläinen & Hämäläinen, 2011). Accordingly, the programme supports education providers in offering continuing professional development to their education personnel and ensuring staff opportunities to improve their professional competence.

4.4.3 Teacher development in Iceland

Iceland, too, has no comprehensive legislation for the professional development of teachers; however, the country does emphasise on ensuring that teachers have access to continuous professional development, in-service education or further education. The organisation of professional development of educational personnel varies across the three school levels. Teacher education institutions usually offer such programmes in collaboration with the Ministry of Education, local educational offices, individual schools or subject associations of teachers (European Commission, 2015.).

Educational authorities at both the state and municipal level have adopted the policy of encouraging teachers to introduce innovations and initiate developmental projects and in-service projects on their own. CDP has thus remained optional for teachers in Iceland but most teachers' contracts of employment state that they are expected to spend time on in-service training, preparation and other duties outside of school (European Commission, 2015).

Since its establishment in 2013, the Council of Continuous Professional Development of Teachers has focused on policy making, i.e. mapping and analysing the field of professional development in Iceland. The

council also plans to publish a general CPD framework for teachers and principals in Iceland in the spring of 2016 and present a proposal to the Minister of Education, Science and Culture.

4.4.4 *Teacher development in Norway*

In Norway, university colleges of education offer teacher education, while other higher education institutions provide continuing and in-service training. The school administrator have a system for providing teaching personnel, school leaders and personnel with special responsibilities in the school system with opportunities for competence enhancement, with a view to refreshing and extending their professional and educational knowledge. Municipal education authorities can then decide whether the staff may participate or not. The tradition of local autonomy is strong in the Norwegian administrative system. National educational authorities are currently discussing ways to avoid differences in participation in in-service programs (European Commission, 2014).

In-service training and continuing education courses are organised at the local, regional and national levels. Organisers consist of local education authorities, teachers' associations, associations for special subjects in higher education institutions, regional officer's educational departments, county education committees, national councils or national education authorities. In addition, separate continuing training programmes are provided for mentors (i.e. those mentoring newly qualified teachers) and newly employed head teachers/principals and school deputies (European Commission, 2014).

4.4.5 *Teacher development in Sweden*

In Sweden, competence development of teachers is a key instrument for achieving equivalence and high-quality teaching standards. It also aims at strengthening the attainment of national and local goals and developing school activities. The organiser of education is obliged to ensure that competence development opportunities are available to the teaching staff. Contract education for this purpose can be purchased from various

providers. Higher education institutions and regional development centres are the primary providers of publicly funded competence development. Other organisers include the public broadcasting service, which produces educational broadcasting, teachers' trade unions, other state authorities inclusive of higher education institutions and independent educational companies. An essential component of the work profile of the Swedish National Agency for Education is skill development of school staff. The agency owns the responsibility for national-level training of head teachers and for the professional development for teachers. (European Commission, 2011).

Although teachers' competence development has mainly been the education provider's responsibility, the state has also launched a number of initiatives in the area. For example, Teacher Boost 1 and 2 are two heavy-investment programmes aimed at teachers' further education. The first of these was launched in 2007 to increase teachers' theoretical and didactical competence in the subjects they teach. The second was introduced in 2012, and is aimed at teachers with teaching degrees but who lack accreditation in some or all the subjects they teach (European Commission, 2011).

In Sweden, all new teachers have a one-year induction period. The induction serves two purposes: it introduces new teachers to the profession and assesses their suitability. During their induction period, new teachers receive support and assistance from a mentor. Posts such as First Teacher and Senior Subject Teacher make the teaching profession more attractive and ensure good teaching for the pupils. At the moment, this reform is being followed up and analysed (European Commission, 2011).

In summary, teachers in the Nordic countries have opportunities to participate in various kinds of developmental activities. In Denmark, several diploma degrees and master programmes are offered. In Norway, university colleges of education also offer teacher education. In Finland and Iceland, teachers have been entrusted with the responsibility of developing their own expertise. In Sweden, the competence development of teachers is seen as a key instrument for attaining high-quality teaching standards.

4.5 Implementations of TALIS

The components of teacher development covered in TALIS include the induction phase, mentoring and different types of professional development activities, such as courses, seminars and observation visits. TALIS defines induction programmes for teachers as a range of structured activities implemented at a school to support entry into the school and/or into the teaching profession for new teachers. TALIS examines both the availability of induction and mentoring programmes as well as teacher participation rates in these programmes. Moreover, the professional development received by teachers, their professional development needs, and the barriers to participation are also measured (OECD, 2014a).

Information about teachers' professional development was taken from the OECD TALIS 2013 cycle (OECD, 2014a, 2014b). Data were collected using paper-based and online questionnaires during the spring term of 2012. A representative sample of 20 teachers from 200 schools was randomly selected for the study from each participating country. Approximately 106,000 lower secondary teachers, representing more than 4 million teachers in 34 participating countries and economies, responded to the survey. A total of 1,649 teachers in Denmark, 2,739 in Finland, 1,430 in Iceland, 2,981 in Norway and 3,319 in Sweden completed the TALIS questionnaires. Their responses were analysed using descriptive statistics and regression analysis. The following sections present the survey results on induction, mentoring, professional development activities and the relationship of these factors to teachers' perceptions of self-efficacy and job satisfaction.

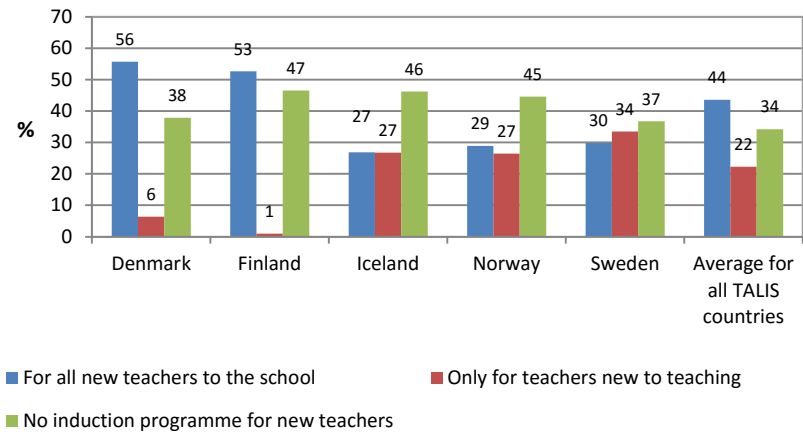
4.6 Results

4.6.1 Teachers' professional development – results from TALIS 2013

The results described here are based on ISCED level 2, as all the Nordic countries participated at this level. Denmark, Finland and Norway also participated in ISCED level 1 and level 3 surveys, while Iceland only participated in ISCED level 3. Data from ISCED level 3 has not been included because of the disparities between the countries in the structure of their upper secondary education and the extent to which they integrate or separate general or academic programmes and vocational programmes. In the TALIS survey, the purpose was not to separate general and vocational schools. In the TALIS data, only 16% of the upper secondary teachers worked in schools that offer vocational programs.

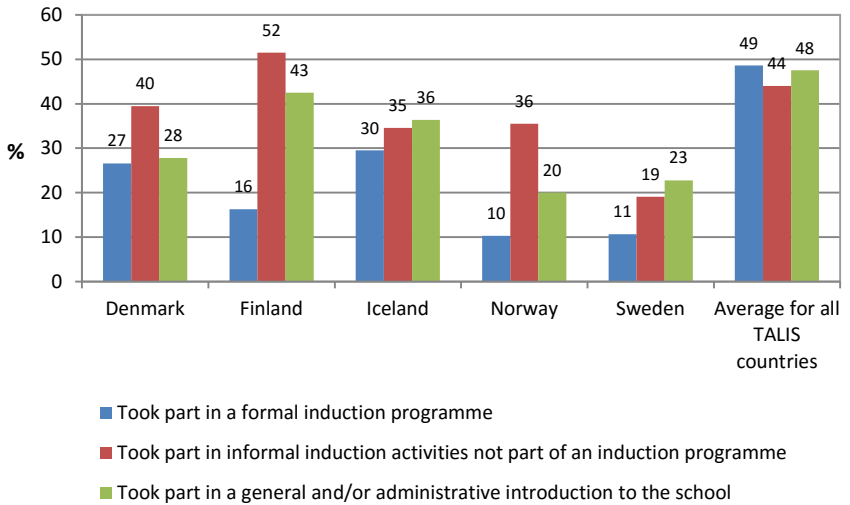
Figure 24 shows the percentages of schools offering induction programmes for new teachers. According to the responses of school principals, on average, less than half the schools in Nordic countries offer an induction programme for new teachers, which is a cause for concern. More than 50% the teachers in Denmark and Finland work in schools that offer an induction programme for all new teachers compared to just 30% in Norway, Sweden and Iceland. In terms of schools offering induction programmes Nordic countries are below other TALIS countries, on average. In fact, almost half the schools in Finland and Iceland do not have a formal induction programme of any kind.

Figure 24: Schools offering formal induction programmes for teachers, as reported by principals



The participation of teachers in formal, informal and general induction activities, as a whole, is low in the Nordic countries (Figure 25). Less than half of the teachers take part in any of these activities. Participation in formal induction activities is less common than informal activities. Further, the differences in teacher participation, especially between Finland and Norway, are considerable. Thus, on the basis of this evidence, I can conclude that teachers’ professional development in the Nordic countries is viewed more as a teacher’s personal responsibility than the duty of the school. This interpretation is in line with Nordic countries’ disparate approaches to organising CPD for teachers, discussed earlier in this article.

Figure 25: Teachers' participation in induction activities



Globally, mentoring is among the most commonly used methods (e.g. Wei *et al.*, 2009) for encouraging new teachers to grow professionally. Figure 26 shows that the use of mentoring in Denmark matches the international TALIS average. Specifically, this means that approximately three out of four principals reported running a mentoring programme at their school. Sweden and Norway have fewer such programmes than Denmark, but Finland has the least of all Nordic countries. In Iceland, over 90% of the teachers work in schools where more experienced teachers support their less experienced colleagues.

Figure 26: Availability of a mentoring system for teachers in their school

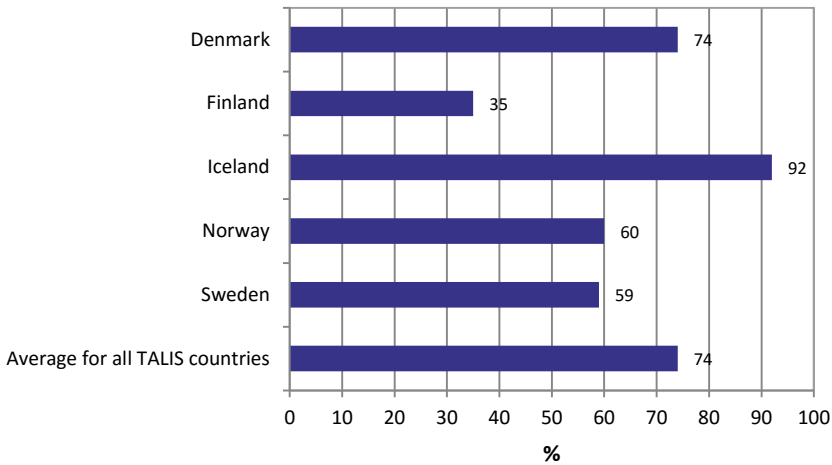
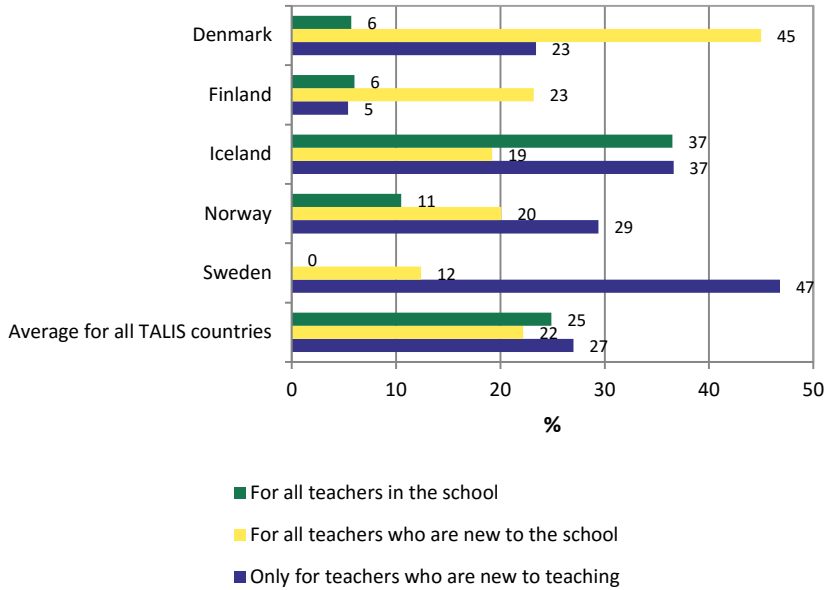


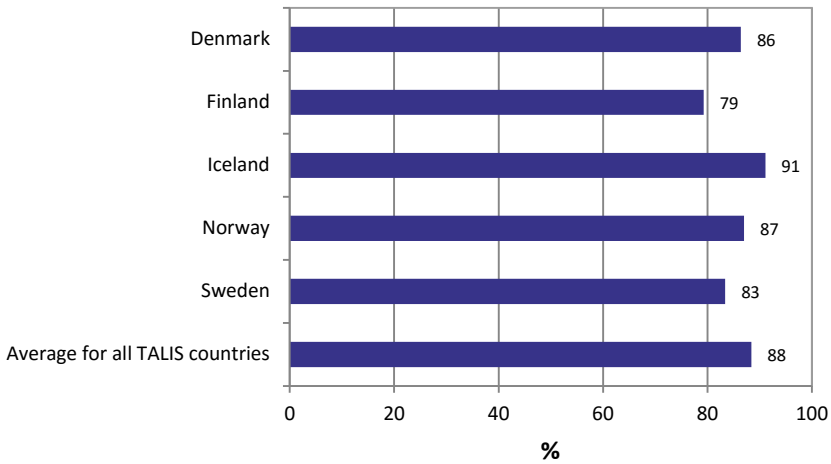
Figure 27 shows which teacher groups receive mentoring in the Nordic countries. In Iceland, Sweden, Norway and Denmark the availability of mentoring is more common than in Finland. In Finland, the responsibility for mentoring rests to a large extent with individual teachers (Jokinen, Taajamo & Välijärvi, 2014). In Sweden, mentoring is provided to 47% of teachers who are new to teaching. In Denmark, mentoring is provided to 45% of all teachers who are new to the school.

Figure 27: Availability of mentoring programmes by teacher groups



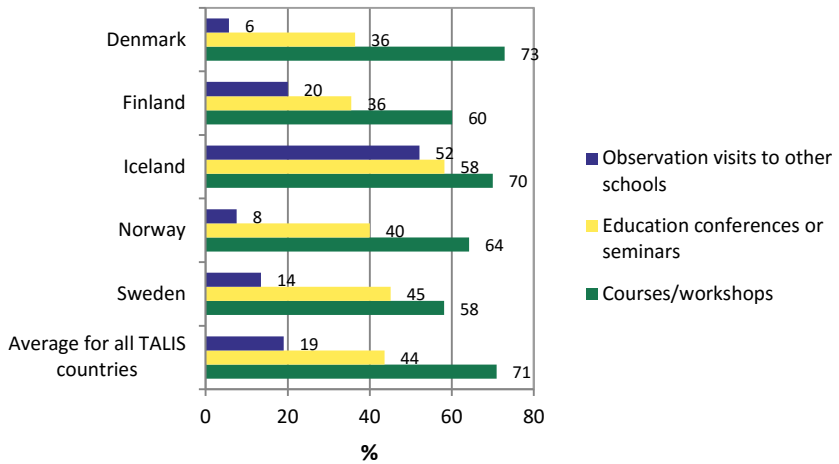
Teachers' overall participation in professional development activities is shown in Figure 28. Across the TALIS countries, 88% of the teachers on average reported participating in activities to develop their professional competence over the previous 12 months. In Finland, this percentage is 79. In Iceland, Sweden, Norway and Denmark, the percentage is slightly higher than in Finland.

Figure 28: Teachers' participation in professional development activities



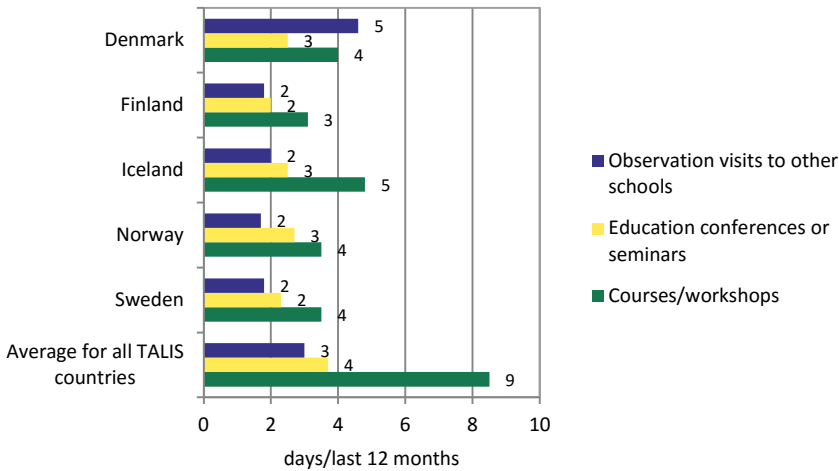
A study of the different forms of professional development in TALIS countries reveals that 71% of the teachers participated in courses or workshops that dealt with subject content or teaching methods. In addition to educational conferences or seminars, courses and workshops are the methods most commonly used to facilitate professional development in the Nordic countries. Observation visits to other schools are not frequently used, except in Iceland, where over 50% of the teachers observed teaching in other schools (Figure 29).

Figure 29: Participation in different types of professional development activities



Although the majority of the respondents from Nordic countries seemed to have participated in developmental activities in the previous 12 months, the total number of participation days is low. Only a few days are spent in observation visits to other schools or in participating in education conferences or seminars. Participation in courses and workshops is slightly more common but below the TALIS average (Figure 30).

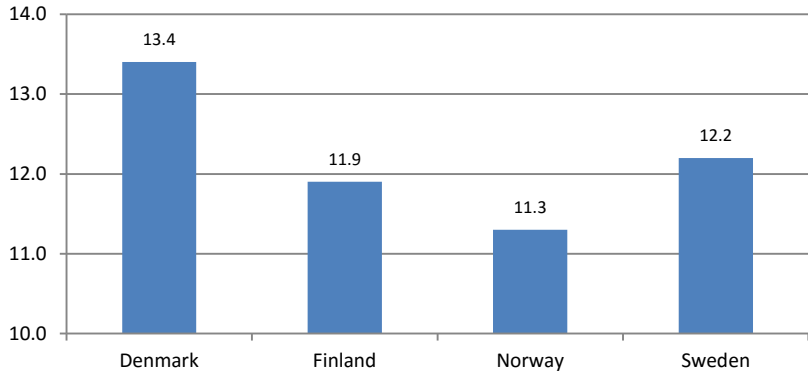
Figure 30: Duration of participation in professional development activities



4.6.2 Professional development activities and teachers' self-efficacy and job satisfaction

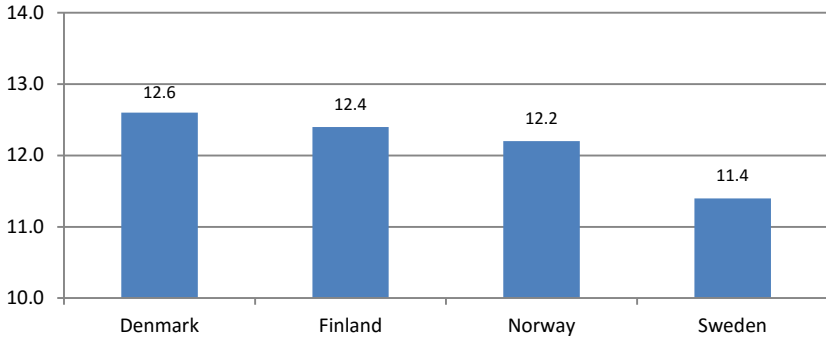
In addition to professional development activities, TALIS focused on teachers' experiences of self-efficacy and job satisfaction. Most teachers in the TALIS countries reported beliefs that suggest high levels of both (OECD, 2014a). In the Nordic countries, mean self-efficacy is highest in Denmark (13.4), followed by Sweden (12.2) (Figure 31). In all four countries mean self-efficacy is significantly higher than the scale midpoint, which is set at 10 (OECD, 2014b, 157). Job satisfaction, on the other hand, is lowest in Sweden (Figure 32). The standard deviation in Sweden is 2.0, the same as the international standard deviation of the scale. The standard deviation in Denmark and Finland is 1.8.

Figure 31: Teachers' self-efficacy in Nordic countries ($p < .05$)



Note: To protect respondents' privacy, Iceland did not authorise the release of its data as part of the International Database (OECD, 2014b).

Figure 32: Teachers' job satisfaction in Nordic countries ($p < .05$)



Note: To protect respondents' privacy, Iceland did not authorise the release of its data as part of the International Database (OECD, 2014b).

To examine the relationship between professional development and self-efficacy, a multiple linear regression has been used. Explanatory factors include different forms of professional development. Table 19 shows the results of the regression model for self-efficacy, and Table 20 shows the results for job satisfaction. In the tables, all six explanatory variables are dichotomous.

Serving as a mentor is a predictor of self-efficacy in Denmark, Finland and Norway (Table 19), whereas in Iceland and Norway self-efficacy is predicted by participation in a formal induction programme. Participation in mentoring and/or peer observation and coaching as part of a formal school arrangement is associated with self-efficacy only in Sweden. In Finland participation in courses, conferences or seminars shows a slight association with self-efficacy.

Table 19: Predictors of teachers' self-efficacy (p<.05)

Teachers' self-efficacy		Denmark	Finland	Iceland	Norway	Sweden
Explanatory factors						
Participation in formal induction programmes	β S.E.			0.24 0.12	0.40 0.20	
Participation in informal induction activities	β S.E.					
Teachers having a mentor	β S.E.					
Teachers serving as mentors	β S.E.	0.26 0.11	0.48 0.20		0.40 0.16	
Participation in mentoring and/or peer observation and coaching, as part of a formal school arrangement	β S.E.					0.38 0.09
Participation in courses/workshops, education conferences or seminars	β S.E.		0.23 0.09			

Participation in informal induction activities explains the variation in teachers' job satisfaction in Finland, Norway and Sweden, whereas participation in a formal induction programme predicts teachers' job satisfaction in Denmark and Norway (Table 20). It is noteworthy that the Swedish teachers who reported having a mentor or serving as a mentor themselves are markedly more satisfied with their job. Participation in mentoring and/or peer observation and coaching as part of a formal school arrangement also has a positive association with Swedish teachers' job satisfaction. Serving as a mentor has a positive association with

job satisfaction in Norway, but this association is somewhat weaker than in Sweden (Table 20).

Table 20: Predictors of teachers' job satisfaction (p<.05)

Teachers' job satisfaction		Denmark	Finland	Iceland	Norway	Sweden
Explanatory factors						
Participation in formal induction programmes	β	0.30			0.39	
	S.E.	0.12			0.15	
Participation in informal induction activities	β		0.24		0.29	0.30
	S.E.		0.10		0.12	0.12
Teachers having a mentor	β					0.73
	S.E.					0.28
Teachers serving as mentors	β				0.37	0.62
	S.E.				0.16	0.22
Participation in mentoring and/or peer observation and coaching as part of a formal school arrangement	β					0.39
	S.E.					0.11
Participation in courses/workshops, education conferences or seminars	β					
	S.E.					

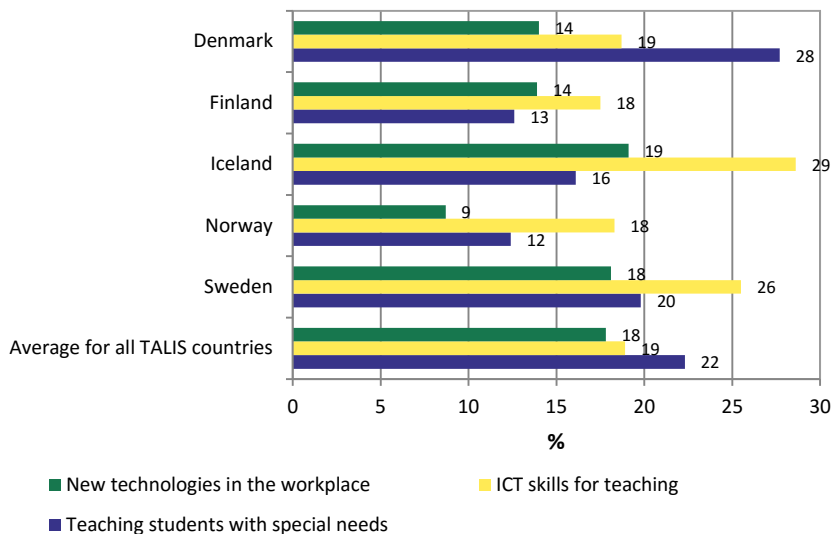
Tables 19 and 20 indicate that self-efficacy and job satisfaction are to some extent explained by different variables. Overall, a formal induction appears to be related more to self-efficacy and an informal induction more to job satisfaction. In Sweden, the factors “teachers having a mentor” and “teachers serving as mentors” very strongly predict job satisfaction, and “participation in mentoring” predicts self-efficacy. In Norway, participation in formal and informal induction programmes predicts both self-efficacy and job satisfaction.

4.6.3 Teachers' professional development needs and barriers

TALIS 2013 also examined teachers' key professional development needs. The response options consisted of (1) knowledge and understanding of subject field, (2) pedagogical competencies in teaching, (3) knowledge of the curriculum, (4) student evaluation and assessment practice, (5) student behaviour and classroom management, (6) approaches to individualised learning, (7) teaching students with special needs, (8) teaching in a multi-cultural or multi-lingual setting, (9) teaching cross-curricular skills, (10) approaches to developing cross-occupational competencies for future work or future studies and (11) student career guidance and counselling. Possible barriers to participation in professional development activities comprised the following response options: (1) do not have the prerequisites (e.g. qualifications, experience, seniority), (2) professional development is too expensive/unaffordable, (3) lack of employer support, (4) conflicting work schedules, (5) do not have time because of family responsibilities, (6) no relevant professional development offered and 7) no incentives for participating in such activities.

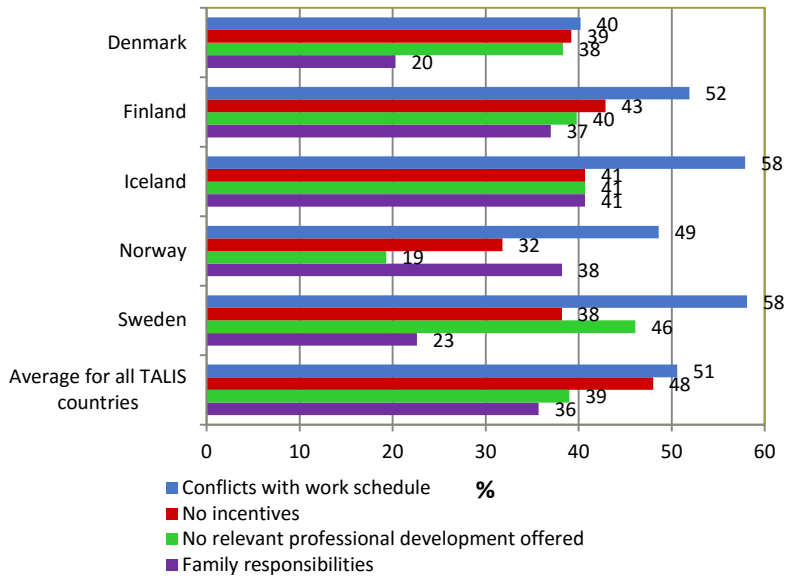
Figure 33 shows the most commonly identified areas of developmental needs. In the TALIS countries, the need for professional development seems particularly acute in the field of special education. Approximately 22% teachers felt that they need more knowledge in this area. The second most common development issue is ICT skills in teaching (19%), followed by the use of new technologies in the workplace (18%). In the Nordic countries, the teachers felt a need for more knowledge in all these areas, but the distribution is slightly different. Iceland, Norway, Sweden and Finland are mainly concerned with ICT skills whereas Denmark teachers seek support for teaching students with special needs.

Figure 33: Teachers' needs for professional development



Teachers in TALIS countries (ISCED level 2) report that the most common obstacles to participation in professional development activities are work schedules (51%) and lack of incentives (48%) (Figure 34). In the Nordic countries, work schedules are the biggest obstacles to participation in continuous professional development. Other obstacles include lack of incentives, lack of relevant professional development offered and lack of time due to family responsibilities. Lack of employer support, not having the prerequisites (e.g. qualifications and experience) and the high cost of professional development programmes also contribute to the obstacles.

Figure 34: Barriers to teachers' participation in professional development activities



4.7 Discussion

Despite the abundance of research literature on teachers' professional development, relatively little is known about its relationship with job satisfaction and self-efficacy (Caprara *et al.*, 2006; Henson, 2001; Skaalvik & Skaalvik, 2011). In this article, these important aspects of a teacher's work are addressed from a Nordic perspective, with the help of data from TALIS 2013. The article examines how professional development is prioritised in the Nordic countries and the kind of impact it has on job satisfaction and self-efficacy.

In the Nordic countries teachers mainly rely on courses, workshops and educational seminars or conferences to develop themselves professionally. The time spent on these activities is minimal – only a couple of days per year. Nevertheless, teachers find such professional development activities beneficial. Moreover, according to the TALIS 2013 results,

these activities are associated with job satisfaction and self-efficacy. Formal induction activities are linked more to teachers' self-efficacy, informal induction activities to teachers' job satisfaction and serving as a mentor predicts teachers' self-efficacy. In Finland and Norway serving as a mentor, and in Sweden participation in mentoring, are associated with self-efficacy. In Sweden, having a mentor and serving as a mentor both have a positive association with job satisfaction. The results provide strong evidence for the benefits of mentoring. While it is impossible to delineate how these factors affect each another, it is good to know that most teachers in the Nordic countries have high levels of self-efficacy and job satisfaction. In light of these findings, there is clearly a need to drive home the importance of professional development activities among these teachers.

In-service training in the Nordic countries seems fragmented and offered by many different organisations. Apart from these, other challenges related to teacher professional development include following: inadequate integration of the basic and in-service training of teachers, lack of a career-long continuum in professional development and lack of coordination between competence development and strategic thinking. These problems impede the cost-effective implementation of in-service training.

The findings of the TALIS study showed that the organisation of teacher development in the Nordic countries meets the average level of the TALIS countries. An exception to this seems to be the organisation of induction practices during the first year of employment, where the Nordic countries' average is lower than that of the TALIS countries. The reason might be that schools do not consider it necessary to offer induction programmes because of the high level of teacher education and most teachers have a master's degree. The TALIS 2013 results also indicate that on average less than a half of the schools in the Nordic countries provide an induction programme for all new teachers. This probably contributes to the low level of participation in formal, informal, and general induction activities.

It is important to ensure that a new teacher is well assimilated into the school community and participates in its diverse activities. Ideally, not only should a new teacher receive support on entry into the work community, other teachers should also be interested in their fresh views.

Typically, mentoring is provided only to new teachers in the induction phase. However, mentoring provided at different career stages for teachers has proven to be an effective way of supporting professional development (Heikkinen *et al.*, 2012; Heikkinen, Swachten, & Akyol, 2015).

Unfortunately, as Timperley (2011) notes, it seems that many professional development activities do not hold much meaning for teachers. She calls for paradigmatic shifts in conventional professional development approaches by actively involving teachers in their learning. She says these shifts entail “moving from professional development to professional learning, focusing on students, attending to requisite knowledge and skills, engaging in systematic inquiry into the effectiveness of practice, being explicit about underpinning theories of professionalism and engaging everyone in the system in learning” (Timperley, 2011, 4). In brief, professional learning requires teachers to be seriously engaged in their learning, not merely present at development activities. Putting students at the centre means being committed through professional learning to creating conditions in which everyone learns. Knowledge and skills must be foregrounded rather than viewed as methods of delivering professional learning. On the whole, professional learning should be an active process of systematic inquiry into the effectiveness of practice for student learning (See also Johnson & Sondergeld, 2015).

Timperley (2011) also emphasises that teachers must identify their knowledge needs and practices they should adopt be more effective with different kind of students. According to the TALIS 2013 results, teachers in the Nordic countries need to hone their skills in using ICT for teaching, teaching students with special needs and using new technologies at the workplace. The need and focus of in-service training and other professional development activities will change according to the different stages of a teacher’s career. Therefore, it is important to map educational needs and to make sure that teachers’ different career-specific needs are better taken into account than they are at present. Also, it is important to re-think the forms of professional development would best serve these purposes. For example, traditional training sessions outside the school could be replaced or augmented with work-related development programmes and online-based collaborative activities (e.g. Hökkä & Vähäsantanen, 2014).

Although at the official level systematic support is available for teachers in every Nordic country, there remains room for improvement. Cooperation between education providers could serve to bridge this gap considerably. For instance, the Nordic countries can exchange experiences on good practices. Or support for a new teacher can incorporate inputs from teacher training, in-service training and the employer (Heikkinen, Aho, & Korhonen, 2015).

Attention should be paid to in-service training and new forms of professional development and learning that are long-lasting and closely related to the work of teachers and to the development of the school community. It is important to secure an unbroken continuum from initial teacher training through the induction phase and to subsequent continuous professional development. Thus, the path for the development of professional competence should be created at the beginning of a teacher's career, consolidated through initial teacher education and supported until retirement. Continuum thinking of this kind emphasises the importance of synchronising and connecting basic and continuing teacher education.

4.8 References

- Anagnostopoulos, D., Smith, E. R., & Basmadjian, K. G. (2007). Bridging the university-school divide: Horizontal expertise and the "Two-Worlds Pitfall." *Journal of Teacher Education*, 58(2), 138–152. <http://dx.doi.org/10.1177/0022487106297841>
- Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education*, 27(1), 10–20. <http://dx.doi.org/10.1016/j.tate.2010.08.007>
- 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. <http://dx.doi.org/10.1016/j.jsp.2006.09.001>
- Crossman, A., & Harris, P. (2006). Job satisfaction of secondary school teachers. *Educational Management Administration and Leadership*, 34(1), 29–46. <http://dx.doi.org/10.1177/1741143206059538>

- Danielson, C. (2007). Enhancing professional practice: A framework for teaching. In: V. A. Alexandria, *Association for supervision and curriculum development (ASCDE)*. eBook collection (EBSCOhost), EBSCOhost, viewed on 20 October, 2015, from <http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=189709&site=ehost-live>
- Darling-Hammond, L. (2000). *Teacher quality and student achievement: A review of state policy evidence*. Education Policy Analysis Archives. Retrieved on 17 June, 2015, from <http://dx.doi.org/10.14507/epaa.v8n1.2000>
- de Vries, S., van de Grift, W. J. C. M., & Jansen, E. P. W. A. (2014). How teachers' beliefs about learning and teaching relate to their continuing professional development. *Teachers and Teaching*, 20(3), 338–357. <http://dx.doi.org/10.1080/13540602.2013.848521>
- European Commission. (2011). *Sweden: Continuing professional development for teachers working in early childhood and school education*. Retrieved on November 5, 2014, from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Sweden:Continuing_Professional_Development_for_Teachers_Working_in_Early_Childhood_and_School_Education
- European Commission. (2012a). *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Rethinking education: Investing in skills for better socio-economic outcomes*. Strasbourg, COM (2012) 669 final.
- European Commission. (2012b). *Denmark: Continuing professional development for teachers working in early childhood and school education*. Retrieved on 13 January, 2016, from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Denmark:Continuing_Professional_Development_for_Teachers_Working_in_Early_Childhood_and_School_Education
- European Commission. (2013). *Finland: Continuing professional development for teachers working in early childhood and school education*. Retrieved on 5 November, 2014, from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Finland:Continuing_Professional_Development_for_Teachers_Working_in_Early_Childhood_and_School_Education
- European Commission. (2014). *Norway: Continuing professional development for teachers working in early childhood and school education*. Retrieved on 5 November, 2014, from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Norway:Continuing_Professional_Development_for_Teachers_Working_in_Early_Childhood_and_School_Education
- European Commission. (2015). *Iceland: Continuing professional development for teachers working in early childhood and school education*. Retrieved on 25 November, 2015, from https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Iceland:Continuing_Professional_Development_for_Teachers_Working_in_Early_Childhood_and_School_Education

- Ferrari, A., Cachia, R., & Punie, Y. (2009). *Innovation and creativity in education and training in the EU member states: Fostering creative learning and supporting innovative teaching. Literature review on innovation and creativity in E&T in the EU member states (ICEAC)*. European Commission. Joint Research Centre. Institute for Prospective Technological Studies. Retrieved on 15 June, 2015, from ftp://ftp.jrc.es/users/publications/public/EURdoc/JRC52374_TN.pdf
- Geeraerts, K., Tynjälä, P., Heikkinen, H., Markkanen, I., Pennanen, M., & Gijbels, D. (2015). Peer-group mentoring as a tool for teacher development. *European Journal of Teacher Education*, 38(3). <http://dx.doi.org/10.1080/02619768.2014.983068>
- Goddard, J. T., & Foster, R. Y. (2001). The experiences of neophyte teachers: A critical constructivist assessment. *Teaching and Teacher Education* 17(3), 349–365. [http://dx.doi.org/10.1016/S0742-051X\(00\)00062-7](http://dx.doi.org/10.1016/S0742-051X(00)00062-7)
- Harland, J., & Kinder, K. (2014). Teachers' continuing professional development: framing a model of outcomes. *Professional Development in Education*, 40(4), 669–682. <http://dx.doi.org/10.1080/19415257.2014.952094>
- Hegarty, S. (2000). Teaching as a knowledge-based activity. *Oxford Review of Education*, 26(3/4), 451–465. <http://dx.doi.org/10.1080/713688541>
- Heikkinen, H. L. T., Aho, J., & Korhonen, H. (2015). *Opetus ja Saa Oppia [Teacher learns not] Development of teacher education continuum*. University of Jyväskylä. Finnish Institute for Educational Research.
- Heikkinen, H., Swachten, L., & Akyol, H. (2015). *Bridge over troubled water: New perspectives on induction*. Ankara: Gazi University. (In press.)
- Heikkinen, H.L.T., Jokinen, H. & Tynjälä, P. (2012). *Peer-group mentoring for teacher development*. London: Routledge.
- Henson, R. K. (2001). Teacher self-efficacy: Substantive implications and measurement dilemmas. Retrieved on 23 September, 2015, from <http://www.uky.edu/~eushe2/Pajares/EREkeynote.PDF>
- Hämäläinen, K. & Hämäläinen, K. (Eds.). (2011). *Opetustoimen henkilöstökoulutus osaamisen voimavarana. Muistio opetustoimen henkilöstökoulutuksen hyviin käytäntöihin ja kehittämistöimiin*. [Professional development for education personnel as a competence resource. A report on good practices and development measures in professional development for education personnel]. Opetus- ja kulttuuriministeriön julkaisuja 9 [Ministry of Education and Culture, publication 9].
- Hökkä, P., & Eteläpelto, A. (2014). Seeking new perspectives on the development of teacher education: a study of the Finnish context. *Journal of Teacher Education*, 65(1), 39–52. <http://dx.doi.org/10.1177/0022487113504220>
- Hökkä, P., & Vähäsantanen, K. (2014). Agency-centred coupling—a better way to manage an educational organization? *International Journal of Leadership in Education*, 17(2), 131–153. <http://dx.doi.org/10.1080/13603124.2013.783932>

- Johnson, C. C., & Sondergeld, T. A. (2015). Effective STEM education. In C.C. Johnson, E.E. Peters-Burton, & T. J. Moore (eds.) *STEM road map. A framework for integrated STEM education*. New York: Routledge, 203–216.
- Jokinen, H., Taajamo, M., & Välijärvi, J. (2014). *Pedagoginen asiantuntijuus liikkeessä ja muutoksessa – huomisen haasteita* [Pedagogical expertise in transition: challenges of tomorrow]. Jyväskylän yliopisto. Koulutuksen tutkimuslaitos.
- Klassen, R. M., & Chiu, M. M. (2010). Effect on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741–756. <http://dx.doi.org/10.1037/a0019237>
- Leask, M., & Younie, S. (2013). National models for continuing professional development: the challenges of twenty-first-century knowledge management. *Professional Development in Education, 39*(2), 273–287. Retrieved on 12 January, 2015, from <http://dx.doi.org/10.1080/19415257.2012.749801>
- OECD (2012). *Education at a Glance 2012: Highlights*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/eag_highlights-2012-en
- OECD (2014a). *TALIS 2013 results. An international perspective on teaching and learning*. OECD Publishing. <http://dx.doi.org/10.1787/9789264196261-en>
- OECD (2014b). *TALIS 2013 technical report*. OECD Publishing. Retrieved from <http://www.oecd.org/edu/school/TALIS-technical-report-2013.pdf>
- Skaalvik, E. M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession: Relations with school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education, 27*(6), 1029–1038. <http://dx.doi.org/10.1016/j.tate.2011.04.001>
- Snoeck, M., Eisenschmidt, E., Forsthuber, B., Holdsworth, P., Michaelidou, A., Dahl, J., & Pachler, N. (2010). *Developing coherent and system-wide induction programmes for beginning teachers - a handbook for policy makers*. Brussels: European Commission.
- Timperley, H. S. (2008). Teacher professional learning and development. Educational Practices Series-18. *International Academy of Education*. Retrieved from http://www.ibe.unesco.org/fileadmin/user_upload/Publications/Educational_Practices/EdPractices_18.pdf
- Timperley, H. S. (2011). *Realizing the power of professional learning*. New York: Open University Press.
- Tynjälä, P. & Heikkinen, H. (2011). Beginning teachers' transition to working life: Theoretical perspectives and best practices. *Zeitschrift für Erziehungswissenschaft, 14*(1), 11–33. <http://dx.doi.org/10.1007/s11618-011-0175-6>
- Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review, 3*(2), 130–154. <http://dx.doi.org/10.1016/j.edurev.2007.12.001>
- Tynjälä, P., & Gijbels, D. (2012). Changing world: changing pedagogy. In: P. Tynjälä, M-L. Stenström, & M. Saarnivaara (eds.) *Transitions and transformations in learning and education* (pp. 205–222). Dordrecht: Springer. http://dx.doi.org/10.1007/978-94-007-2312-2_13

Virtanen, A., Tynjälä, P., & Eteläpelto, A. (2014). Factors promoting vocational students' learning at work: study on student experiences. *Journal of Education and Work*, 27(1), 43-70. <http://dx.doi.org/10.1080/13639080.2012.718748>

Wei, R. C., Darling-Hammon, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession. A status report on teacher development in the U.S. and abroad*. Dallas: National Staff Development Council.

5. Can collegial work and school leader feedback improve teachers' self-efficacy in Nordic classrooms?

By Joakim Caspersen, NTNU Social Research, Norway

5.1 Summary

Collaboration with colleagues and feedback on teaching approaches, classroom work, and interaction with pupils create opportunities for teachers to improve their teaching practices. This can in turn have a powerful impact on student learning and outcomes. However, it is important to analyse and discuss how different types of feedback, and from whom, play a role in improving practices. In this chapter, the TALIS 2013 survey is used to shed light on the following question: Can collegial work and school leader feedback improve teachers' self-efficacy in Nordic classrooms?

Teachers' self-efficacy can be understood as their perception of their ability to organise and execute tasks necessary for certain kinds of performance. Recent research has identified teacher self-efficacy as an important mediating variable for teacher effects on student outcome, achievement and motivation. It has also been linked to predicting teacher attrition and burnout, teacher behaviour in the classroom, the efforts teachers invest in teaching and the goals they set. Self-efficacy is also shown to be a positive predictor of gains in children's vocabulary development in the right contexts. TALIS helps us distinguish between three

aspects of teacher self-efficacy: classroom management, instruction and student engagement.

A second theme addressed in this chapter is whether novice and experienced teachers differ in their self-efficacy levels and in how they collaborate and interact with colleagues and leaders. Although few differences are found between how novice and experienced teachers collaborate with colleagues, the differences in the feedback they receive and from whom are prominent. These differences indicate that collegial work and school leader feedback can improve teachers' self-efficacy in Nordic classrooms. However, the variation in the results between countries, between groups of teachers with different experience, between schools within countries and also between countries highlights the need for a differentiated policy approach, both across and within countries.

5.2 Introduction

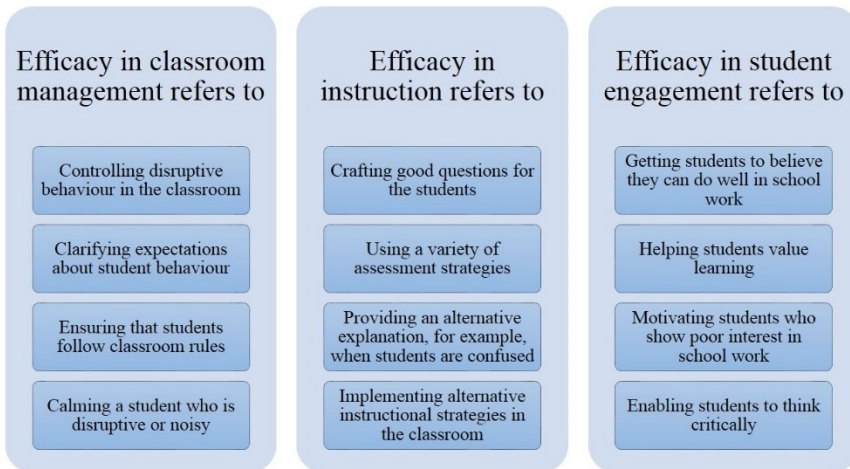
Feedback on teaching approaches, classroom work and interaction with pupils creates opportunities for teachers to improve their teaching practices. This can in turn have a powerful impact on student learning and outcomes (Reynolds *et al.*, 2014). However, the sources of feedback and support can vary, and many have argued that collegial and collective feedback and discussion are equally, if not more, important than feedback from superiors or leaders (Hargreaves, 2000; Hargreaves & Fullan, 2012; Hipp, Huffman, Pankake, & Olivier, 2008; Little, 1990). It has also been found that teachers in different stages of their career may benefit differently from feedback from colleagues and superiors (Caspersen & Raaen, 2014; Tschannen-Moran & Hoy, 2007). Caspersen and Raaen (2014) find that for novice teachers, contextual factors such as colleagues and collegial cooperation exert a stronger influence on teacher certainty and self-efficacy than support from superiors. On the other hand, for experienced teachers, contextual factors have a weaker relation with certainty and self-efficacy, possibly because they are likely to rely more heavily on memories and interpretations of similar past teaching experiences and are less dependent on the current context (Tschannen-Moran & Hoy, 2007).

In international research on teachers and schools, collaborative culture has been described as a key component for developing schools and teaching, especially as policy systems progress towards increased accountability (Hargreaves & Fullan, 2012; Mausethagen & Smeby, forthcoming). However, cross-country comparisons of teachers and schools within different policy contexts have been scarce mainly because of the unavailability of comparative data for statistically generalisable analyses. Some large-scale projects such as Teaching and Learning International Survey (TALIS), Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) have improved this situation by creating new opportunities for comparative research. Although TALIS, as well as the other surveys, is based on teachers' self-reported data and makes use of broad categories, the growing scope of projects such as TALIS has offered new insights into teachers' working conditions and practices across countries.

In this chapter, data from TALIS 2013 is used to explore answers to the following question: Can collegial work and school leader feedback improve self-efficacy in Nordic classrooms? Data analysis is followed by a discussion on the policy implications of the findings. Although schools are primarily about children's development, TALIS does not consider learning outcomes at the student level (except for countries that choose the TALIS-PISA link option). Instead, TALIS focuses on teacher variables known to be of importance for student outcomes. Self-efficacy has been identified as one such variable in the research literature on teachers and teaching. Teachers' self-efficacy can be defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" (Bandura, 1986, p. 391). Recent research evidence suggests that teacher self-efficacy is an important mediating variable for teacher effects on student outcome, achievement and motivation (Caprara, Barbaranelli, Steca, & Malone, 2006; Granger *et al.*, 2012; Midgley, Feldlaufer, & Eccles, 1989; Multon, Brown, & Lent, 1991; Pajares, 1996; Tschannen-Moran & Barr, 2004; Tschannen-Moran & Hoy, 2001). It is also reported to predict teacher attrition and burnout, teacher behaviour in the classroom, the efforts teachers invest in teaching and the goals they set (Skaalvik & Skaalvik, 2007; Tschannen-Moran & Hoy,

2001). Self-efficacy is also shown to be a positive predictor of gains in children’s vocabulary development in the right contexts (Guo, Piasta, Justice, & Kaderavek, 2010). TALIS items allow us to distinguish between three aspects of teacher self-efficacy: efficacy in classroom management, efficacy in instruction and efficacy in student engagement.

Figure 35: Three-dimensions of self-efficacy in TALIS 2013



It is evident from country-specific reports that there are large national differences between the feedback and appraisal teachers receive and their consequences. For instance, 26% of Finnish teachers report that they have never received any formal evaluation, while only 4% of Swedish teachers report the same. However, these national reports do not factor the differences at the school level within the countries in a systematic manner, which leaves room for further exploration. This chapter uses data from TALIS 2013 to examine whether different types of appraisal and feedback can be linked to different aspects of self-efficacy, and then attempts to map the findings to the different forms of feedback and appraisal used in Nordic countries.

Further, the chapter focuses on comparing the self-efficacy and feedback outcomes between novice teachers and experienced teachers, as self-effi-

cacy and its antecedents have been shown to be related differently to different career-stages. The difficulties faced by novice teachers have been described as a “shock” by many researchers over the years (Corcoran, 1981; Gaede, 1978; Ryan, 1970; Smagorinsky, Gibson, Bickmore, Moore, & Cook, 2004; Stokking, Leenders, de Jong, & van Tartwijk, 2003), although newer research contributions are more nuanced and focus more on how novice teachers’ coping can be accommodated and supported by contextual factors in addition to individual resources (Eteläpelto, Vähäsantanen, & Hökkä, 2015; Mansfield, Beltman, & Price, 2014; Ozder, 2011). Studies have shown that novice teachers’ coping do not differ significantly from those of experienced teachers (Caspersen, 2013b), but novice and experienced teachers benefit differently from the support extended by superiors and colleagues. A general challenge with research on novice teachers, however, is that the analyses are not performed with contrast groups and that novice teachers are a priori assumed to differ from more experienced teachers (Caspersen, 2013a). As teacher certainty seems to be related to age/experience (Munthe, 2001), it is important to distinguish between different stages of a teacher’s career. In this chapter, teachers with 0–3 years’ experience are compared to teachers with 11–14 years’ experience.

5.3 Research questions and approach

Essentially, this chapter attempts to address the following three questions:

1. How are the Nordic countries (Norway, Sweden, Denmark, Iceland and Finland) different and similar in terms of feedback and appraisal received by teachers in different stages of their careers, from superiors and colleagues?
2. How much of the variation in feedback and appraisal received by teachers from superiors and colleagues, in different stages of their careers, is related to the school they are employed in?
3. What are the relationships between (1) collegial feedback and appraisal, (2) feedback and appraisal from superiors and (3) teacher self-efficacy in Nordic countries?

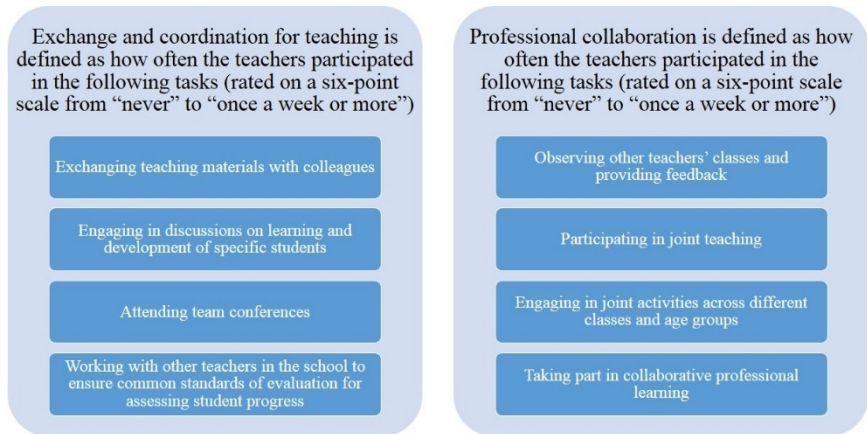
Data for the analyses are obtained from survey questionnaires completed by teachers from lower secondary grade during TALIS 2013, which saw participation from all the Nordic countries. As discussed, it is distinguished between teachers with 0–3 years' experience (novice) and those with 11–14 years' experience (experienced). According to the weighted results of the TALIS 2013 data, the number of novice teachers in Finland and Norway were 1,862 and 3,085, respectively, and the number of experienced teachers in Finland and Sweden were 2,166 and 5,830, respectively. The entire population of teachers from Iceland participated in the survey; therefore, non-weighted data for the number of novice and experienced were 203 and 269, respectively.

The answer to the first question is derived from descriptive analyses of teacher responses. To answer question 2, the intra-class correlation coefficient (ICC) is calculated for each group of questions. ICC describes how strongly each teacher within a school resembles other teachers from the same school, and thus provides a measure of how the specific school or school culture can explain teachers' answers. Calculating ICC is also the first step of a multi-level analysis, which is used to answer question 3 (multi-level linear regression). This means that the analyses controls for school affiliation for all teachers, when examining the relationship between feedback and appraisal and self-efficacy. In addition, the multi-level analyses include control for relevant teacher background variables, i.e. full time/part time and gender, as well as for the differences between and within Nordic countries.

5.4 Collegial work, feedback and appraisal in Nordic schools

Feedback and appraisal from colleagues is captured by two different measures in TALIS, which pertain to collaborative or collegial culture: *exchange and coordination for teaching* and *professional collaboration*.⁵

Figure 36: Two dimensions of teacher collaboration in TALIS 2013

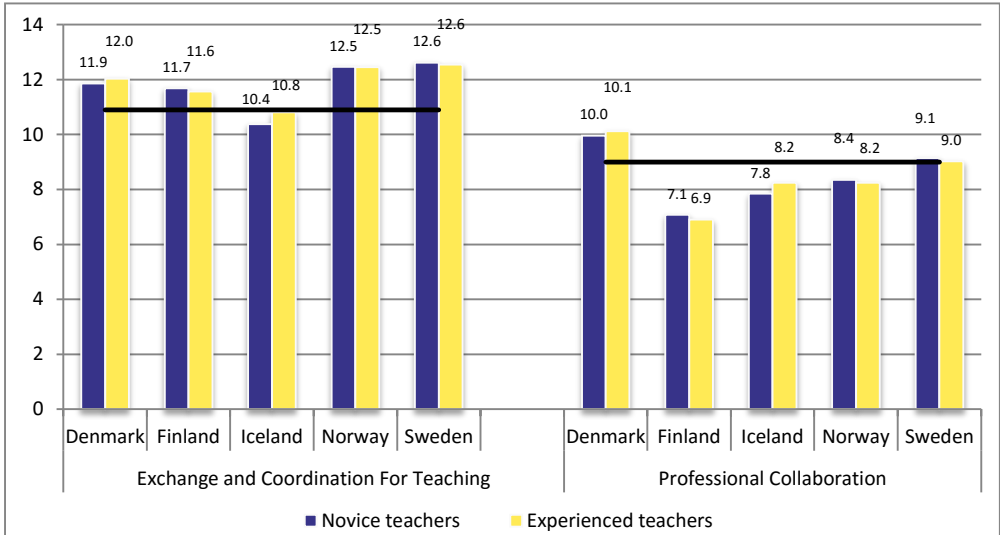


The two measures emphasise different aspects of how teachers work together and provide one another with the necessary feedback, and together they measure cooperation along two different dimensions. The professional collaboration dimension is related to progressive forms of professionalism (OECD, 2014: 166) and linked to more binding and profound sharing and development among teachers (Clement & Vandenberghe, 2000; Little, 1990). Exchange and coordination, on the other hand, is seen as a less binding type of collaboration and is more common among the OECD countries (OECD, 2014). Thus, although the two categories are broad, they cover two important dimensions of teacher collaboration.

⁵ The statistical properties of these measures are presented in the TALIS 2013 Technical Report (<http://www.oecd.org/edu/school/TALIS-technical-report-2013.pdf>).

The mean scores of these dimensions for each of the Nordic countries are presented in Figure 3. Each measure is standardised such that a score of 10 denotes the mid-point of the response scale.

Figure 37: Mean scores on teacher collaboration dimensions in the Nordic countries. Black horizontal line indicates international average

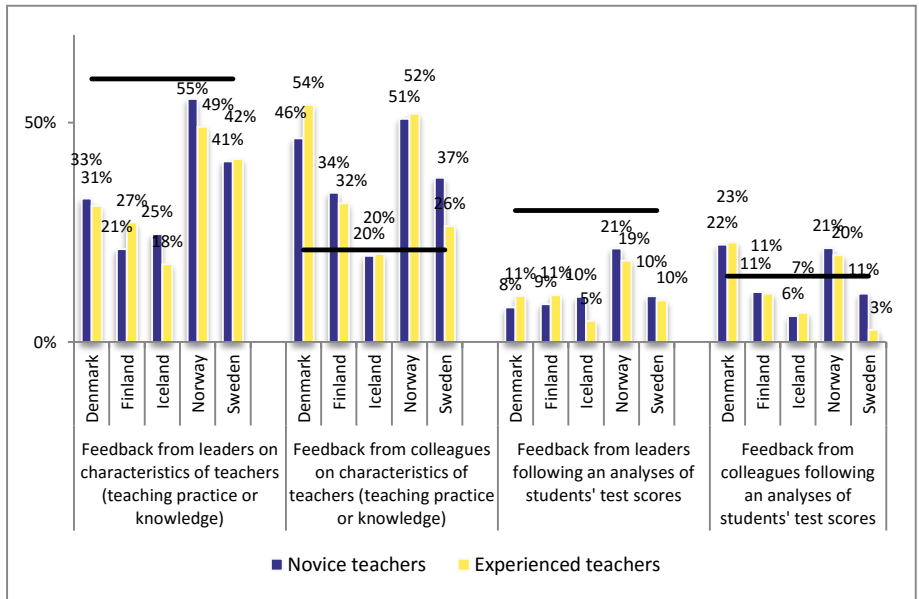


Scores reveal large variations between the Nordic countries in terms of teacher collaboration.⁶ However, hardly any differences are seen in the scores of novice and experienced teachers, a finding that is in line with more recent studies on the experiences of novice teachers. Iceland is an exception to this trend, as novice teachers participate less in exchange and coordination for teaching or professional collaboration. Their scores on exchange and coordination for teaching are the lowest, while those of the Swedish teachers are the highest. Finnish teachers have the lowest level of professional collaboration, while Danish teachers have the highest.

⁶ The differences between countries are examined using ANOVA, with post hoc-tests (Scheffe) in order to identify significant differences between countries. All differences commented on in the chapter are significant at .05-level. Differences between experienced and novice teachers were examined using t-tests for independent samples.

As part of TALIS, teachers are also asked to specify their sources of feedback and the information basis for the feedback. Response options for the former include school principal/other members of the school management team and other teachers/assigned mentors. Response options for the latter consist of feedback on teaching characteristics (practice and knowledge),⁷ and on student outcomes.⁸

Figure 38: Percentage of teachers in the Nordic countries who received feedback from leaders or colleagues on teaching characteristics (practice/knowledge) or student outcomes. Black horizontal line indicates international average



As with Figure 3, clear differences are seen between the Nordic countries in terms of the type and sources of feedback. In Iceland, Norway and Sweden, teachers mostly receive feedback from superiors and school management on teaching characteristics (practice and content knowledge), whereas in Denmark and Finland, it is mostly colleagues who provide feedback on this aspect. Norwegian teachers, and to some extent Swedish

⁷ Feedback based on direct observation of classroom teaching, students' evaluation of the teaching, or an assessment of teachers' content knowledge.

⁸ Feedback based on an analysis of students' test scores.

teachers, receive high level of feedback on their teaching and content knowledge from both colleagues and school leaders.

Further, Norwegian teachers top the list of teachers receiving feedback on the basis of an analysis of students' test scores from school leadership, whereas Danish teachers top the list of those who receive such feedback from their colleagues. Interestingly, Danish teachers report receiving little feedback from school leaders on students' test scores, but have the highest amount of feedback from colleagues on the same topic. All in all, Norwegian and Swedish teachers seem to receive most feedback, while Icelandic and Finnish teachers receive relatively less feedback.

Scores in Figure 2 also highlight differences in the sources and types of feedback received by novice and experienced teachers. Novice teachers in Finland receive less feedback from leaders on teaching practice and content knowledge than their experienced colleagues, while those in Iceland and Norway receive more feedback than the experienced teachers on their teaching practice and content knowledge from the school leaders. In Denmark, experienced teachers receive more feedback on their teaching practice and content knowledge from their colleagues than novice teachers, while it is (clearly) the other way around among Swedish teachers.

In terms of test scores, the only distinct difference between novice and experienced teachers is found in Sweden: 11% of the novice teachers report that they receive feedback from their colleagues on their students' test scores, while only 3% of the experienced teachers report the same.

Finally, it is interesting to note that the Nordic countries, in many cases, deviate considerably from the international average (Figure 1 & 2). In Nordic countries, feedback from leaders on teaching characteristics and students' test scores seems less frequent than in other countries, perhaps indicating a more collegial and non-hierarchical distribution of work in schools.

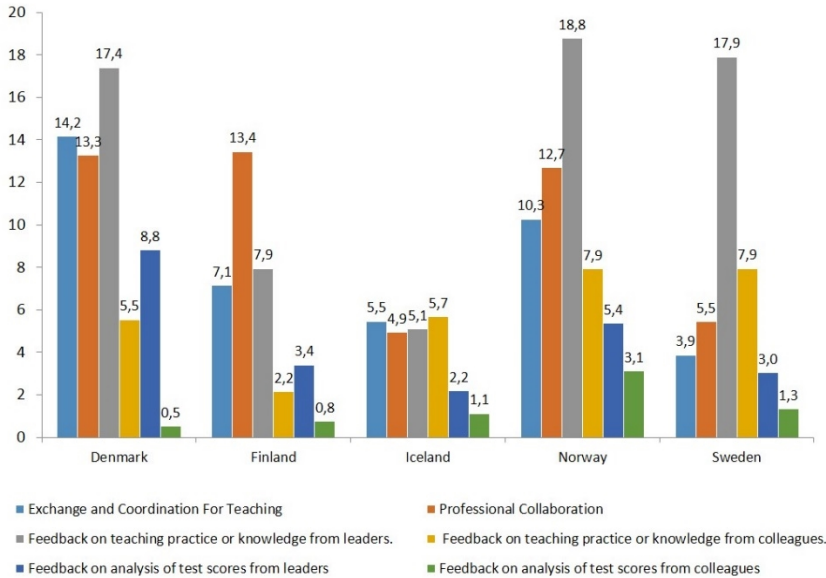
5.5 Variation in feedback within countries

Variation in feedback and appraisal between the countries points to the different experiences of novice and experienced teachers. In this section, we examine how much of this variation is related to the school in which the teachers are employed. According to Hansen, Gustafsson and Rosén (2011: 25) “variability in the level of performance between different schools is of great interest both from research and policy perspectives”. This is because large observed differences between schools may indicate differences in how teachers’ work is carried out within the same country. While the differences between Nordic countries are known, the question that remains to be addressed is “How large are the differences *within* the different Nordic countries?”

Accordingly, we have calculated the ICC for all the measures presented in Figures 3 and 4 (Figure 5).⁹ The ICC expresses the proportion of variance that can be attributed to a certain school. Large variation between schools translates into a high ICC.

⁹ For the four dichotomous variables – measuring feedback from colleagues and superiors on students’ test scores and teaching practice and knowledge – the ICCs have been calculated in the same way as for the two ordinal variables.

Figure 39: Proportion of variance explained by school of employment for indicators on collaboration, feedback and appraisal



The average level of variance explained by the school is high, but it also varies between the different Nordic countries. On average, Danish teachers report the highest ICCs, suggesting that systematic variation between schools is the highest in Denmark, closely followed by Norway.

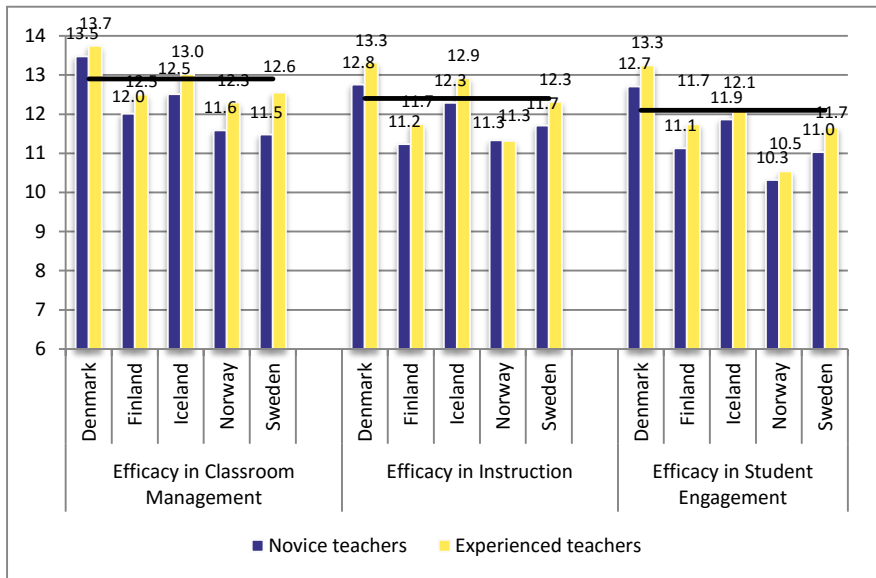
On the other end, Icelandic teachers have the lowest ICC scores, indicating that the variation between schools is the least in Iceland, when it comes to teacher collaboration and feedback. Finland and Sweden are somewhere in the middle but have a high degree of systematic variation between schools.

With regard to the measures analysed, feedback on teaching practice and content knowledge from leaders has the highest ICCs in Denmark, Norway and Sweden. Finland has the highest systematic variation between schools in professional collaboration, while in Iceland this variation is generally low.

5.6 Self-efficacy in Nordic classrooms – differences and variation

Next, we explore the third research issue: the relationship between feedback and appraisal from superiors and colleagues on the one hand and teacher self-efficacy on the other. Self-efficacy has been identified as a key mediating variable for a variety of outcomes. TALIS includes three measures of self-efficacy: efficacy in classroom management, efficacy in instruction and efficacy in student engagement. Figure 6 shows the mean scores for each of the three efficacy measures.

Figure 40: Mean scores on efficacy measures for teachers in the Nordic countries. Black horizontal line indicates international average



Danish teachers report the highest level of self-efficacy on all three measures. While the differences between the other four Nordic countries are not as distinct, it appears that Icelandic teachers have somewhat higher mean self-efficacy scores, while Norwegian teachers have somewhat lower self-efficacy scores. Finnish and Swedish teachers are somewhere in the middle. It is important to note that all five Nordic

countries have average scores higher than the scale midpoints (10), indicating that they are more positive than negative in assessing their own ability to make a difference in the classroom. Interestingly, Danish teachers' self-efficacy scores are higher than the international average. Of the three measures of self-efficacy, Nordic countries report the highest means for efficacy in the classroom and the lowest for efficacy in student engagement.

Significant differences are observed between novice and experienced teachers, in that experienced teachers have higher self-efficacy than novice teachers in all the Nordic countries. However, differences in efficacy in instruction among Norwegian teachers and efficacy in student engagement among Icelandic teachers are non-significant.

ICCs for self-efficacy variables are, in general, small in all the five Nordic countries (smaller than five percent between school variance for all three measures). This is probably related to the fact that self-efficacy is a more individualistic feeling, unlike other measures that are oriented towards collective practices at schools. Thus, it is easier to find variation at the school level on practices that are, per se, linked to the school than to find variation at the school level on individual characteristics.

5.7 The relationship between feedback and self-efficacy in the Nordic countries

To examine the relationship between the different types of collegial feedback, appraisal and teacher self-efficacy, three multi-level regression models have been fitted, one for each of the three self-efficacy measures. Collaboration and feedback variables, in addition to gender and full-time/part-time employment, are the independent variables. The multi-level model allows for including school-level variance (Figure 3) in the model, i.e. controlling for the systemic variation between schools. The model uses the data from all the five Nordic countries together; however, to control for variation between the five Nordic countries, we have also included control for the five Nordic countries. Thus, we present a three-level model, where teachers are nested within schools within countries.

The effects¹⁰ of the different variables are presented only graphically for easy interpretation.¹¹

Table 21: Multi-level regression results on three types of teacher self-efficacy. Significant at $p < 0,05$ level. Positive results are marked with +. School-level and country-level results not presented

	Self-efficacy in class-room management		Self-efficacy in instruction		Self-efficacy in student engagement	
	Novice	Experienced	Novice	Experienced	Novice	Experienced
Female		+		+		+
Full time	+	+		+		+
Exchange and coordination for teaching		+	+	+		+
Professional collaboration	+				+	+
Feedback on teaching practice or knowledge from leaders						
Feedback on teaching practice or knowledge from colleagues						
Feedback on analysis of test scores from leaders		+	+	+	+	+
Feedback on analysis of test scores from colleagues						

¹⁰ A caveat to be acknowledged is that the analyses do not examine any causal effect. The relationship between dependent and independent variables are in some ways arbitrary, except for the fact that gender and experience cannot be a result of self-efficacy. However, it is fully possible to hypothesize that high levels of self-efficacy in teachers lead them to collaborate more. The findings should therefore be interpreted as relations, and not as effects, although it is common to use the word “effect” when presenting results from regression analyses.

¹¹ It should be noted that the multi-level regression model used here is sensitive to errors of measurement in the independent variables. More error of measurement in a variable will cause its effect to be underestimated. Thus, although “feedback on analysis of test scores from leaders” emerges as the strongest school-level predictor of teacher self-efficacy, this may partly be related to the fact that it might be the most reliable variable. More detailed statistics can be requested from the author.

Figure 7 shows that teachers' gender does not influence any of the three types of self-efficacy for novice teachers, but influences all types of self-efficacy among experienced teachers. Experienced female teachers tend to have higher self-efficacy than experienced male teachers.

Novice and experienced teachers working full time (more than 90% full time hours) have higher self-efficacy in classroom management than teachers working part time. Self-efficacy in instruction and self-efficacy in student engagement are found to be higher among experienced teachers who work full time. The reasons for working full time or part time could be based in choice or circumstances. Ultimately, the variation only implies that full-time teachers are more connected to the daily work and routines in schools and, probably, therefore, have higher levels of self-efficacy.

As mentioned, teacher collaboration is measured using two variables: (1) exchange and coordination for teaching and (2) professional collaboration, with the former involving less commitment than the latter. Exchange and coordination for teaching has a positive relationship with all three dimensions of self-efficacy for experienced teachers, and with self-efficacy in instruction for novice teachers. Professional collaboration has a positive relation with self-efficacy in classroom management for novice teachers, and with self-efficacy in student engagement for both experienced and novice teachers.

Barring feedback from leaders based on analyses of students' test scores, the feedback variables do not seem to have a strong and clear relationship with self-efficacy. Feedback from leaders on students' test scores has a positive relation with all three dimensions of self-efficacy, except for self-efficacy in classroom management for novice teachers.

Instead of comparing results between analyses of novice and experienced teachers, if we were to include teachers' age/experience as an independent variable, the results would be slightly different. Age (or experience) would have a significant positive relationship with all three self-efficacy measures. Professional collaboration would have a positive relationship with self-efficacy in instruction, and just with self-efficacy in student engagement. Feedback on analyses of students' test scores from leaders would still have a positive relation with all three self-efficacy di-

mensions, and feedback on analyses of students' test scores from colleagues would have a positive relation with self-efficacy in instruction and in student engagement.

5.8 Impact of collaboration and feedback on novice and experienced teachers

Findings from our analysis of TALIS 2013 data point to large variations between the Nordic countries, in terms of the type of feedback received by teachers, the sources of the feedback and the teachers' stage in their teaching career at the time of receiving feedback. From a policy perspective, variations between the Nordic countries are perhaps best interpreted as different areas for improvement. It indicates that school improvement policies need not follow the one-size-fits-all principle, regardless of national context. Further, the large variation between schools within each country suggests that cross-nation policies may be hard to implement, as they would resonate in very different ways with local challenges and situations.

Research on international school policy and reform has shown that reforms and policy initiatives should align with local needs to survive beyond the classroom threshold (Cuban, 1998). Moreover, the traditional view within organisational studies suggests that different institutional logics govern the process of translating reforms to local implementation (Thornton & Ocasio, 2008; Thornton, Ocasio, & Lounsbury, 2015). In other words, the uptake of reforms is influenced by idiosyncratic, socially constructed, historic patterns of approaches, assumptions and values. Thus, school development that does not align with local needs is not translated into practice at schools, and the content of these reforms tends to be repeated in later reforms and policy initiatives without evidence-based analysis (Cuban, 1990). Comparative analyses of variation within and between countries, such as in this chapter and volume, can play an important role in developing policy and reform practice as they highlight the need for tailoring solutions.

The results from the regression analyses point to many different linkages between collegial work and feedback on the one hand, and the three self-efficacy dimensions on the other hand. As discussed earlier, the differences in novice and experienced teachers' self-efficacy is likely related to the teachers' participation in collegial practices at schools, although such differences between novice and experienced teachers are small. This finding sheds light on an important facet of comparative research strategies: variables can have different impact levels in different groups, despite equal baseline levels.

Experienced teachers with high self-efficacy in classroom management and instruction are mostly female, work full time, participate in exchange and collaboration for teaching (exchange teaching materials with colleagues, engage in discussions about the learning development of specific students, work with other teachers in their school to ensure common standards in evaluation for assessing student progress and attend team conferences), and receive feedback from their leaders (principal or other members of school management team) on students' test scores. Experienced teachers who have high self-efficacy in student engagement seem to have the same characteristics as those with high self-efficacy in classroom management and instruction, but they also participate in professional collaboration (joint teaching, observing other teachers' classes and providing feedback, engaging in joint activities across different classes and age groups and taking part in collaborative professional learning).

Novice teachers with high self-efficacy in classroom management work full time and participate in professional collaboration with their colleagues. Novice teachers with high self-efficacy in instruction participate in exchange and coordination for teaching and receive feedback based on analyses of students' test scores from leaders. Novice teachers with high levels of self-efficacy in student engagement also receive feedback on analyses of test scores from leaders and participate in professional collaboration with colleagues.

Overall, novice teachers seem to benefit more from participation in professional collaboration – the in-depth form of teacher collaboration – while experienced teachers seem to benefit more from participation in exchange and coordination for teaching – the lighter form of collaboration. This finding is in line with previous findings (Caspersen &

Raaen, 2014). Whether this trend is linked to the length of experience or to the fact that they belong to a younger cohort is difficult to ascertain from the analyses in this chapter. Many have argued that new generations of teachers approach their tasks differently than their older colleagues, especially when it comes to issues such as autonomy, accountability, involvement and interference in their work (Mausethagen, 2013; Stone-Johnson, 2013; Wilkins, 2011). The movement towards professional collaboration cultures to strengthen the quality of schooling (Hargreaves, 2000; Hargreaves & Fullan, 2012) may also be supported and endorsed through a generational shift among teachers. Both experienced and novice teachers benefit from receiving feedback on analyses of test scores from leaders.

The positive relation between feedback from leaders on students' test scores for all the teachers on all three dimensions of self-efficacy (except self-efficacy in classroom management for novice teachers) can be interpreted in different ways. Of course, a strict interpretation would imply that *students' test scores* are of paramount importance. School leaders who discuss test scores with teachers ensure higher levels of self-efficacy. A broad interpretation is that teachers who discuss students' progress with their leaders, and use tests as mapping tools in their teaching, also have a higher level of self-efficacy.

Using tests for assessing students' development and learning is an inherent part of a teacher's role. Discussing these results with school leaders would imply that the school leadership is positively involved in the educational practice of teachers (see also Tajaamo, this volume). Thus, policy initiatives aiming to strengthen the involvement of school leaders in teachers' work would be a way forward in school development, and this resonates with many policy initiatives introduced in the Nordic countries and abroad (see chapter 4 in Lysø, Stensaker, Aamodt, & Mjøen, 2011).

5.8.1 Differences across the Nordic countries

Figure 4 shows that professional collaboration is positively related to self-efficacy in student engagement for both experienced and novice teachers. However, Figure 1 shows that in Finland, the level of reported professional collaboration is low for both experienced and novice teachers. Whether

such differences should be interpreted as a negative result for the country in question is difficult to determine. One reason is that the broad categories used in international surveys such as TALIS make it difficult to undertake fine-grained analyses. Another reason is the inherent limitation of self-reported data. However, given that the findings are based on self-reported data, one could speculate, as self-efficacy is known to be associated with different kinds of school outcomes, that Finnish schools should utilise the potential to improve the collaboration between teachers. Finnish teachers do not have higher mean scores on self-efficacy in student engagement than other Nordic teachers. Thus, it seems plausible that increased focus on professional collaboration in schools could improve Finnish teachers' self-efficacy in student engagement. On the other hand, Danish teachers' high level of professional collaboration and high level of self-efficacy in student engagement indicate that school development initiatives should perhaps adopt a different direction in Denmark. For instance, development of the school leader role and leadership involvement in pedagogical discussions pertaining to particular students (as indicated by the question "feedback from leaders following an analyses of students' test scores") seem to display more potential for improvement in Denmark (based on the limited numbers of variables included in the analyses in this chapter). However, more detailed investigations with nuanced research methods are needed to confirm such interpretations.

5.8.2 *Variation between schools within the Nordic countries*

The findings on the proportion of variance explained at the school level (Figure 3) and the self-efficacy measures (Figure 4) indicate that large systemic differences between schools are present in all Nordic countries except Iceland. First, this implies that it is important to consider teachers' responses on feedback and collegial work at the school level. Second, the policy implications of such differences should be addressed. If national policies are geared towards streamlining teachers' work, large variations between schools imply that national policies have not been adequately implemented or enforced. However, if national policies grant schools the autonomy to work towards national goals in the way they find best, large

variations might indicate that schools have applied different means to reach their goals. Such a policy approach, again, suggests that school leadership and overall school involvement are important for reaching local educational goals. If local needs are to be met, they must be identified and catered to, and the national systems for developing schools should be geared towards meeting needs from a bottom-up perspective as much as from a top-down perspective (as a national reform would imply in most cases).

5.9 Conclusion

The motive behind this chapter was to examine whether collegial work and school leader feedback can improve teachers' self-efficacy in Nordic classrooms. Overall, the findings indicate that feedback and collaboration "styles" are different among the Nordic countries. This chapter sheds light not only on the differences in how novice and experienced teachers collaborate with colleagues but also on the differences in the types and sources of feedback that the Nordic teachers receive. Novice teachers in Finland receive less feedback from leaders on teaching practice and content knowledge than their experienced colleagues, while in Iceland and Norway novice teachers receive more feedback than their experienced counterparts on their teaching practice and content knowledge from the school leaders. In Denmark, experienced teachers receive more feedback from their colleagues on their teaching practice and content knowledge, while it is (clearly) the other way around in Sweden. These differences, as well as the overall country-specific variations, highlight the need for a differentiated policy approach between and within countries. The analyses of self-efficacy show that novice teachers have lower self-efficacy than experienced teachers and that feedback and collegial work is differently related to the self-efficacy of experienced and novice teachers.

Thus, the answer to the question of whether collegial work and school leader feedback can improve teachers' self-efficacy in the Nordic classrooms is most definitely "yes", but the more important question is "how". The variation in results between countries and between groups of teachers with different experience requires well-balanced initiatives that

promote teacher self-efficacy. The variations between the schools within each country add another layer of complexity, suggesting that national policy initiatives must take into account local variations.

5.10 References

- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, N. J.: Prentice-Hall.
- 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. <http://dx.doi.org/10.1016/j.jsp.2006.09.001>
- Caspersen, J. (2013a). *Professionalism among novice teachers. How they think, act, cope and perceive knowledge*. (Ph.d. dissertation). Oslo: Oslo and Akershus University College. <http://hdl.handle.net/10642/1710>
- Caspersen, J. (2013b). The valuation of knowledge and normative reflection in teacher qualification: A comparison of teacher educators, novice and experienced teachers. *Teaching and Teacher Education, 30*(0), 109–119. <http://dx.doi.org/10.1016/j.tate.2012.11.003>
- Caspersen, J., & Raaen, F. D. (2014). Novice teachers and how they cope. *Teachers and Teaching: theory and practice, 20*(2), 189–211. <http://dx.doi.org/10.1080/13540602.2013.848570>
- Clement, M., & Vandenberghe, R. (2000). Teachers' professional development: A solitary or collegial (ad) venture? *Teaching and Teacher Education, 16*(1), 81–101. [http://dx.doi.org/10.1016/S0742-051X\(99\)00051-7](http://dx.doi.org/10.1016/S0742-051X(99)00051-7)
- Corcoran, E. (1981). Transition shock: The beginning teacher's paradox. *Journal of teacher education, 32*(3), 19–23. <http://dx.doi.org/10.1177/002248718103200304>
- Cuban, L. (1990). Reforming again, again, and again. *Educational Researcher, 19*(1), 3–13. <http://dx.doi.org/10.3102/0013189X019001003>
- Cuban, L. (1998). How schools change reforms: Redefining reform success and failure. *The Teachers College Record, 99*(3), 453–477.
- Eteläpelto, A., Vähäsantanen, K., & Hökkä, P. (2015). How do novice teachers in Finland perceive their professional agency? *Teachers and Teaching, 21*(6), 660–680. <http://dx.doi.org/10.1080/13540602.2015.1044327>
- Gaede, O. F. (1978). Reality shock: A problem among first-year teachers. *The Clearing House, 51*(9), 405–409. <http://dx.doi.org/10.1080/00098655.1978.9957085>
- Granger, E. M., Bevis, T. H., Saka, Y., Southerland, S. A., Sampson, V., & Tate, R. L. (2012). The efficacy of student-centered instruction in supporting science learning. *Science, 338*(6103), 105–108. <http://dx.doi.org/10.1126/science.1223709>

- Guo, Y., Piasta, S. B., Justice, L. M., & Kaderavek, J. N. (2010). Relations among pre-school teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching and Teacher Education, 26*(4), 1094–1103. <http://dx.doi.org/10.1016/j.tate.2009.11.005>
- Hansen, K. Y., Gustafsson, J.-E., & Rosén, M. (2011). School performance differences and policy variations in Finland, Norway and Sweden. In: K. Y. Hansen, J.-E. Gustafsson, M. Rosén, S. Sulkunen, K. Nissinen, P. Kupari, R. F. Ólafsson, J. K. Björnsson, L. S. Grønmo, L. Rønberg, J. Mejding, I. C. Borge, & A. Hole (Eds.), *Northern lights on TIMSS and PIRLS. Differences and similarities in the Nordic countries*. Nordic Council of Ministers. <http://dx.doi.org/10.6027/TN2014-528>
- Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers & Teaching, 6*(2), 151–182. <http://dx.doi.org/10.1080/713698714>
- Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York/Ontario: Teachers College Press.
- Hipp, K., Huffman, J., Pankake, A., & Olivier, D. (2008). Sustaining professional learning communities: Case studies. *Journal of Educational Change, 9*(2), 173–195. <http://dx.doi.org/10.1007/s10833-007-9060-8>
- Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teachers College Record, 91*(4), 509–536.
- Lysø, I. H., Stensaker, B., Aamodt, P. O., & Mjøen, K. (2011). *Ledet til ledelse : nasjonal rektorutdanning i grunn- & videregående skole i et internasjonalt perspektiv*. [Led to leadership: evaluation of the national training program for school leaders]. Oslo: NIFU.
- Mansfield, C., Beltman, S., & Price, A. (2014). 'I'm coming back again!' The resilience process of early career teachers. *Teachers and Teaching, 20*(5), 547–567. <http://dx.doi.org/10.1080/13540602.2014.937958>
- Mausethagen, S. (2013). Accountable for what and to whom? Changing representations and new legitimation discourses among teachers under increased external control. *Journal of Educational Change, 14*(4), 423–444. <http://dx.doi.org/10.1007/s10833-013-9212-y>
- Mausethagen, S., & Smeby, J.-C. (forthcoming). Contemporary education policies and teacher professionalism. In: M. Dent, I. Bourgeault, J.-L. Denis, & E. Kuhlmann (Eds.), *Routledge Companion on Professions and Professionalism*.
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*(2), 247. <http://dx.doi.org/10.1037/0022-0663.81.2.247>
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30. <http://dx.doi.org/10.1037/0022-0167.38.1.30>

- Munthe, E. (2001). Professional uncertainty/certainty: How (un)certain are teachers, what are they (un)certain about, and how is (un)certainty related to age, experience, gender, qualifications, and school type? *European Journal of Teacher Education*, 24(3), 355–368. <http://dx.doi.org/10.1080/02619760220128905>
- OECD. (2014). *TALIS 2013 results: An international perspective on teaching and learning*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264196261-en>
- Ozder, H. (2011). Self-efficacy beliefs of novice teachers and their performance in the classroom. *Australian Journal of Teacher Education*, 36(5), 1–15. <http://dx.doi.org/10.14221/ajte.2011v36n5.1>
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543–578. <http://dx.doi.org/10.3102/00346543066004543>
- Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): a state-of-the-art review. *School Effectiveness and School Improvement*, 25(2), 197–230. <http://dx.doi.org/10.1080/09243453.2014.885450>
- Ryan, K. (1970). *The first year of teaching*. Chicago: The University of Chicago Press.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burn-out. *Journal of Educational Psychology*, 99(3), 611–625. <http://dx.doi.org/10.1037/0022-0663.99.3.611>
- Smagorinsky, P., Gibson, N., Bickmore, S. T., Moore, C. P., & Cook, L. S. (2004). Praxis shock: Making the transition from a student-centered university program to the corporate climate of schools. *English Education*, 36(3), 214–245.
- Stokking, K., Leenders, F., de Jong, J., & van Tartwijk, J. (2003). From student to teacher: reducing practice shock and early dropout in the teaching profession. *European Journal of Teacher Education*, 26(3), 329–350. <http://dx.doi.org/10.1080/0261976032000128175>
- Stone-Johnson, C. (2013). Parallel professionalism in an era of standardisation. *Teachers and Teaching*, 20(1), 74–91. <http://dx.doi.org/10.1080/13540602.2013.848514>
- Thornton, P. H., & Ocasio, W. (2008). Institutional logics. In P. H. Thornton, W. Ocasio, & R. Greenwood (Eds.), *The Sage handbook of organizational institutionalism* (pp. 99–129). London, England: Sage. <http://dx.doi.org/10.4135/9781849200387.n4>
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2015). *The institutional logics perspective. Emerging trends in the social and behavioral sciences*. John Wiley & Sons, Inc. <http://dx.doi.org/10.1002/9781118900772.etrds0187>
- 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. <http://dx.doi.org/10.1080/15700760490503706>
- Tschannen-Moran, M., & Hoy, A. W. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944–956. <http://dx.doi.org/10.1016/j.tate.2006.05.003>

- Tschannen-Moran, M., & Hoy, A. W. H. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. [http://dx.doi.org/10.1016/S0742-051X\(01\)00036-1](http://dx.doi.org/10.1016/S0742-051X(01)00036-1)
- Wilkins, C. (2011). Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system. *Professional Development in eEducation, 37*(3), 389–409. <http://dx.doi.org/10.1080/19415257.2010.514204>

6. Characteristics of high-performing students in mathematics

An exploratory analysis of PISA data from the Nordic countries

By Peter Nyström, University of Gothenburg, Sweden

6.1 Summary

This article aims at identifying characteristics of high-achieving students in mathematics in the Nordic countries, using data from student questionnaires in PISA 2003 and PISA 2012. A deeper understanding of the backgrounds, beliefs, attitudes, and learning strategies of these students can hint to factors important for the teaching and learning of mathematics.

Questionnaire items, and indices derived from items, were screened in order to identify differences between the 5% best performing students in each country and the 5% of students performing at and around the country median. The screening of 181 variables from PISA 2003 and 389 variables from PISA 2012, gave a multitude of results, and a fairly comprehensive selection of these results are presented in the article.

Many differences between high- and median-performing mathematics students were identified. As expected, the socioeconomic and educational background is higher for high-performing students. The high-performing groups of students have a lower percentage of girls compared to the median-performing groups for all Nordic countries, but no general difference is detected concerning percentages of students born in the country and

students speaking the language of the test at home. High-performing students study and learn mathematics differently, and differ in their perception of control, attribution of success and other motivational beliefs. They report higher perseverance and more positive attitudes to mathematics. These factors are most likely very important for student learning, and pedagogically significant for all students and teachers.

A set of characteristics of high-performing students in mathematics emerges through this analysis, but also aspects that do not differentiate the high-performing groups from the median-performing groups in the Nordic countries. One part of the study also examines possible changes in how high-performing students answer student questionnaires, from PISA 2003 to PISA 2012. No discernible patterns of change were identified, which indicates that the high-performing group has not changed substantially regarding the aspects measured in both PISA questionnaires.

By identifying factors characterising high-performing students in mathematics, this study can also indicate aspects of mathematics teaching and learning that might help all students perform well. Through the study of students' answers to individual questions in the questionnaire, it is also possible to go beyond structural factors such as socio-economic status and take a closer look on attitudes, beliefs and behaviours that students bring to school. These are factors which can be addressed by teachers and other actors relevant for mathematics teaching and learning.

6.2 Introduction

High-performing students in mathematics have not been given much attention from a policy or research perspective in Sweden, and possibly not in the other Nordic countries, despite the fact that several countries have experienced a decline in TIMSS and PISA results for the most able students. There is however a growing concern over recruitment of students to tertiary science, technology, engineering and mathematics (STEM) streams. Initiatives have been launched with the intention to offer high-performing students better opportunities to grow according to their potential, and to increase the number of high-performing students in the Nordic countries. This chapter describes a study of the characteristics of

high-performing students in mathematics in the Nordic countries based on data from PISA 2003 and PISA 2012. Although the analysis compares high-performing students with students performing around the median in each country, this study is not primarily about “gifted” students. The overall aim is to identify what characterises successful students to inform the development of mathematics education in the direction of helping all students perform well. Furthermore, the aim is to go beyond constructs and factors that school cannot affect directly (such as socio-economic status) and consider what the students bring to school that the teachers and others working in schools actually can do something about.

Even though the PISA design, method and combination of instruments offer rich opportunities for the study of factors connected to the teaching and learning of mathematics (see e.g. Niss, Emanuelsson, & Nysström, 2013), a study based on PISA cannot determine the causal relationships between student characteristics and achievement. This study attempts, however, through a study of associations between performance and student characteristics, to offer a starting point for the identification of factors that can be important to improve the performance of students across all achievement levels. It also attempts to study changes in the characteristics of high-performing students over time, which might shed some light on why the achievement of the most able students in the Nordic countries has declined, according to PISA.

6.3 Background

The study presented here is motivated by the general potential value of identifying factors related to high achievement in mathematics, but it is also strongly motivated by the declining PISA and TIMSS results of all students in most of the Nordic countries, and of the top performing students in particular.

Table 22 summarizes the change in scores for different percentiles of student performance from PISA 2003 to PISA 2012.

Table 22: Change in percentiles between PISA 2003 and PISA 2012 (OECD, 2014)

	10th percentile		25th percentile		75th percentile		90th percentile	
	Score diff.	S.E.	Score diff.	S.E.	Score diff.	S.E.	Score diff.	S.E.
Denmark	-2	6.4	-9	5.3	-22	4.6	-25	5.2
Finland	-29	4.7	-25	3.9	-26	3.8	-23	4.6
Iceland	-24	4.3	-23	4.3	-22	4.0	-17	4.9
Norway	-3	5.6	-5	4.5	-8	5.0	-10	5.3
Sweden	-27	5.9	-31	4.6	-33	4.6	-34	5.1

Table 22 shows that the achievement of high-performing students, defined here as the lowest PISA score within the top 10% of the students in each country, has changed in the negative direction from PISA 2003 to PISA 2012 for all Nordic countries, and particularly for Sweden.

Using percentile is one way of identifying students performing at different levels. The PISA score at a particular percentile indicates the level reached by a certain percentage of students in a country. An advantage of studying percentiles is that it makes it possible, for example, to compare the best performing students in one country with the best performing students in another country. PISA also offers the possibility of identifying students performing at different levels through six proficiency levels which are based on specific scores on the PISA scale. These cut-scores are the same for all countries, and comparisons between countries and over time are possible based on the percentage of students reaching these scores. The use of such proficiency levels offers the possibility to analyse what students at each level know and can do. The study presented here is based on percentiles because it focusses on the best performing students in each country, irrespective of whether these students are high-performing compared to other countries or compared to a certain standard. Using percentiles to identify high-performing students also yields more similar sample sizes across countries than the use of proficiency levels.

6.3.1 Studies of high-performing students

Characteristics of high-performing students in the Nordic countries were also studied by Sulkunen, Nissinen and Kupari (2014), focusing achievement in reading and mathematics in grade 4. They compared questionnaire data for students from three achievement tiers based on the TIMSS and PIRLS international benchmarks, which are similar to proficiency levels in PISA. Top performers were defined as those reaching the advanced international benchmark, i.e. achieving at least 625 score points. Sulkunen *et al.* (2014) analysed the factors associated with low and top performance and used a preliminary collection of variables from the PIRLS and TIMSS background questionnaires (student, teacher and parent questionnaires) for the exploratory analysis. One of their objectives was to determine the background variables that significantly distinguish top performers from intermediate performers. For mathematics, they found no significant association between top performers and data from teacher and parent questionnaires. However, their analyses of data from student questionnaires revealed quite a few factors associated with both low and top performance in mathematics in all the four Nordic countries included in their study (Finland, Denmark, Norway and Sweden). Sulkunen *et al.* concluded that the results are consistent with those reported in previous studies and summarised that top performers in mathematics are characterised by high confidence in mathematics, a favourable home background and good numeracy and reading skills.

In order to further set the scene for interpreting and discussing the empirical study presented here, a simple literature review was conducted. The databases Web of Science, Scopus and MathEduc were searched using the descriptor “high-achieving mathematics”. A total of 87 unique hits were returned. On the basis of the titles and abstracts, 28 publications were considered potentially relevant, and 25 of these were possible to retrieve as full texts.

The identified research includes a study that empirically identified a positive relationship between students’ achievement and student self-confidence, enjoyment and value within countries, but a negative relationship between these variables across countries (Kim & Kim, 2010). Chung (2008) concluded that the following factors are associated with

mathematics achievement specifically for high-performing students in Korea: teacher's expectations, parent's involvement in school activity, self-confidence in the subject, attitude to the subject, reading time, parent's educational level and expectations. Similar results were found by Lim and Chapman (2015) in their study of Grade 12 students in a top pre-tertiary school in Singapore. Self-confidence and ease with mathematics were key factors in explaining the variation in mathematics achievement. Furthermore, Liu and Meng (2010) used data from TIMSS 2003 to investigate low- and high-performing students' mathematics self-concepts in three East Asian societies and in the USA. Across the four societies, the self-concept of high-performing students was significantly higher than that of low-performing students.

Sulkunen *et al.* (2014) and the present study focus on Nordic countries, while other studies have focussed on other sets of culturally related countries. Areepattamannil and Caleon (2013) investigated learning strategies among students in Shanghai-China, Hong Kong-China, Korea, and Singapore, using data from PISA 2009. They found that the frequency of memorisation strategies for learning mathematics was negatively associated with mathematics achievement, and that the frequency of control strategies was positively associated with mathematics achievement. The frequency of elaboration strategies, i.e. how often the students relate what they are learning to what they already know and can do, was negatively associated with mathematics achievement in two of the investigated countries, and not at all associated in the other two. Choi, Choi and McAninch (2012) compared the top 5% students with the other 95% in nine countries and also the Korean results (top 5%) with the results of top 5% students from the other countries. The high-performing group of students demonstrated more positive attitudes to mathematics, valued mathematics more and demonstrated higher self-confidence in their ability to learn mathematics. Further, compared to students from other countries (also high-performing), Korean students displayed lower self-confidence and valued mathematics less. In another study identified as relevant in this review, Cheng (2014) compared five Asian countries by examining the types of mathematics instructional practices that affected students' mathematics learning, using data from TIMSS 2011. The study used data from teacher questionnaires with a focus on high-performing

countries, but also used a sample of high-performing mathematics teachers and their students in these systems. Students in classes with an average score of at least 500 were selected, along with their teachers. The study found mixed results for the relationship between teacher actions and students' achievement. Lower achievement in Japan was, for example, associated with teachers frequently asking students to work out problems by themselves or with peers, and in Korea, it was associated with the teacher frequently asking students to explain their answers. Higher achievement was in Singapore and Japan associated with teachers frequently asking students to decide on their own procedures for solving complex problems.

Other studies on high-performing students have focussed gender differences. The study presented by Carr, Steiner, Kyser and Biddlecomb (2008) involved students in grade 2. They found that the percentage of girls was lower in the high-performing group compared to the low-performing group, and that boys with scores in the exceeds proficiency range on the mathematics competency test were more likely to use cognitive strategies correctly whereas high-performing girls were more likely to use manipulatives correctly. Several studies have concluded that gender differences in mathematics achievement are primarily found in the high-performing group. On the basis of several other studies Ganley *et al.* (2013, p. 1886) concluded that "A particular concern is that a larger gap exists at the top end of the distribution. That is, the highest performing boys significantly outperform the highest performing girls." Stoet and Geary (2013) analysed students' mathematics and reading performance in four PISA rounds (2000, 2003, 2006 and 2009), across all the participating countries. They found no gender difference in mathematics among the lowest performing students, and the gender difference was the largest among the highest performing students. For the 33 countries that participated in all four of the studied PISA assessments, the ratio of boys to girls among students achieving above the 95th percentile was in the range of 1.7–1.9:1, and among students scoring above the 99th percentile, the ratio was 2.3–2.7:1.

About half of the potentially interesting studies identified from the database search were not considered relevant to the study presented here. Most of these were excluded because they focussed on very young

children, on domains other than mathematics (reading or science) or on high-performing countries and schools rather than high-performing students. The review indicates that studies on high-achieving students in mathematics are scarce and that research focussing on the characteristics of these students in the Nordic countries is lacking.

6.4 Aim

The study presented here aims to identify characteristics of the best performing students in mathematics in the Nordic countries, based on data from PISA 2003 and 2012. The study is explorative and can best be described as a screening of a large amount of variables from PISA student questionnaires that addresses the following two research questions:

- How do the best performing students in mathematics in each country differ from median-performing students in terms of characteristics measured in PISA student questionnaires?
- Have the characteristics of the high-performing group in PISA changed from 2003 to 2012, and can these changes explain the decline in performance of the high-performing group?

6.5 Design and method

The overall study design involved comparing questionnaire results for the best performing 5% students in each country with the 5% of students performing at and around the country median. This yielded a high contrasting effect combined with large enough groups for meaningful inferential analysis. A sub-sample of 5% equalled an average of 220 students in PISA 2003 and 292 students in PISA 2012 for these countries. Only Iceland had a sub-sample of less than 200 students (168 and 176 for 2003 and 2012).

The focal groups were identified from an average of the five plausible values that constitute overall mathematics achievement in PISA, similar to the method used by Liu and Meng (2010). Thus, the high-performing groups of students were those with an average plausible value in the top 5% of the sample from each country. Similarly, the median-performing groups consisted of students with an average plausible value between the percentiles of 47.5 and 52.5.

To identify the differences between the groups and between the high-performing groups of PISA 2003 and 2012, statistical inference techniques were used to investigate whether sample differences could be expected to represent differences in the student populations represented by the samples. This was done by stating a null hypothesis and then investigating if the hypothesis could be rejected or not. For all variables analysed in this study, the null hypothesis stated that there was no difference between the way the high-performing group of students' answered questionnaire items and the way the median-performing group of students answered the same items. Differences between the groups were registered with two degrees of certainty, when the null hypothesis was rejected with 99 % certainty or more ($p < 0.01$) and when it was rejected with 95% certainty or more ($p < 0.05$). The inferential analysis was conducted using SPSS and comprised three steps. In step one data files with questionnaire data from PISA 2003 and PISA 2012 were prepared with newly added variables identifying high- and median-performing students. This step also included the calculation of proper weights to be used in the analyses. In step two inferential statistics was used to analyse data from the prepared data files. For questionnaire items with the highest measurement levels (interval and ordinal), a non-parametric method suitable for multiple ordinal options was used (Kolmogorov-Smirnov). Questionnaire items with a nominal measurement level were investigated using a chi-square test. Step three involved the comparison of PISA 2003 and PISA 2012 data and primarily identified and selected questionnaire variables found in both studies. Data for these variables were merged into the same file after modifying a few of the scales to make them comparable. The difference between the high-performing groups of students from the two PISA cycles were investigated as described in step 2.

6.6 Results

The screening character of this study generated a multitude of results, which could not be completely summarised within the limitations of this chapter. In the following sections an extensive selection of results are presented, primarily focusing on consistent results across the Nordic countries.

Each section begins with a table showing the results for some basic background data and the most interesting indices found in the PISA databases. These basic comparisons are supplemented by text and tables showing differences for individual items in the PISA student questionnaires, particularly questions that relate to aspects of students' beliefs and views and are relevant for mathematics instruction. Individual items are more narrow than indices, but they are generally easier to interpret and more stable across different PISA rounds.

6.7 PISA 2003

For PISA 2003, the analysis of questionnaire data for high- and median-performing students included 181 variables in total. Table 23 presents results for the most interesting indices, background variables and group descriptors, most derived from sets of questions in the questionnaire.

Table 23: Characteristics of high-performing students in mathematics compared to median-performing students in PISA 2003. Selection of indices for background, beliefs and attitudes. ↑ represents significantly higher values for high-performing students compared to median-performing students, and ↓ represents significantly lower values. Significance levels are indicated by ** (p<0.01) and * (p<0.05)

	Country				
	Fin	Swe	Nor	Den	Ice
Female students	↓**	-	↓**	↓*	-
Students born in the country	-	-	-	↑**	-
Mother born in the country	-	-	-	-	-
Father born in the country	-	-	-	↑**	↑**
Speaks language of test at home	-	-	-	-	-
Attended ISCED 0 for one year or less	-	-	-	↓**	↓**
Educational level of mother	↑**	↑**	↑**	↑**	↑**
Educational level of father	↑**	↑**	↑**	↑**	↑**
Highest parental education in years of schooling	↑**	↑**	↑**	↑**	↑**
Expected educational level of student	↑**	↑**	↑**	↑**	↑**
Expects to complete ISCED 5A or 6	↑**	↑**	↑**	↑**	↑**
Cultural possessions at home	↑*	↑**	↑**	↑**	↑**
Home educational resources	↑*	↑**	↑**	↑**	↑**
Index of home possessions	↑**	↑**	↑**	↑**	↑**
Index of socio-economic and cultural status	↑**	↑**	↑**	↑**	↑**
Positive attitude towards school	↑**	↑**	↑**	-	-
Positive student-teacher relation at school	↑**	-	↑**	↑**	-
Sense of belonging to school	-	-	-	-	-
Minutes of mathematics per week	↑**	↑**	↑*	-	↓**
Ratio of time for mathematics instruction to total instruction time	↑**	-	-	-	↓**
Relative time spent on mathematics homework	-	↓**	↓**	-	↓**
Perceived teacher support in lessons	↑*	-	-	-	-
Positive disciplinary climate in mathematics lessons	↑**	↑**	-	-	-
Interest in mathematics	↑**	↑**	↑**	↑**	↑**
Instrumental motivation in mathematics	↑**	↑**	↑**	↑**	↑**
Mathematics self-efficacy	↑**	↑**	↑**	↑**	↑**
Mathematics anxiety	↓**	↓**	↓**	↓**	↓**
Mathematics self-concept	↑**	↑**	↑**	↑**	↑**
Preference for control strategies when learning mathematics	↑**	-	↑**	-	-
Preference for elaboration strategies	↑**	↑**	↑**	↑**	↑**
Preference for memorisation strategies	-	↑**	↑**	-	-
Preference for competitive learning situations	↑**	↑**	↑**	↑**	↑**
Preference for co-operative learning situations	↓**	-	↓**	-	↓**

Not surprisingly, and consistent with the expectations based on previous research, the high-performing group demonstrates higher educational background, more cultural capital and higher socioeconomic status. This result is consistent across all the studied countries. For Finland, Sweden and Norway, we also found that high-performing students have more minutes of mathematics per week and that they have a more positive attitude to school. High-performing students in Finland, Norway and Denmark report a more positive student-teacher relation at school than the median-performing group, but no difference is found in terms of feelings of belonging to school for any of the five countries. Further, as expected, high-performing students are more interested in mathematics, consistently across countries, displaying higher instrumental motivation in mathematics, higher mathematics self-efficacy and self-concept. They report lower degrees of mathematics anxiety.

In the PISA 2003 questionnaire students were asked about their preference for different strategies in learning mathematics. In all countries, elaboration strategies are highly valued by the high-performing students. Elaboration strategies entail, for example, making connections to related areas and thinking about alternative solutions when solving problems in mathematics. Control strategies (planning, monitoring and regulation) are preferred by only the high-performing groups in Finland and Norway, and memorisation strategies are preferred only by this group of students in Sweden and Norway. Norway is the only country where all three strategies are preferred more by high-performing than by median-performing students. In all countries the high-performing students report a higher preference for competitive learning situations, and in three countries, they report a lower preference for co-operative learning situations than median-performing students.

The percentage of female students in the high-performing group is lower than that in the median-performing group in three countries (Finland, Norway and Denmark). This result aligns with the overall gender differences in mathematics achievement for Finland and Denmark, but not for Norway. According to the PISA 2003 international report (OECD, 2004), male students overall performed significantly better than female students in Finland, Sweden and Denmark; no overall gender differences

in achievement were detected for Norway; and in Iceland female students performed significantly better.

With regard to immigration and language, the analysis provides no, or very weak, indications of differences between high- and median-performing students in mathematics. Only in Denmark the high-performing group has a higher percentage of students born in country compared to the median-performing group. No such differences are found for mother's birth country, and differences for father's birth country are only found in Denmark and Iceland. The percentage of students speaking the language of the test at home is the same for both groups in all countries.

To better understand what these indices represent and to explore the characteristics of high-performing students, high- and median-performing students' answers to some individual questions will also be presented.

The student questionnaire contained questions about students' views of their school. High-performing students in all five countries report to a lesser extent than median-performing students that school has been a waste of time, and for three countries (Finland, Sweden and Norway), the former group felt that school has done little to prepare them for adult life. No difference is detected concerning feelings of being an outsider or being left out of things, but in all countries except Sweden the high-performing students agree to a lesser extent that school is a place where they make friends easily. In Finland, Sweden and Norway, the high-performing students also agree to a lesser extent with the statement that they arrived late for school in the past two weeks.

Student-teacher relations are viewed differently by the studied groups of students in Finland, Norway and Denmark. The pattern is most pronounced in Finland, where high-performing students show a more positive attitude towards their teachers on all five questionnaire items (Table 24). For Sweden and Iceland, no differences are found for any of the five questionnaire items on student-teacher relations.

Table 24: View of teachers among high-performing students in mathematics compared to median-performing students in PISA 2003. ↑ represents significantly higher values for high-performing students

	Country				
	Fin	Swe	Nor	Den	Ice
Students get along well with most teachers at my school	↑**	-	↑**	-	-
Most teachers at my school are interested in students' well-being	↑**	-	-	-	-
Most of my teachers really listen to what I have to say	↑**	-	↑**	↑**	-
If I need extra help, I will receive it from my teachers	↑**	-	-	-	-
Most of my teachers treat me fairly	↑**	-	↑**	↑*	-

Results pertaining to hours spent on schoolwork and other potentially important learning activities displayed some expected results. For example, high-performing students spend less hours in remedial classes in school, both overall and in mathematics (Table 25). For all countries except Finland, high-performing students in mathematics also spend less hours on homework in mathematics. The results show that these students in Finland and Norway participate more in enrichment classes at school in mathematics, but not more in out-of-school mathematics classes or in working with a mathematics tutor. For Norway and Iceland, the results even indicate lower participation in these activities among high-performing students.

Table 25: Mathematics work out of class for high-performing students in mathematics compared to median-performing students in PISA 2003. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
Hours spent on homework (all subjects)	-	-	-	↓*	-
Hours spent on homework in mathematics	-	↓**	↓**	↓**	↓**
Hours spent in remedial classes for mathematics at school	↓*	↓**	↓**	↓*	↓**
Hours spent in enrichment classes for mathematics at school	↑**	-	↑**	-	↓*
Hours spent working with a mathematics tutor	-	-	↓*	-	↓**
Hours spent attending out-of-school mathematics classes	-	-	↓*	-	↓*

Students were also asked how many minutes of mathematics class they had per week, and high-performing students in Finland, Sweden and Norway report more time in mathematics class than median-performing students in these countries. In Iceland, the difference is reversed. There is no difference in total number of instructional time per week for any country.

Students' ways of studying mathematics could be of specific interest when determining the characteristics relevant for the success of high-performing students. Table 26 shows that high-performing students report stronger agreement with thinking of new ways to get the answer when they are solving problems in mathematics. In all the countries except Denmark, they also emphasise trying to figure out which concept they still have not understood properly. Furthermore, the high-performing students try to understand new concepts in mathematics by relating them to things they already know, and when solving mathematics problems they search for more information to clarify the problem and often think about how the solution might be applied to other interesting questions. Interestingly, no difference is found between high- and median-performing students in terms of how much they think about how the mathematics they have learnt can be used in everyday life.

Table 26: Learning strategies in mathematics for high-performing students compared to median-performing students in PISA 2003. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
When I study for a mathematics test, I try to work out what the most important parts to learn are	-	-	-	↓*	-
When I am solving mathematics problems, I often think of new ways to get the answer	↑**	↑**	↑**	↑*	↑**
When I study mathematics, I make myself check to see if I remember the work I have already done	↑*	-	↑*	-	-
When I study mathematics, I try to figure out which concepts I still have not understood properly	↑**	↑**	↑**	-	↑**
I think about how the mathematics I have learnt can be used in everyday life	-	-	-	-	-
I go over some problems in mathematics so often that I feel that I could solve them in my sleep	-	↑**	↑**	-	-
When I study for mathematics, I learn as much as I can by heart	-	↑**	↑**	↑**	-
I try to understand new concepts in mathematics by relating them to things I already know	↑**	↑**	↑**	↑**	↑**
In order to remember the method for solving a mathematics problem, I go through examples again and again	-	↓**	↓**	-	↓**
When I cannot understand something in mathematics, I always search for more information to clarify the problem	↑**	↑**	↑**	↑**	-
When I am solving a mathematics problem, I often think about how the solution might be applied to other interesting questions	↑**	↑**	↑**	↑**	-
When I study mathematics, I start by working out exactly what I need to learn	↑*	↓**	-	↓*	-
To learn mathematics, I try to remember every step in a procedure	↑*	↑**	-	-	-
When learning mathematics, I try to relate the work to things I have learnt in other subjects	↑**	↑**	-	-	-

In Finland, Sweden and Iceland, high-performing students report higher number of students in the mathematics class than median-performing students. There are also several differences between the groups concerning how they think about their mathematics classes (Table 27). High-performing students display more competitive attitudes and report less enjoyment and expectations from group work.

Table 27: Attitudes to mathematics learning among high-performing students in mathematics compared to median-performing students in PISA 2003. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
I would like to be the best in my class in mathematics	↑**	↑**	↑**	↑**	↑**
In mathematics, I enjoy working with other students in groups	↓**	-	↓**	↓*	-
I try very hard in mathematics because I want to do better in the exams than the others	↑**	↑**	↑**	↑**	↑*
When we work on a project in mathematics, I think that it is a good idea to combine the ideas of all the students in a group	-	↑*	-	-	-
I make a real effort in mathematics when I work with other students	↑**	↑**	↑**	↑**	↑**
I do my best work in mathematics when I work with other students	↓**	-	↓**	↑*	↓**
In mathematics I always try to do better than the other students in my class	↑**	↑**	↑**	↑**	↑**
In mathematics, I enjoy helping others to work well in a group	↑**	↑**	↑**	-	↑*
In mathematics I learn most when I work with other students in my class	↓**	↓**	↓**	↓**	↓**
I do my best work in mathematics when I try to do better than others	↑**	↑**	↑**	↑**	↑**

6.8 PISA 2012

The number of student questionnaire items was substantially increased from PISA 2003 to PISA 2012. This was accomplished by a rotated design for some of the items in the questionnaire, i.e. a substantial part of the questionnaire was only answered by two thirds of the sample. The analyses for the purpose of this study therefore included 389 variables in total, and only a smaller set of these are presented here. Table 28 summarises differences between high- and median-performing students for a set of indices, background variables and basic characteristics.

Table 28: Characteristics of high-performing students in mathematics compared to median-performing students in PISA 2012. ↑ represents significantly higher values for high-performing students compared to median-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
Female students	↓**	↓*	-	↓**	↓**
Student born in the country	-	-	-	↑**	-
Mother born in the country	↑*	↑**	-	↑*	-
Father born in the country	-	↑*	-	↑**	-
Speaks language of test at home	-	-	-	↑**	-
Number of years in ISCED 0	-	-	-	↑**	-
Age when starting ISCED 1	↓**	-	-	-	-
Educational level of mother	↑**	↑**	↑**	↑**	↑**
Educational level of father	↑**	↑**	↑**	↑**	↑**
Highest educational level of parents	↑**	↑**	↑**	↑**	↑**
Highest parental education in years	↑**	↑**	↑**	↑**	↑**
Highest parental occupational status	↑**	↑**	↑**	↑**	↑**
Wealth	↓**	-	↓**	-	-
Father International Socio-Economic Index (ISEI)	↑**	↑**	↑**	↑**	↑**
Mother International Socio-Economic Index (ISEI)	↑**	↑**	↑**	↑**	↑**
Index of economic, social and cultural status	↑**	↑**	↑**	↑**	↑**
Number of books at home	↑**	↑**	↑**	↑**	↑**
Cultural possessions	↑**	↑**	↑**	↑**	↑**
Home possessions		↑**	-	↑**	↑*
Attitude towards school: Learning outcomes	↑**	↑**	↑*	↑**	↑*
Attitude towards school: Learning activities	-	↑**	-	↑*	↑*
Sense of belonging to school	-	-	-	-	-
Learning time for mathematics (minutes per week)	-	-	-	↓**	-
Out-of-school study time	-	-	-	-	↓**
Teacher–student relations	-	↑**	↑*	↑**	↑**
Teacher support	↑*	↑**	↑**	↑**	↑*
Interest in mathematics	↑**	↑**	↑**	↑**	↑**
Instrumental motivation in mathematics	↑**	↑**	↑**	↑**	↑**
Mathematics self-efficacy	↑**	↑**	↑**	↑**	↑**
Mathematics anxiety	↓**	↓**	↓**	↓**	↓**
Mathematics self-concept	↑**	↑**	↑**	↑**	↑**
Experience with applied mathematics tasks at school	↑**	↑**	↑*	↑**	↑*
Experience with pure mathematics tasks at school	↑**	↑*	↑**	↑**	↑*
Attributions to failure in mathematics	↑**	↓**	↓**	↓**	↓**
Perseverance	↑**	↑**	↑**	↑**	↑**
Mathematics behaviour	↑**	↑**	↑**	↑**	↑*
Mathematics intentions	↑**	↑**	↑**	-	-
Mathematics work ethic	↑**	↑**	↑**	↑**	↑**
Openness to problem solving	↑**	↑**	↑**	↑**	↑**
Number of days of arriving late for school	↓**	↓**	↓**	-	-
Number of whole school days skipped	↓**	↓*	↓**	↓**	-
Number of classes skipped	↓**	↓**	↓**	↓**	-

As expected, and similar to the results of PISA 2003, parents educational level, cultural capital, and socioeconomic status are higher for students

in the high-performing group compared to those of median-performing group. Similar to TIMSS 2003, the percentage of students and parents who are born in the country is larger in the high-performing group than the median-performing group in Denmark. No other differences concerning students' birth country are observed. However, differences in the percentage of parents born in the country are found for Finland and Sweden. The percentage of students speaking the language of the test at home does not show any consistent pattern of difference over the studied countries. In fact, only Denmark displays such a difference, with a higher percentage of Danish speaking students in the high-performing group.

In the PISA 2012 questionnaire, students were asked to answer yes or no to a set of statements about how they study and learn mathematics. For several of those statements, differences between the high- and median-performing groups are observed. The statements that show consistent differences between the groups over three or more countries are presented in Table 29.

Table 29: Learning strategies in mathematics (selection) of high-performing students compared to median-performing students in PISA 2012. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
When I study for a mathematics test, I try to work out what the most important parts to learn are	↓**	↓*	↓*	-	-
When I study for a mathematics test, I try to understand new concepts by relating them to things I already know	↑**	↑*	↑*	-	-
When I study mathematics, I think of new ways to get the answer	↑**	↑**	-	↑*	-
When I study mathematics, I make myself check to see if I remember the work I have already done	↓**	↓**	-	↓*	-
In order to remember the method for solving a mathematics problem, I go through examples again and again	↓**	↓**	↓**	↓**	-
When I cannot understand something in mathematics, I always search for more information to clarify the problem	↑**	↑**	↑**	↑**	-

On the statement *I think about how the mathematics I have learnt can be used in everyday life*, no differences are found between the groups, except

in Finland where the percentage of students agreeing is significantly higher in the high-performing group.

As expected, the high-performing group has a more positive view of mathematics, e.g. *I look forward to my mathematics lessons*. Questions about the views of mathematics held by friends and parents, however, do not show any such pattern. Consistent positive differences between the groups are observed for the following statements: *My parents believe that mathematics is important for my career* (significant for all countries except Iceland) and *My parents like mathematics* (significant for all countries, but only at $p > 0.05$ for Iceland). Agreement with the statement *Most of my friends work hard at mathematics* shows no difference for any of the countries, and the statement *Most of my friends do well in mathematics* shows differences only in Sweden. Confidence to perform particular mathematics items (self-efficacy) is also higher, for all items and all countries. Furthermore, the high-performing group displays more positive views of studying mathematics, consistently over questionnaire items and countries, for example in response to the statement *In my mathematics class, I understand even the most difficult work*.

Table 30 shows the different views of high-performing and median-performing students with regard to perceived control, attribution of success and work ethics. For example, high-performing students agree more that if they invest enough effort, they can succeed in mathematics. Further, these students do not attribute failure to bad guessing and they keep on studying until they understand mathematics material, more than median-performing students.

Table 30: Perceived control, attribution of success and work ethics for high-performing students in mathematics compared to median-performing students in PISA 2012. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
Perceived control. Agrees with the statement					
If I put in enough effort I can succeed in mathematics	↑**	↑**	↑**	↑**	↑**
Whether or not I do well in mathematics is completely up to me	↑**	↑**	↑**	↑**	↑**
Family demands or other problems prevent me from putting a lot of time into my mathematics work	↓**	↓**	↓*	↓**	↓**
If I had different teachers, I would try harder in mathematics	↓**	↓**	↓**	↓**	↓**
If I wanted to, I could do well in mathematics	↑**	↑**	↑**	↑**	↑**
I do badly in mathematics whether or not I study for my exams	↓**	↓**	↓**	↓**	↓**
Attribution. Agrees with the statement					
I'm not very good at solving mathematics problems	↓**	↓**	↓**	↓**	↓**
My teacher did not explain the concepts well this week	↑*	-	-	↑**	-
This week I made bad guesses on the quiz	↓**	↓**	↓**	↓**	↓**
Sometimes the course material is too hard	↓**	↓**	↓**	↓**	↓**
The teacher did not get students interested in the material	-	-	↓**	-	-
Sometimes I am just unlucky	-	-	-	↓**	↓**
Work ethics. Agrees with the statement					
I finish my homework in time for mathematics class	↑**	↑**	↑**	↑**	↑**
I work hard on my mathematics homework	-	-	↑**	-	-
I am prepared for my mathematics exams	↑**	↑**	↑**	↑**	↑**
I study hard for mathematics quizzes	-	-	-	-	↑*
I keep studying until I understand mathematics material	↑**	↑**	↑**	↑**	↑**
I pay attention in mathematics class	↑**	↑**	↑**	↑**	↑**
I listen in mathematics class	↑**	↑**	↑**	↑**	↑**
I avoid distraction when I am studying mathematics	↑**	-	↑**	-	↑**
I keep my mathematics work well organised	↑**	-	↑**	↑**	↑*

Student records of frequency of class work and outside-school work show some interesting differences between high- and median-performing students (see Table 31). In all five countries, high-performing students help friends with mathematics more frequently than their median-performing counterparts. Further, in several countries, high-performing students talk more frequently about mathematics with friends, take part in mathematics competitions and play chess. Programming computers is not more frequent in the high-performing group, according to this analysis. Finally, high-performing students report attending fewer hours of out-of-school lessons in the test language, mathematics and science for all studied countries (results not shown in table).

Table 31: Frequency of activities at and outside of school for high-performing students in mathematics compared to median-performing students in PISA 2012. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
I talk about mathematics problems with my friends	↑**	↑**	↑**	-	-
I help my friends with mathematics	↑**	↑**	↑**	↑**	↑**
I do mathematics as an extracurricular activity	-	-	↑**	↑*	-
I take part in mathematics competitions	↑**	↑**	↑*	-	↑**
I do mathematics more than 2 hours a day outside of school	-	-	-	-	↓*
I play chess	↑**	-	↑**	↑*	-
I program computers	-	-	-	-	↓*

The results presented in Table 32 show discernible patterns of differences between the high- and median-performing groups in terms of availability of ICT (information and communication technology) devices at home and at school, but high-performing students seem to find limitations of the computer as a tool for school learning more problematic. Another set of questions in the PISA 2012 questionnaire addressed the occurrence of different kinds of computer use by teachers and students during the last month. For most of these questions, no strong discernible patterns of differences are observed. One exception is that a significantly higher percentage of students in the high-performing group in Sweden, Norway and Denmark state that neither students nor teachers rewrote algebraic expressions or solved equations using a computer during the last month of mathematics instruction.

Table 32: ICT use and attitude among high-performing students in mathematics compared to median-performing students in PISA 2012. ↑ represents significantly higher values for high-performing students, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
Frequency of ICT use for entertainment	↓**	-	↓**	-	-
Frequency of ICT use at home for school-related tasks	-	-	-	-	↓**
Negative attitude focusing on limitations of the computer as a tool for school learning	↑**	↑**	-	↑**	↑**
Positive attitude towards computers as tools for school learning	-	-	↓*	↑**	-
ICT availability at home	-	-	-	-	-
ICT availability at school	-	-	-	↓**	-
Time of computer use	↓**	-	↓**	-	↓*
ICT use in mathematics lessons	-	-	-	-	-
ICT use at school	-	-	↓**	↓**	↓**
Age when first using a computer	↓**	↓**	↓*	-	↓**
Age of first access to the Internet	-	↓*	↑*	-	↓*

6.9 PISA 2003 and PISA 2012 comparison

As previously described, the final part of the analysis attempts to compare the characteristics of high-performing students in mathematics between PISA 2003 and PISA 2012. From the two databases, a total of 92 variables were identified as relevant to the comparative analysis. Differences in overall descriptions of the high-performing group, background characteristics and derived indices are presented in Table 33.

Table 33: Difference in descriptives of high-performing students in mathematics between PISA 2003 and PISA 2012. ↑ represents significantly higher values for high-performing students in PISA 2012 compared to PISA 2003, and ↓ represents significantly lower values

	Country				
	Fin	Swe	Nor	Den	Ice
Mother's highest schooling	↓**	↑**	-	-	↑**
Father's highest schooling	↓**	↑*	↑*	-	-
Number of books at home	-	↓**	↓*	-	-
Highest educational level of parents	↑**	↑*	-	↑**	↑**
Educational level of mother	↑**	↑*	-	↑**	↑**
Cultural possessions	↑**	↓**	↓**	↓**	↑**
Index of economic, social and cultural status	-	-	↓**	↑**	-
Family structure	↑**	-	-	-	-
Educational level of father	-	↑**	-	-	↑**
Highest parental occupational status	↑**	↑**	↑**	↑**	↑**
Home possessions	↓**	↓**	↓**	↓**	↓
Highest parental education in years	↑**	↑**	↓**	↑**	↑**
Home educational resources	↓**	-	-	↑**	↑**
Number of class periods per week in mathematics	-	-	-	-	-
Number of class periods per week, overall	-	-	↓**	↑**	↓**
Mathematics anxiety	↑**	↑**	↑**	↑**	-
Attitude towards school: Learning outcomes	-	-	-	-	-
Sense of belonging to school	-	-	-	-	↑**
Disciplinary climate	↓**	-	↑**	↑*	-
Instrumental motivation in mathematics	↑*	↓*	↓*	↓**	-
Interest in mathematics	-	-	↓**	-	-
Mathematics self-efficacy	-	↓*	-	-	-
Learning time for mathematics (minutes per week)	-	-	↑**	-	-
Mathematics self-concept	↓**	-	-	-	-
Teacher–student relations	-	-	-	-	↑*
Teacher support	-	-	-	↑*	↑**
Female students	-	-	-	-	-
Mother lives at home	↑**	-	↑*	-	-
Father lives at home	-	↑**	-	-	-
Mother completed ISCED6 or 5A	↑**	↑**	↑*	↑**	↑**
Father completed ISCED6 or 5A	-	-	-	↑**	↑*
Student born in country	-	-	-	-	-
Mother born in country	-	-	↓*	-	↓*
Father born in country	↓**	-	-	↓*	↓*
Speaks language of test at home	-	-	-	-	-

Although the results are mixed, they suggest that the educational level of parents is even more pronounced for the high-performing students in mathematics in PISA 2012 compared to PISA 2003.

Mathematics anxiety is higher in PISA 2012, and instrumental motivation is lower in three countries. Significant changes in the scores of single countries must be interpreted with caution, but it might be relevant that Norway is the only country where the learning time in mathematics has increased between 2003 and 2012. No significant changes can be detected in the percentage of students born in the country or the students speaking the language of the test at home, from 2003 to 2012.

Individual questionnaire items found in both PISA 2003 and PISA 2012, relevant to mathematics, deal with attitudes towards learning mathematics, confidence in solving specific mathematics tasks and students' thoughts about studying mathematics. These results are not presented in detail, but very few changes were actually found in this analysis. One exception is that high-performing students in Denmark expressed higher confidence in solving particular tasks in mathematics in 2012. No similar pattern is found for the other Nordic countries. Another pattern, consistent across all five countries, is that in PISA 2012, students expressed higher agreement with the following: the teacher shows an interest in every student's learning, gives extra help when students need it, continues teaching until the students understand, and gives students an opportunity to express opinions.

In Norway, Denmark and Iceland, high-performing students in mathematics report fewer problems with noise and disorder in the classroom, and in these countries students also report decreasing problems with teachers having to wait a long time for students to quiet down. In Finland and Sweden, no such changes can be detected.

6.10 Analysis and discussion

This study has investigated differences between high- and median-performing students' responses to student questionnaires in PISA, with the aim of determining the characteristics of high-performing students and the changes in these characteristics from PISA 2003 to PISA 2012. The

high-performing group is defined as the 5% highest achieving students in mathematics in the PISA samples of Finland, Sweden, Norway, Denmark and Iceland. This means that the patterns found concern the best-performing students in each country, which is not necessarily the best performing students overall, because of different achievement levels in different countries. The analysis has primarily focussed on patterns across countries and to a limited extent on patterns within countries.

As expected, across the studied countries and in the two PISA rounds, high-performing students have a stronger cultural and educational background and higher socio-economic status than students in the median-performing group. This supports the conclusions of many studies (see e.g. Chung, 2008; Sulkunen, *et al.*, 2014). Further, high-performing students are also more interested in mathematics, more motivated, have higher mathematics self-efficacy and self-concept and a lower degree of mathematics anxiety than the median-performing group of students. This is also consistent with results from previous research (Lim & Chapman, 2015; Liu & Meng, 2010). While we can assume some kind of causal mechanism between home background and good performance, the relationship is not deterministic and does not help teachers in terms of the steps that they can take to make students learn. More information is needed about what students bring to school, possibly from their home background, that the schools can build on. The causality between the general positive view of mathematics, motivation, self-confidence etc. of high-achieving students and their achievement is less obvious, but possibly more relevant. Performing well in mathematics can of course be expected to have a positive influence on interest and motivation, but the interest and motivation can also boost learning and performance in mathematics. The results in this study support the idea of motivational beliefs being an important part of high achievement in mathematics.

The high-performing groups have a higher percentage of boys than girls. The percentages of girls in the high-performing groups is significantly smaller than that in the median-performing groups in Finland, Norway and Denmark according to PISA 2003, and in all countries except Norway according to PISA 2012. No significant changes are seen in the percentage of boys and girls in the high-performing group between PISA 2003 and PISA 2012, for any country. The underrepresentation of

girls in the high-performing groups is also an expected result, based on previous research. Carr *et al.* (2008) found a similar difference in their study of grade 2 students, and Ganley *et al.* (2013) arrived at the same conclusion after reviewing several other articles. Even though the overall gender differences are small or insignificant in most Nordic countries, there is a difference among high-performing students. Mathematics has been viewed as a gendered domain, and at least in some respects a male-dominated one (Brandell, Leder, & Nyström, 2007; Brandell, Nyström, & Sundqvist, 2004). This is an issue that calls for change. This study shows no significant change in these gender differences over a period of 9 years, which is noteworthy.

Language and immigrant status are two student variables of substantial current interest. In this study very small differences were found between the high- and median-performing groups with regard to immigrant status (students and their parents born in country), in both PISA 2003 and PISA 2012. Only in Denmark, the percentage of students born in the country was higher in the high-performing group. The high-performing groups did not differ substantially from the median-performing groups in the language spoken at home. Across all five countries, possibly with the exception of Denmark, belonging to the high-performing group did not seem to have any specific connection to students being born in the country or the students speaking the language of the test at home, and this situation does not seem to have changed from 2003 to 2012.

Closer to the classroom, a noteworthy difference between high- and median-performing students pertains to how they study mathematics. In PISA 2003, for example, the high-performing students thought of new ways of answering a problem and tried to relate new concepts to what they already knew. For both PISA 2003 and PISA 2012 no difference was found, however, between the groups with regard to the importance of connecting mathematics to everyday life. High-performing students are less likely to revisit the examples repeatedly to learn mathematics. In PISA 2003 the high-performing students reported a more competitive attitude and lower expectations of group work. Results from PISA 2012 showed that high-performing students have more positive views of their mathematics classes, but even more interesting is that they report higher persistence in working with problems.

The causality between these findings and students achievement has not been established in this study. However, it is challenging to think about the possible positive effects that these views and behaviours could have on lower-achieving students and how teachers could introduce such a change. There are similarities between the ways high-performing students say that they learn mathematics and what is often regarded as best practice in the mathematics education community (see e.g. National Council of Teachers of Mathematics (NCTM), 2000). Results related to classrooms and teachers are interesting but also rather difficult to interpret because many of the high-performing students in the Nordic countries study mathematics in mixed ability classes. This could explain the lower expectations of group work found among these students, as they might find their peers to be unable to challenge their thinking. High-performing students' views about classrooms and teachers can also be interpreted as their personal perception of what happens in the classroom and what the teacher does.

This perspective is also relevant to another pattern found in this study: high-performing students are more positive towards their teachers, at least in some of the countries included. With PISA 2003, this difference was most pronounced in Finland as it can be seen for all five questionnaire items in this category. No difference was found for any of the items for Sweden and Iceland. The analysis of PISA 2012 shows a slightly different pattern. Several differences between high- and median-performing students were observed pointing to a more positive view of teachers among the high-performing students, also for Sweden. Only Iceland showed no significant difference on any of the questionnaire items about students' views of the teacher. This change is further supported by the comparison of data from PISA 2012 and PISA 2003 (Table 33). High-performing students' views of their mathematics teachers is significantly more positive 2012. In fact, this change in view of teachers is one of the few significant changes over time found in this study. It seems that the high-performing students' views of their mathematics teacher has improved e.g. concerning teachers' interest in every student's learning. The more positive view of mathematics teachers expressed by the high-performing students can be viewed as a confirmation of the teachers' importance for student learning, which has been emphasized in recent

years (see e.g. Hattie, 2009). High-performing students' improved view of teachers is even more interesting, given the fact that the achievement of these students has declined during the same time. It seems as though the more positive view of mathematics teachers is not accompanied by an improvement in achievement, but rather the opposite.

Specifically in PISA 2012, disturbances such as students cannot work well and there is noise and disorder seem to be less frequent in the mathematics classes where we find the most high-performing students. This pattern is not as pronounced in PISA 2003, where e.g. noise and disorder was significantly less frequent only for the Swedish high-performing group, compared to the median-performing group. This points to another classroom variable that might be significant for understanding the environment that supports learning, but again this study does not support any strong claims about causal inferences.

The analysis of PISA 2003 showed that high-performing students in Finland, Sweden and Norway reported engaging in more minutes of mathematics per week than their median-performing peers. However, compared to the median-performing group, the high-performing students in all the countries except Finland spent less time on mathematics homework, while not spending less time on homework in general (possibly with the exception of Denmark). It seems as though the high-performing students manage very well without homework, but this could also be interpreted as high-performing students managing to do what is expected of them during mathematics lessons. Not working on homework can be a sign of low expectations and less challenging learning opportunities for these students. In addition, the results from the analysis of PISA 2003 show that high-performing students attend enrichment classes in school more, in Finland and Norway. The number of students in mathematics class is not smaller for high-performing students. The role of tutoring is another current topic for discussion, at least in Sweden, but in 2003 there was no sign of high-performing students meeting with tutors out of school, at least not more than median-performing students. Positive effects of tutoring have been shown in other countries (Huang, 2013), but at least in 2003 there were no signs of this, according to this study.

Results related to ICT use are specific to PISA 2012, supporting the more general analysis of the relationship between computer use and achievement published recently (OECD, 2015). The high-performing group tends to find the use of computers for school learning more limiting than the median-performing group. Computer use in mathematics class does not seem to differ between the groups, however, which is expected given the previously discussed assumption that most high-performing students study mathematics in mixed ability classes.

The PISA 2012 student questionnaire also included a set of questions on perceived control of mathematics learning, attribution of success and failure and work ethics. Persistently, across questionnaire items and across countries, students exhibited a view of being in control of their performance, avoiding attributing failure to bad guessing and inherent ability and putting an effort into working hard to learn mathematics. This finding is in line with the theories of motivation and other beliefs supporting learning and achievement.

To close this section highlighting and discussing a selection of results from the screening of high-performing students we give some attention to the part of the study addressing differences between PISA 2003 and PISA 2012. The number of variables possible to include in this analysis was limited and the results must of course be viewed with that in mind. Only two salient patterns of change were found in the questionnaire data for high-performing students from PISA 2003 to PISA 2012. The change towards a more positive view of the mathematics teachers has been reported and discussed above. In addition, parents' education level was higher in 2012 than 2003, though the measure of possessions in the home was lower. The results on indices (such as the possessions index) are problematic and it is unclear whether these indices can actually be compared from one PISA study to another. It is possible that the individual items constituting the index may change, and the scaling of the index is not made with trend measurement in mind. The analysis of individual questionnaire items is safer, since the students have actually been presented the same question, with the same answer options. The attempt to analyse changes in students' answers to the PISA questionnaire has revealed how problematic this is, primarily due to changing questionnaires.

6.11 Further research

The screening character of this study has affordances and constraints and can hopefully prompt other research projects. The design and methodology used here can be applied to other groups of students. It would be interesting to compare questionnaire data for a low-performing group of students with those for the median-performing group. This could not only confirm some of the results in this study, which are equally important for that achievement range, but also reveal other patterns and characteristics of low-performing students that could be addressed in school. Further, it is important to probe deeper into the factors that might have an influence on achievement and empirically determine the factors that need to be considered to boost achievement for all students. This is based on the assumption that the characteristics of high-performing students are relevant to their success and can be achieved by all students to some extent. A deeper study also includes connecting variables and conducting more theory-based studies, which might provide insights into the mechanisms for improving student learning.

6.12 Conclusions

To conclude, this study has identified many differences between high- and median-performing mathematics students, using PISA data. Students with very good results in the PISA tests generally have a more privileged background, such as parents with a higher educational background. Girls are consistently underrepresented in the high-performing groups across the five countries included in this study. High-performing students also study and learn mathematics differently, and these variables may be candidates for successful approaches to the learning and teaching of mathematics. In addition, high performing students' perception of control, attribution of success and other motivational beliefs, as well as their perseverance and positive attitudes to mathematics, have been confirmed by this study. These factors are most likely very important for student learning and pedagogically significant for all stu-

dents and teachers. There are also areas where no differences were observed, such as the percentages of students born in the country and students speaking the language of the test at home.

Through these descriptions a set of characteristics of high-performing students emerge, answering one of the research questions in this study. The other research question concerned changes in the characteristics of the high-performing group between two PISA cycles. Such changes could help us understand why the achievement of the highest performing students in the Nordic countries has decreased. The comparison has limitations, primarily because of the limited number of questionnaire items recurring in the two PISA rounds. However, the analysis of the common items and indices does not reveal a discernible pattern that can explain the decrease in achievement for high-performing students, leaving the question to be answered.

6.13 References

- Areepattamannil, S., & Caleon, I. S. (2013). Relationships of Cognitive and Metacognitive Learning Strategies to Mathematics Achievement in Four High-Performing East Asian Education Systems. *Journal of Genetic Psychology, 174*(6), 696–702. <http://dx.doi.org/10.1080/00221325.2013.799057>
- Brandell, G., Leder, G., & Nyström, P. (2007). Gender and mathematics: recent development from a Swedish perspective. *Zentralblatt für Didaktik der Mathematik, 39*(3), 235–250. <http://dx.doi.org/10.1007/s11858-007-0025-4>
- Brandell, G., Nyström, P., & Sundqvist, C. (2004, Juli). *Mathematics – a male domain?* Paper presented at the International conference on mathematics education (ICME-10), Köpenhamn.
- Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61–75. <http://dx.doi.org/10.1016/j.lindif.2007.04.005>
- Cheng, Q. (2014). Quality Mathematics Instructional Practices Contributing to Student Achievements in Five High-Achieving Asian Education Systems: An Analysis Using TIMSS 2011 Data. *Frontiers of Education in China, 9*(4), 493–518.
- Choi, K., Choi, T., & McAninch, M. (2012). A comparative investigation of the presence of psychological conditions in high achieving eighth graders from TIMSS 2007 mathematics. *ZDM – International Journal on Mathematics Education, 44*(2), 189–199. <http://dx.doi.org/10.1007/s11858-012-0401-6>

- Chung, P. (2008). Effect of Educational Contextual Variables on Mathematics Achievement of the Different Ability Groups in Korea. *Journal of Educational Evaluation*, 21(3), 23–41.
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An Examination of Stereotype Threat Effects on Girls' Mathematics Performance. *Developmental Psychology*, 49(10), 1886–1897. <http://dx.doi.org/10.1037/a0031412>
- Hattie, J. (2009). *Visible learning*. London: Routledge.
- Huang, M.-H. (2013). After-School Tutoring and the Distribution of Student Performance. *Comparative Education Review*, 57(4), 689–710. <http://dx.doi.org/10.1086/671346>
- Kim, K., & Kim, S. (2010). A Cross-national Analysis on the Relationship Between Achievement and Affective Characteristics in Mathematics and Science. *The Journal of Curriculum and Evaluation*, 13(3), 179–208.
- Lim, S. Y., & Chapman, E. (2015). Identifying affective domains that correlate and predict mathematics performance in high-performing students in Singapore. *Educational Psychology*, 35(6), 747–764. <http://dx.doi.org/10.1080/01443410.2013.860221>
- Liu, S., & Meng, L. (2010). Re-examining factor structure of the attitudinal items from TIMSS 2003 in cross-cultural study of mathematics self-concept. *Educational Psychology*, 30(6), 699–712. <http://dx.doi.org/10.1080/01443410.2010.501102>
- National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston: NCTM.
- Niss, M., Emanuelsson, J., & Nyström, P. (2013). Methods for studying mathematics teaching and learning internationally. In: M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick & F. K. S. Leung (Eds.), *Third International Handbook of Mathematics Education* (Vol. 27, pp. 975–1008). New York: Springer.
- OECD. (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD. <http://dx.doi.org/10.1787/9789264006416-en>
- OECD. (2014). *PISA 2012 results: What students know and can do - Student performance in mathematics, reading and science* (Volume I, Revised edition, February 2014). Paris: PISA, OECD Publishing.
- OECD. (2015). *Students, Computers and Learning: Making the connection*. Paris: OECD Publishing. <http://dx.doi.org/10.1787/9789264239555-en>
- Stoet, G., & Geary, D. C. (2013). Sex Differences in Mathematics and Reading Achievement Are Inversely Related: Within- and Across-Nation Assessment of 10 Years of PISA Data. *Plos One*, 8(3). Retrieved from <Go to ISI>://WOS:000316849200022. <http://dx.doi.org/10.1371/journal.pone.0057988>
- Sulkunen, S., Nissinen, K., & Kupari, P. (2014). Characteristics of low and top performers in reading and mathematics. Exploratory analysis of 4th grade PIRLS and TIMSS data in the Nordic countries *Northern lights on TIMSS and PIRLS 2011* (Vol. TemaNord 2014:528). Copenhagen: Nordic Council of Ministers. <http://dx.doi.org/10.6027/TN2014-528>

7. International large-scale educational assessments: Elephants at the gate?

By Johan Braeken, University of Oslo, Norway

7.1 Summary

International large-scale educational assessments such as PISA or TIMSS can be regarded as huge elephants arriving at the gates of our national educational system, capable of generating a potentially useful knowledge base for research and policy making, but at the same time also capable of trampling the current media and policy landscape, leading to controversy and hasty ill-formed reactions. To make the most out of these assessment-elephants and clarify some of the controversies, all stakeholders can benefit from a better understanding of the methodological challenges and complexities involved in large-scale educational assessment, and the criticisms that naturally emerge as a consequence of these complexities. Using PISA 2012 as a working example, a brief but balanced executive summary is given here framed around three key aspects of large-scale studies: design & data, statistical analysis, and communication. If these international large-scale assessments are to serve as supporting instruments for evidence-based educational policy decisions, we should strive to engage in an open-source transparent model of operations supported by meta-research and constructive debate on exactly these three aspects.

7.2 Introduction

In the recent years, there has been a huge increase in publicity and influence of international large-scale educational assessment projects, which in turn has led to considerable discussion and debate among the various stakeholders (i.e. policy makers, media, schools, teachers, students, and researchers within and outside academia). The Programme for International Student Assessment (OECD: PISA) can perhaps be considered a prime example. Unfortunately, the amount of constructive criticism on these projects is rather scarce, and the explosive combination of woolly official reports, oversimplified media-reporting, and speculative criticism has polarised the public debate: PISA is by some seen as utterly useless and misleading, whereas others defend it as if it were sacred. To be able to move away from this polarised dispute towards a more constructive debate, a better understanding is needed of both (i) the methodological challenges and complexities involved in large-scale educational assessment and (ii) the criticisms that naturally emerge as a consequence of these complexities.

Here, a *framework* is outlined that can offer guidance when assessing potential contributions and limitations of large-scale assessment projects. The framework, partially inspired by a discussion paper by Goldstein & Spiegelhalter (1996) on accountability of public institutions, makes a distinction between three key aspects: (1) design and data, (2) statistical analysis, and (3) communication. Although the framework can be applied to a range of projects involving performance indicators and accountability in the private and public domains, the PISA project has been used as a working example. Given that the technical reports of these large-scale projects run up to 400+ pages (see e.g. OECD, 2014), I do not claim to cover all the aspects in detail, but instead intend to provide an executive summary of common procedures and context for understanding general problems or critiques.

It is important to acknowledge from the start that international large-scale educational studies are not a simple endeavour. Research is always a compromise between what would be ideal and what is actually feasible given practical constraints in terms of time, expertise, and other resources. This implies that a number of alternatives and conscious choices

are considered in the development, implementation, and evaluation phases of the different parts of the larger project. It is imperative that these steps, processes, and decisions are well-justified, logical, transparent and open to scrutiny by all stakeholders. Clear and accurate documentation is the cornerstone of open communication and should explicitly cover what is being done, for what reasons, inherent strengths and limitations, and what can and cannot be done.

7.3 Design & Data

7.3.1 *Who gets to be tested?*

A constant concern of the public with regard to an international large-scale assessment project is that not every country is equally fair when setting up the sample of students that are selected to participate. Arguments such as “They do well on PISA because only their top students are being included, whereas we make an effort to cover the whole range of students” are all too common among teachers and administrators. Such comments cast doubt on whether the sample of students is indeed representative of the country’s student population and question the sampling design and procedures of the project. This can be partially to a lack of clarity in or understanding of these procedures.

In general, to ensure a representative sample of students, a complex multi-stage sampling procedure is used in these large-scale assessments: First schools are selected at random in proportion to their size (i.e. larger schools have a greater chance of being selected); Then, from within the selected schools, students are randomly sampled. PISA is for instance targeted at students aged 15 years and aims to survey roughly 4,500 students and 150 schools per country.

We should acknowledge that it remains difficult to completely standardise this procedure across so many different countries and varied educational systems (e.g., inclusive school systems or systems with specialised schools for children with special needs), and that the projects rely on the participation of the selected schools and students. There is a limited local flexibility with regard to inclusion of individual

students at the school level for reasons of emotional or cognitive disability (e.g. dyslexia, dysgraphia or dyscalculia); Similarly, there is a limited local flexibility with regard to inclusion of specific schools at the national level (e.g. national coordinators can decide on excluding schools for the blind) (OECD: PISA 2012 Sampling guidelines, p. 21–24). However, these exceptions are only applicable to individual cases and are documented. Furthermore, control measures and procedures are in place to enforce adherence to design and prevent countries from systematically tailoring their own sample (e.g., the sampling of schools is done by a central OECD commission based on official school statistics and not by local authorities, and there are sampling monitoring systems, verification checks, and required participation rates). This operational system is designed to ensure that representativeness can be obtained within acceptable and realistic boundaries.

7.3.2 What are they tested on?

Another obvious question to ask with respect to large-scale assessments is what are they actually testing? The most common criticism around this aspect is that “The assessment neither tests all relevant content domains nor all the important educational outcomes.” This demand for complete coverage is of course not a realistic expectation, but it is partially a consequence of differences in perspective and ideology and partially a reaction to the loud sales-and-marketing pitch associated with these projects. The claim that PISA assesses “the extent to which students worldwide can apply knowledge to real-life situations and be equipped for full participation in society” (<http://www.oecd.org/pisa/aboutpisa/>) is an example of the latter. Hence, it is important to not be deceived by short-sighted thinking and look further into how the actual test is build up: What is the conceptual framework behind the test and what do the test items look like (i.e. the operationalisation of the conceptual framework)?

PISA is not formally based on the educational curricula of the different countries, but opted for a more independent conceptual framework. This framework has been developed through a cross-country collaboration among expert groups. For instance, mathematics in PISA is heavily influenced by what is sometimes called realistic arithmetic, a perspective

on mathematics connected to the well-known Freudenthal Institute for Science and Mathematics Education (see e.g., Stacey & Ross, 2015). This implies that the mathematics items in PISA usually involve a common real-life context from which information first needs to be extracted and encoded into mathematical operations. However, this does not imply that PISA items are completely disconnected from the curriculum as this very approach has been adopted as a core principle by many national curricula of mathematics. Those with a strict view on mathematics may find that PISA is more language driven or more of an IQ test. While the former might be true, one could argue that the process of extracting and encoding information as input for mathematical operations is part of the competencies needed for applying mathematics. The latter remark about IQ is slightly misguided because with any general educational test, results across different topics tend to correlate quite highly in the broader population. In fact, it was this observation that gave rise to the idea of IQ and the study of intelligence in the first place (Spearman, 1904). There are not a lot of students who are absolutely brilliant in mathematics, but are completely at loss in language.

When discussing what a test measures, it is important to look beyond the label of an educational outcome and be aware of the conceptual framework that underlies the test and how it has been operationalised in form of actual test items that students need to solve. Differences in implementation choices can also partially explain some differential results across large-scale assessment projects. For instance, Western countries perform relatively better on PISA mathematics than on the mathematics component of the Trends in International Mathematics and Science Study (IEA: TIMSS), whereas the reverse holds for Asian and East-European countries. This difference is partially due to the definition of mathematics in the two frameworks: PISA has a larger representation of data-analysis and probability questions and less algebra than in TIMSS (see e.g. Wu, 2010).

Test items in these assessments are usually the result of joint research development efforts with item writers from different countries, field trials to verify the psychometric quality of test items, blueprints to ensure coverage across the chosen conceptual framework and so on. Still, the final selection of items is bound to generate discussion or arguments,

as with any test. Ask any teacher who has designed an exam test or read Champagne and Pearson (2003) who vividly discuss the reception that the release of items in the National Assessment of Educational Progress (NAEP) typically meets in the USA. While these discussions are unavoidable, it is important to ensure that they do not turn into meaningless criticism or item bashing. A constructive debate towards the optimisation of future items is vital. To this end, all large-scale projects, not only PISA, may want to invest in a more systematic generation of the item pool as in more modern test design (see e.g. Gierl & Haladyna, 2012).

Furthermore, you can argue that even a systematic follow-up is not a default part of the large-scale assessment projects. To further validate the conceptual framework and individual test items, they rely on the efforts of individual researchers in individual countries. This is a pragmatic time-constraint-driven choice, but perhaps also unfortunate when combined with the sometimes excessive marketing talk associated with these projects. Can they deliver what they promise? Hence, more validation studies are needed if you want to pinpoint test labels and link them to real-world outcomes and behaviour. Having said that, all additional steps needed to optimise measurement do require extra time, research and investments, which in turn might delay the regularity of the assessments. Balancing optimality and practicality is never an easy task.

7.3.3 Design Choices & Grain size of information

People are usually less aware of the fact that students participating in international large-scale assessments do not complete the full set of items in the test. While this may seem like an odd testing practice from an individual perspective, it is important to remember that PISA is targeted at the higher system level. Because one is less interested in the performance of an individual student, the test design can instead be tailored towards efficient and effective assessment of the average population effects at the country level.

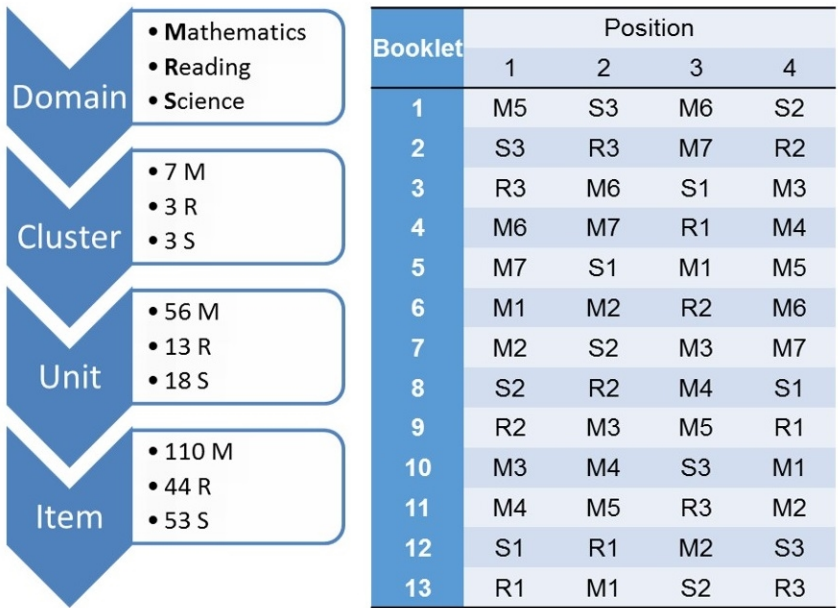
Ideally, one would like to assess all students in every country on the whole test. However, practical and logistics constraints imply that it is more realistic to assess a representative sample of students in each

country. The same practical and logistic constraints make it more realistic to assess each student on a sample of items from the whole test. PISA, for instance, has the ambition to assess three main content domains – mathematics, reading, and science. Any test that intends to cover these three broad topics in a representative, reliable and valid way will necessarily consist of a large number of test items. A huge test of 200+ items will place a high burden on an individual student and risk to confound student proficiency with test fatigue (i.e. items appearing later in the test are not solved at full cognitive capacity), and participating schools might have to waive several full days from their regular programme. This produces some tension between adequate test coverage and realistic test time.

Design Choices

This tension is resolved by making use of efficient design schemes for assessment. A reasonable test time frame is set (e.g. PISA is set at 2 hours test time) and items are distributed across different test booklets that each can be solved within this time frame. PISA has a so-called rotating matrix design for the test booklets: the PISA 2012 cycle, for instance, had 13 booklets consisting of 4 item clusters each (see Figure 41). Each cluster appears exactly once in each of the four possible booklet positions (i.e. counter balancing test fatigue – positioning effects) and each cluster pair occurs only once together in a booklet. The distribution of the items across booklets should ensure a sufficient and efficient exposure of items to the student sample. The application of this design results in about 1,300 responses per item in a country (ca. 30% coverage of total number of students) and 64 items solved by each student (ca. 30% coverage of total number of items).

Figure 41: Booklet Design of PISA



7.3.4 Information grain size

Thus by design, an individual student does not take the whole test. The booklet designs offer a way to efficiently and effectively capture information at the country level by partially overlapping data blocks across all individuals. Statistical models will be able to derive the average country performance based on the available data given sufficient information overlap. An analogy of how we are able to grasp the full meaning of a sentence even when some of the letters in different words are missing can help to conceptually understand the approach:

Mice do not like c___kies as much as ch___se because c___kies hurt their t___th.

Thus, the fact that individual students do not fill in the whole test is not crucial for establishing the overall system level performance. If the gaps in individual-level test performance are strategically placed, the higher order pattern of results can still be derived without first having to explicitly fill in values for individual missing item responses.

The biggest consequence of such designs is that the amount of information, and therefore the measurement precision, decreases when moving through the hierarchy (country, school, student). Average country performance is measured with the highest precision, and individual student performance with the least precision. This is also the reason why international large-scale assessment programmes typically do not report or provide feedback on the performance of an individual student or school. Under the matrix sampling design for items, an individual only responds to a small subset of items from the whole test, and these few item responses of that single individual would form the only information basis for making an estimate of the individual's proficiency. The same reasoning applies to individual schools, which only have responses from a small set of students for different subsets of items of the whole test. In contrast, for a country we have item responses of a whole sample of students across the full item set, allowing for a much more precise measure of average proficiency at the country level.

7.4 Statistical Analysis

The large-scale aspect of projects such as PISA (i.e. handling a lot of data and a multitude of possible comparisons with complex statistical models efficiently under time constraints) involves having to work outside the comfort zone of typical statistical applications. There will be some tension between the ideal theoretical approach to the analysis and the actual pragmatic solutions that are feasible in practice. The design choices made will also be directly reflected in the chosen statistical approach and procedures. The grain size of information will be directly related to the level of inferences that can be drawn given the collected data. Before providing a global overview of the procedures in PISA (with national and international calibrations of a measurement model with structural links to background

variables and the generation of plausible values for secondary data-analysis), the debate on truthfulness and defensibility of the statistical modeling is redirected towards usefulness and robustness of the modeling results, with special attention to transparency and reproducibility.

7.5 Models as useful approximate summaries

In a working paper Svend Kreiner, a retired biostatistics professor at the University of Copenhagen, strongly criticised the statistical measurement model in PISA arguing that it should be abolished because the model did not fit the data in any country, remaining items still showed variable patterns across countries, and resulting country rankings were too unstable to be trusted. The paper (Kreiner & Christensen, 2014) was picked up by the academic journal *Psychometrika* to generate more public debate around the methodology used in large-scale educational assessments. Yet, PISA's reaction was rather shallow and mostly dismissive, which left a number of stakeholders in doubt. The response from the academic community was also rather scarce. This disinterest was due to a multitude of reasons.

First, Kreiner's critique overplays the problems as he purposefully narrowed down the information window. The analysis was based on a limited dataset consisting of only those students who filled in one test booklet containing 20 reading items and disregarded all the other available information on students' performance or background. Thus, it comes as no surprise that the resulting country rankings were fairly unstable in his analysis. Slightly overstated, if we had to relate this back to our example of the sentence with missing letters, it would amount to singling out two words in isolation, omitting the context, but still having to guess the full sentence. Secondly, the criticism that the measurement model does not perfectly fit the data is no big surprise to anyone active in statistical modelling of real data and is mostly an attack on a straw man. A quote credited to the famous statistician George Box states that "All models are wrong...but some are useful." While critics usually focus on the first part of the quote, without offering practical alternatives, defenders highlight the second part of the quote, without offering tangible evidence.

In contrast, the relevant question to ask is “How wrong do they (models) have to be to not be useful?” (Box, G.). In this regard, Kreiner’s paper does have important signal value: Currently the projects mostly focus on optimising efficiency (i.e. “do more with less”), and methodological choices seem motivated by pragmatic reasons. The question then is how far they may go without putting validity of inferences at risk and which screening and modeling procedures should be used. Supporting evidence from robustness and simulation studies is needed and should be prioritised by all stakeholders (i.e. parties organising the large-scale projects, academics and funding agencies).

In a *robustness* study a small deliberate change is introduced either in the data or in the analytical procedure, after which it is checked whether the overall pattern of results remains stable. Obviously, the credibility of the inferences and conclusions severely decreases if the results do not prove to be robust to such small tweaks in the methodological procedure. In a statistical *simulation* study, artificial data are generated according to a statistical model that gives an a priori prescription of the outcome results. Because, in contrast to real-life applications, the underlying true results are now already known from the start, the simulated dataset can be used to evaluate the performance of the default analytical procedures against an objective true standard, something which can never be established with actual data. Hence, a simulation study verifies to what extent we can retrieve what we generated, hereby providing a performance standard under ideal conditions (or even under ill-defined conditions if the simulation model is theoretically inconsistent with the analytical procedure). Both robustness and simulation studies provide a framework for designing controlled experiments that evaluate the performance of analytical procedures and the stability of related results. However, to facilitate such studies and encourage an open and constructive debate, an increase in transparency is needed with respect to the procedural system used in PISA.

7.6 Reproducibility and Transparency

A technical report (OECD, 2014) describes PISA's divide-and-conquer computational analysis strategy, which is common to most assessments that deal with big data problems (e.g. 50+ countries, 500,000+ students according to the report). This strategy involves recursively breaking down a huge problem into many smaller sub-problems such that they become more manageable. Solutions to sub-problems are then combined to assemble a solution to the overarching initial problem. The side-effect of such a pragmatic approach is that it is susceptible to the "but in the ideal-case" criticism and that it has the tendency to end up looking like a non-transparent black box. Although the general methodology and principles are well-known and established, the lack of operational details in the available technical reports makes it hard for even expert outsiders to exactly replicate the procedures used to produce the reported results in the projects. This hinders a proper and fair discussion of the technical operations. This can be partially due to the challenges faced by the specialised companies and institutes involved in these analyses in describing and documenting the whole machinery, procedures, analytical steps, and decisions in full detail.

Yet these large-scale educational assessments are sponsored by local governments and are essentially public goods. Therefore, it is paramount that they do meet the minimum scientific thresholds of reproducibility and transparency. The extensive documentation on the conceptual framework and the easily accessible processed datasets are a good start, but similar standards and details are lacking on the statistical data analysis. Such perceived lack of reproducibility and transparency gives rise to a scepticism nicely summarised by Leamer (1983): "Hardly anyone takes data analyses seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analyses seriously." Thus, to safeguard the credibility of large-scale assessments, sponsoring governments should not hesitate to push through to make public the raw data, statistical analysis code, and psychometric decision steps.

Such openness, as for instance in the development of open-source software, will benefit the project as a whole by encouraging adherence to standards, clean design, reliability and maintainability. As there is no

longer need to speculate on what happens inside the black box, procedural criticism can be judged fairly. Quality control no longer becomes the sole responsibility of experts who are part of the project but is distributed among the larger academic community. The availability of a default working technical system strengthens the market position of the participating governments and encourages potential commercial partners to search for added value when entering the project bids. It does not make sense that both partners have to keep investing in reinventing everything from scratch every new cycle of the project. Ensuring reproducibility and transparency will increase credibility, foster meaningful discussion, and drive research progress and developments that will lead to better value for money and better use of the gathered data.

7.7 Global Procedural Overview

The currently available technical documentation on PISA does not facilitate replication of their in-between steps and analyses, though it does summarise the global procedure that is followed, and note that the statistical methodology and techniques used seem to resemble those in other large assessment projects such as TIMSS or NAEP.

7.7.1 *National calibrations*

In the first phase of PISA analyses, items are screened for psychometric quality, anomalous items are flagged and item profiles across the whole test are compared across the different countries. Decisions and procedures in this phase should ensure that one can proceed with a solid and sufficiently invariant item base that can be used as foundation for further cross-country comparisons. This implies that items that behave very differently in one country compared to other countries (i.e. so-called DIF: differential item functioning) will be omitted as comparison base for that country or for all countries (if the item proves troublesome in several countries).

7.7.2 *International calibration*

The first phase provides a working assumption that a reasonable ruler can be established for the measurement of student performance in similar units relative to a given reference point across all countries. In a second stage, a statistical measurement model is estimated across all countries, enabling straightforward cross-country comparison of estimated average student performance. The measurement model is further extended by a population model component, which derives country-specific system-level structural links between student performance and background information such as gender, socio-economic status or responses to other questions from the individual and teacher survey.

7.7.3 *Plausible values for secondary analyses*

In more traditional small-scale projects, researchers would first compute a performance score for each individual student, after which population statistics would be estimated by, for instance, averaging the scores of students in a particular country or correlating the students' scores to the students' socio-economic status. This works well if each individual student responds to the same large set of items because then considerable information is present per individual, and student performance can be accurately estimated at the individual level.

However, this traditional approach does not work for large-scale assessments, as every student responds only to a small subset of items and his/her responses on other items are missing. Thus, the set of available item responses varies per student, and the information base per individual has become too small for individual scores to be reliably estimated, leading to large measurement error at the individual level.

Conveniently, because the structurally missing gaps do not alter the overall picture (cf. design section), population statistics can still be statistically derived from the representative sample of item responses, without having to estimate a score for every individual student. Less conveniently, most educational researchers interested in pursuing secondary analyses of large-scale assessments data currently do not have the psychometric expertise or statistical machinery to work directly on the item

response data. The absence of aggregate measures of student performance at the individual level forms the core of their problem. Hence, to not impede further applied research, large-scale projects usually provide a dataset with so-called plausible values for student proficiency.

This plausible-values dataset enables researchers to reach valid inferences at the population level by applying simple techniques that are more in line with the conventional approach of first filling in a value for each individual student and then computing the population statistics based on these imputed student scores. Instead of a single unreliable estimate for the score of an individual student, the dataset acknowledges the inherent measurement uncertainty at the individual level by providing a set of 5 plausible values for each student's proficiency. Hence, one will actually have 5 new datasets, where for every individual student one gets an estimated proficiency score. The procedure would involve running the same statistical analysis per dataset and then pooling their results together to get a single outcome that takes into account the uncertainty of working with plugged-in estimated student proficiencies.

The range of such plausible values would still be quite large given that every student has filled in only a few items. To gain precision, strength can be borrowed from the background knowledge available at the population level. That is, if an individual student performed poorly on the small subset of items that s/he completed, the estimated set of plausible values for this student's proficiency would still be shifted upwards if students with similar background characteristics tended to perform much better on average in the population. One might argue that this is unfair to individual students because students with the same test score may be disadvantaged/compensated depending on whether they belonged to a low/high performing group. However, it is important to remember that we are not targeting an individual student, but that the goal of large-scale assessments is to obtain accurate population level estimates. Although occasionally the proficiency of an individual student might be slightly biased, on the whole we will be much more accurate when using collateral background information. The focus is on the population system of countries, not on a single individual student. In fact, adding more information at the population level can never hurt the cause, on the contrary. Hence, plausible values should not be considered as regular

point estimates for individual student proficiency, but as a means to an end to facilitate conventional secondary analyses that do accurately reflect the proficiency of the underlying population (see e.g. Mislevy, Beaton, Kaplan, & Sheehan, 1992).

Although plausible values are put in place for the benefit of researchers doing secondary analyses on the project data, many of these intended end-users are not well-trained in how to handle plausible values. Hence, training on running and reporting proper statistical analyses using the plausible values technique should be provided to secondary data analysts working with the data from large-scale assessment projects (for more information, see e.g. von Davier, Gonzalez, & Mislevy, 2009).

The misapplication of the plausible-values technique notwithstanding, there are some moral and practical dilemmas with respect to the limits of the approach where the general public and even methodologists might not entirely be comfortable with the technique. In principle, you can provide a set of plausible values for mathematics even if a student has not been tested on the mathematics part of PISA; scores would be based on system-level information, and on population-level relations between mathematics performance and the student's background variables and item responses on the science and reading parts of PISA. This is a theoretically sound approach, but it does call for some form of validation studies to provide supporting empirical evidence for the given imputation population model. Only recently a few studies have appeared on the impact of, for instance, missing or low quality background data (for an early example, see Rutkowski & Rutkowski, 2010) on the inferential quality of plausible-value-based analyses.

7.8 Communication

7.8.1 *League Tables*

A major challenge for most international large-scale assessment projects is the effective *communication* of results to the larger public. Unfortunately, simplified league tables and rankings of participating countries according to average performance have dominated communications. The

rankings tend to induce a false sense of differentiation ability, whereas the uncertainty around the countries' average performances does not allow for such strict distinctions. For instance, a country ranked 10th might not be performing significantly different from a country ranked 18th. Goldstein and Spiegelhalter (1996) warn against attaching excessive importance to rankings: "An overinterpretation of a set of rankings where there are large uncertainty intervals, ..., can lead both to unfairness and to inefficiency and unwarranted conclusions about changes in ranks." People are quick in reading causality in rank differences, even if the underlying differences in average performance are small and meaningless. One might argue that the big organisations behind large-scale assessment projects so far have largely neglected the crucial importance of clearly communicating their results. In fact, it does not even have a place in the technical report. It should be acknowledged that some attempts have been made to provide "upgraded" league tables that do communicate uncertainty by means of error bars around the average performances. However, as one of my students so eloquently stated, "We all know that nobody understands error-bars." It would make more sense, perhaps, to consider a presentation format that communicates differential clusters of countries instead of individual rankings.

In principle, we can of course all relate to these rankings with the underlying idea that those that lag behind can learn from the vanguard. In practice, the need to focus on and report these rankings can be questioned as it quickly turns into a conversation about winners and losers, driven by different valorisations, expectations and ulterior motives. One of many examples can be found in the reactions to the initial PISA 2000 results in Norway (for an overview see Koren, 2015 and references therein). Up to that point, international evaluations of the educational system (e.g. Norwegian results in the TIMSS 1995 cycle) had not received much media attention, and the general public and officials alike considered the national schooling system to be excellent. However, in the PISA 2000 rankings, Norway turned out to rank in the middle whereas neighbours from Finland were at the top. This "surprising mediocre" national performance attracted considerable media attention, and the Norwegian minister of education – who had assumed office just after the release of the results – was reported to have said the following: "Norway is a school

loser, now it is well documented. It is like coming home from the Winter Olympics without a gold medal” (Aftenposten). This illustrates that PISA results can and will be used as a pressure-instrument by the media to elicit strong statements from government officials, as a political weapon to disqualify the competence of the previous government, and as a trigger to launch emergency policy actions to improve “rank”. It should be clear that this is not where the core potential contributions of international large-scale assessments lie.

7.8.2 Causality & context

When looking at the results across countries, it is important to realise that relative performances on these assessments primarily have signal value: Differences raise questions, but do not provide direct answers (for a similar issue in a medical context, see e.g. Dickson, 1995). Large differences should be examined and explained by placing them in a broader context. From a similar perspective, differential item functioning across countries is not necessarily a problem for these assessments, but a source of information to further explore patterns of cross-country differences. This is where the core potential contributions of international large-scale assessments are situated: locating differences and raising questions! The focus and discussion should shift from winning or losing towards what we can learn from these international large-scale assessments for our own specific national context.

Explaining these differences is not an easy task because it is unrealistic to identify a single defining cause or believe that one change will miraculously fix the educational system in every and each country, irrespective of their national context. Feniger and Lefstein (2014) formulated the key point quite comprehensively:

“...much greater care should be exercised in the interpretation and uses of international comparative testing programmes such as PISA. While it may be politically attractive and expedient to attempt to imitate the educational policies and structures of high-attaining systems, our analysis reinforces the argument that such cross-national policy borrowing will be ineffective without attending to the historical and cultural contexts in which those policies operate (Alexander, 2012)”.

Common sense dictates that we should guard against simplistic post hoc cherry picking in which we isolate and portray a single contributory determinant as being the necessary and sufficient explanation for cross-country differences. Yet, this is generally not how the interpretation, reporting and communication of results proceed. The excellent results of the Shanghai municipality in China in the latest PISA cycle have, for instance, become a source of huge controversy. Innumerable articles and commentaries have been written, each focusing on a single element in isolation and positioning this as an all-explaining cause: Cheating by the Chinese government (conspiracy theory), a highly selective provincial enrolment system eliminating rural immigrant students (structural system theory with large expected within-system differences) or the Asian culture that has a long history of testing and where education is highly valued and enforced by parents (cultural theory based on comparably excellent results by, for instance, Korea and Japan). Alexander (2012) provides a strong opinionated piece on the use, misuse, and abuse of results of international comparisons by researchers, media and policymakers, whereas Mislevy (1995) provides a more balanced perspective. Both pieces are excellent at raising the discussion about what we can and cannot learn from these international large-scale assessment projects and how they fit in the bigger educational puzzle.

7.9 Research Agenda

In this chapter, I tried to give a balanced overview of the complexities behind international large-scale assessments by summarising common procedures, context and useful references for understanding general problems or critiques associated with these projects. The selection of issues related to design, data, statistical analysis, and communication make it clear that a call for meta-research on exactly these three aspects of large-scale educational assessments is warranted, because – and this is really unfortunate – currently this type of meta-research is lagging behind, leaving three important questions largely unaddressed:

1. To what extent can you *generalise* results of international large-scale assessments?
2. How *robust* are the projects' results and related inferences to small changes and/or anomalies in the analysis framework?
3. How can the results be more effectively *communicated* to policy-makers and the larger public?

The study of international large-scale assessments is already a large-scale research project on its own, but one that should not be neglected if these assessments are to serve as supporting instruments for evidence-based policy decisions.

7.10 Conclusion

It is abundantly clear that international large-scale assessment projects are a huge endeavour where theoretical ideals are challenged by practical constraints and reality, and where one is confronted with difficult methodological choices, analytical complexity, ideological discussions and over-simplified interpretations of the results. A quote from an editorial by Dickson (1995) – originally from a medical accountability setting, but now adapted to the current educational context – provides the perfect summary conclusion: “But it should not be beyond the wit of the research community to produce data which for all their imperfections allow the public and *educational stakeholders* to ask why it should be that one *country* compared to another *consistently scores much higher on Mathematics / Science / Reading*. And it should not be beyond the intelligence of the rest of us to understand the limitations of this information.”

7.11 References

- Alexander, R. J. (2012). Moral panic, miracle cures, and educational policy: What can we really learn from international comparison? *Scottish Educational Review*, 44, 4–21.
- Champagne, A., & Pearson, D. P. (2003). Subject domain: What is being measured? *NAEP Validity Studies: An Agenda for NAEP Validity Research*. National Center for Education Statistics, Working Paper no. 2003-07, 5–11./
- Dickson, N. (1995). League tables: Use for patients. *Quality in Health Care*, 4(1), 1. <http://dx.doi.org/10.1136/qshc.4.1.1>
- Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Educational Policy*, 29(6), 845–855. <http://dx.doi.org/10.1080/02680939.2014.892156>
- Gierl, M., & Haladyna, T. (2013). *Automatic item generation: Theory and practice*. New York: Routledge.
- Goldstein, H., & Spiegelhalter, D. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A*, 159, 385–443. <http://dx.doi.org/10.2307/2983325>
- Koren, J. (2015). Norway's school library development and school reforms 2001–2014. In B. Schultz-Jones, & D. Oberg (eds.), *Global action on school library guidelines*. Saur: De Gruyter. <http://dx.doi.org/10.1515/9783110362664-009>
- Kreiner, S., & Christensen, K. B. (2014). Analyses of model fit and robustness. A new look at the PISA scaling model underlying rankings of countries according to reading literacy. *Psychometrika*, 79(2), 210–231. <http://dx.doi.org/10.1007/s11336-013-9347-z>
- Leamer, E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73, 31–43.
- Mislevy, R. (1995). What can we learn from international assessments? *Educational Evaluation and Policy Analysis*, 17, 419–437. <http://dx.doi.org/10.3102/01623737017004419>
- Mislevy, R., Beaton, A., Kaplan, B., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29(2), 133–161. <http://dx.doi.org/10.1111/j.1745-3984.1992.tb00371.x>
- Nortvedt, G., Pettersen, A., Pettersson, A. & Sollerman, S. (2016). Is PISA 2012 relevant to mathematics education in Norway and Sweden? *Northern Lights on PISA and TALIS*, Nordic Council of Ministers.
- OECD (2014). *PISA 2012 sampling guidelines*, PISA. OECD Publishing.
- OECD (2014). *PISA 2012 technical report*, PISA. OECD Publishing.
- Rutkowski, L. & Rutkowski, D. (2010). Getting it “better”: the importance of improving background questionnaires in international large-scale assessment. *Journal of Curriculum Studies*, 42, 411–430. <http://dx.doi.org/10.1080/00220272.2010.487546>
- Spearman, C. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychology*, 15, 201–293. <http://dx.doi.org/10.2307/1412107>

- Stacey, K., & Turner, R. (2015). *Assessing mathematical literacy: The PISA experience*. New York: Springer. <http://dx.doi.org/10.1007/978-3-319-10121-7>
- von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? In M. von Davier, & D. Hastedt (eds.), *IERI monograph series: Issues and methodologies in large scale assessments (vol. 2)*. IEA-ETS Research Institute.
- Wu, M. (2010). *Comparing the similarities and differences of PISA 2003 and TIMSS*. OECD Education, Working Paper no. 32, OECD Publishing. <http://dx.doi.org/10.1787/5km4psnm13nx-en>

Sammendrag

I den rapporten har nordiske forskere brukt de internasjonale studiene PISA (Programme for International Student Assessment) og TALIS (Teaching and Learning International Survey) til å undersøke forskjeller og likheter i de nordiske landenes utdanningssystemer. Både PISA og TALIS er studier fra OECD.

Sentrale spørsmål som blir besvart i rapporten er:

- Er PISA 2012 relevant for matematikkutdanningen i Norge og Sverige?
- På hvilke måter henger ulike lederstiler blant skoleledere sammen med lærernes holdninger og atferd, og elevenes prestasjoner?
- Hva er sammenhengen mellom faglig utvikling, trivsel og tro på egen mestringsevne?
- Kan samarbeid med kollegaer og skoleleders tilbakemeldinger forbedre lærernes tro på egen mestringsevne?
- Hva kjennetegner høyt presterende elever i matematikk i de nordiske landene?
- Hvordan kan vi bedre forstå metodiske utfordringer og kompleksiteter rundt internasjonale studier?

De nordiske landene deler i stor grad den samme kulturelle bakgrunnen og dette gir en unik mulighet til å analysere utdanningssystemer på tvers av landegrensene. Målet med denne rapporten er å gi relevant kunnskap til å videreutvikle nasjonal utdanningspolitikk. Vi ønsker også å gi et grunnlag for felles nordiske satsinger og videre forskning på utdanning i Norden.



norden

Nordic Council of Ministers

Ved Stranden 18
DK-1061 Copenhagen K
www.norden.org

Northern Lights on PISA and TALIS

- Is PISA 2012 relevant to mathematics education in Norway and Sweden?
- In what ways are the different leadership styles among principals in the Nordic countries related to teachers' attitudes and behaviours and students achievements?
- What are the associations between professional development, job satisfaction and self-efficacy?
- Can collegial work and school leader feedback improve teachers' self-efficacy in Nordic classrooms?
- What characterizes high-performing students in mathematics within the Nordic countries?
- Are international large-scale educational assessments elephants arriving at the gates of our national educational system?

These are some of the questions that are discussed in this collection of articles. The issues are based on the results of the OECD studies PISA and TALIS. The articles aim to provide input for policy discussions and to further policy development within the Nordic countries. Therefore, the main target groups are educational ministers and policymakers at all levels. These analyses will also provide input to the joint Nordic initiatives on educational development.

TemaNord 2016:517
ISBN 978-92-893-4521-7 (PRINT)
ISBN 978-92-893-4523-1 (PDF)
ISBN 978-92-893-4522-4 (EPUB)
ISSN 0908-6692



9 789289 345217