Scandinavian Journal of Educational Research

ISSN: 0031-3831 (Print) 1470-1170 (Online) Journal homepage: http://www.tandfonline.com/loi/csje20

Development of School Achievement in the Nordic Countries During Half a Century

Jan-Eric Gustafsson & Sigrid Blömeke

To cite this article: Jan-Eric Gustafsson & Sigrid Blömeke (2018) Development of School Achievement in the Nordic Countries During Half a Century, Scandinavian Journal of Educational Research, 62:3, 386-406, DOI: 10.1080/00313831.2018.1434829

To link to this article: https://doi.org/10.1080/00313831.2018.1434829

Published online: 07 Mar 2018.

SCANDINAVIAN JOURNAL OF EDUCATIONAL RESEARCH, 2018
VOL. 62, NO. 3, 386–406
https://doi.org/10.1080/00313831.2018.1434829

Development of School Achievement in the Nordic Countries During Half a Century

Jan-Eric Gustafsson[a,b] and Sigrid Blömeke[b]

[a]Department of Education and Special Education, University of Gothenburg, Gothenburg, Sweden; [b]Centre for Educational Measurement, University of Oslo, Oslo, Norway

ABSTRACT
The aim is to describe the development of achievement in compulsory school in the Nordic countries from the 1960s. The study relies on published results concerning literacy and numeracy from the international large-scale assessments between 1964 and 2012. Among others, the following conclusions are drawn: (1) for most countries, a small but consistent increase in the level of achievement was observed from the mid-1970s to around 1990 for both literacy and numeracy; (2) Finland improved literacy performance dramatically between the mid-1980s and the mid-1990s, which is hypothesized to be associated with the introduction of part-time special education; (3) for all countries performance declined from the late-1990s/early-2000s, which is hypothesized to be due to different factors in different countries.

ARTICLE HISTORY
Received 7 October 2017
Accepted 28 January 2018

KEYWORDS
Educational achievement; international large-scale assessments; Nordic countries; TIMSS PIRLS PISA PIAAC

The levels of educational achievement reached by both individual students and educational systems have increasingly come into focus. At the individual level knowledge and skills are important determinants of success in different areas of life, such as in higher education and occupations and for being able to exercise autonomy and choice in all aspects of life (Haller & Portes, 1973; Schmidt & Hunter, 1998). At the societal level there is evidence that the level of knowledge and skills in the population has a causal impact on economic growth (Hanushek & Woessmann, 2015). The international large-scale assessments (ILSAs) have grown in terms of number of participants and frequency of repetition, and their impact on educational policy debates now is substantial. Increasingly, they also have become important sources of information not only about educational achievement differences between countries (Gustafsson, 2008), but also about determinants of these differences (Woessmann, 2016).

One of the most striking findings from the ILSAs is how large achievement differences are between countries (e.g. Hanushek & Woessmann, 2015; Mullis & Martin, 2007). For example, as was observed by Woessmann (2016), the achievement difference between a 15-year-old in the USA, or another average performing country, and in one of the top performing countries, is roughly twice what students usually learn during one year. Another striking finding is that there have been substantial achievement changes over time (Hanushek & Woessmann, 2015; Woessmann, 2016). Questions about how educational achievement has developed over time in different countries are of great interest in themselves, but they are also interesting because investigations of within-country change may provide insight into the mechanisms and factors that influence development of

CONTACT Jan-Eric Gustafsson jan-eric.Gustafsson@ped.gu.se Department of Education and Special Education, University of Gothenburg, P.O. Box 300, Gothenburg 405 30, Sweden

educational achievement (Gustafsson, 2007). The aim of the present study therefore is to investigate within- and between-country change of educational achievement within reading literacy and mathematics for four of the Nordic countries: Denmark, Finland, Norway and Sweden. These four countries are similar culturally (Antikainen, 2006) and, with the exception of Finland, they also have similar languages. The development of educational policy shows both similarities and differences across these countries (Imsen, Blossing, & Moos, 2017), and the relative homogeneity of the four countries in other respects may provide good opportunities to describe and understand changes in educational outcomes over time.

Previous Research

While so far no systematic compilation of information concerning the long-term trends of achievement for the Nordic countries has been published, the different countries have done analyses of their own development. For example, in Sweden the National Agency for Education (*Skolverket*) has published several reports that attempt to analyse and systematize the information in the ILSAs (e.g. Skolverket 2004, 2009, 2014). However, this information is in Swedish only, and it only focuses the development of educational outcomes in Sweden.

A research unit set up in Sweden to conduct evaluations of labour market and educational policies was assigned by the government the task of describing the change in Sweden's educational achievement over time, and to provide explanations for why the decline had occurred. Their report (Holmlund et al., 2014) compiled information about the development of knowledge and skills in Sweden from several different sources. The results indicated that the level of achievement was highest for the birth cohort 1972, which left compulsory school in 1988, and that there has been a successive decline in level of achievement ever since. Holmlund et al. (2014) thus concluded that the decline started before a series of decentralization, deregulation and marketization reforms that took place in Sweden during the early-1990s, and that these reforms therefore could not have been the cause of the decline.

In Finland the main question has been how the great success in the most well-known ILSA, namely PISA, is to be explained. Much has been written on this topic (e.g. Sahlberg, 2011; Välijärvi et al., 2007), and there seems to be agreement that the answer is not to be found in a single factor, but that several factors have been combined to create an educational system of high quality. However, the comprehensive school, which was successively introduced during the 1970s and 1980s, is generally referred to as the main source of such factors: an emphasis on equality of student achievement supported by small school differences that provide students with equal opportunities to learn; instruction that is adapted to the needs of the individual child; highly qualified teachers with a master degree; curricular flexibility and pedagogical freedom; and no national tests or examinations.

However, Sahlgren (2015) rejected these explanations and put forward several alternative hypotheses. He argued that Finland's development of educational achievement was not synchronised with the introduction of comprehensive schools in Finland, and he rejected the explanations referring to changes introduced with the comprehensive school reform, Sahlgren (2015) instead suggested that factors outside of education account for the dramatic improvement of Finland's school results. One of these is the traditionally high social status and quality of teachers, which dates back to the early nineteenth century, when education became an important factor in building Finland's nation and culture. Sahlgren also pointed to the circumstance that Finland developed economically later than the other Nordic countries. Finnish culture therefore kept traditional values longer than the other Nordic countries, and when education was made widely available this made for what Sahlgren (2015) labels a "wealth" effect. The wealth effect implies that economic growth is first associated with rapidly increasing achievement and later with decreasing educational achievement. He also suggested that the decline is due to an increase in the use of pupil-led teaching methods, at the expense of teacher-led instruction.

Empirical studies examining long-term developments are lacking for Norway but Grønmo, Borge, and Hole (2014; see also Grønmo, Onstad, Nilsen, Hole, Aslaksen, & Borge, 2012) identify a decrease in mathematics achievement in 4th and 8th grade between 1995 and 2003, followed by an increase. They assign both developments to changes in opportunities to learn mathematics. Norway underwent two major policy reforms in the mid-90s (L97) and 10 years later (L06: Kunnskapsløftet). They argued that the first reform had put too much emphasis on teaching everyday mathematical problem solving and too little on the type of mathematics students need in further education and professions (p. 116f). The second reform partly strengthened the latter type of mathematics and increased instruction time in mathematics for primary students.

We are not aware of long-term studies examining trends in Denmark. This may be due to sparse participation in the international large-scale assessments before 2000 and to a slower development of educational monitoring compared to the other Nordic countries (Baldersheim & Ståhlberg, 2002).

This brief review suggests that there is great need for detailed and reliable information about patterns of development of educational achievement within the Nordic countries and to expand the timeframe usually examined. The main aim of the current paper therefore is to take advantage of the information available in the reported ILSAs in the domains of mathematics and reading literacy since the 1960s and to describe the development of achievement in these domains in the Nordic countries.

Development of the Nordic Educational Systems

Below we provide a brief overview of the development of the Nordic educational systems after the mid-1900s.

Denmark has a long tradition of municipal autonomy when it comes to regulation of education. A national curriculum exists but the local authorities are in control of the local school curriculum and large parts of the educational budget. As a result, there is a long tradition of multiple school types. Streaming was abolished relatively late in a Nordic perspective, namely in 1993 (Imsen et al., 2017). Today, a comprehensive school system up to grade 9, following one year of pre-school exists. Free school choice is still a core element though, and compulsory schooling has been less homogenous than in the other Nordic countries. As in the other Nordic countries, public schools dominate, but the proportion of private schools is larger, recruiting 14% of the primary students and 26% of the lower-secondary students (OECD, 2014).

Grading happens only from Year 8 on and is the responsibility of the individual teachers throughout the school years (Lundahl & Tveit, 2014). This trust in teachers has a long tradition so that no formal processes of teacher evaluation are in place but feedback is provided informally. Only in 2006, national tests were introduced in Danish, mathematics, English and science for public schools; private schools still can opt out though.

Finland had already abolished tracking at the end of the 1960s to increase equity and implemented successively a 9-yr comprehensive school system throughout the country during the 1970s (Laukkanen, 2008). However, streaming still took place within schools until 1985, when it was abolished. Municipalities have had almost full autonomy in organizing schooling since 1985. Although a national curriculum exists, the municipalities are in control of the local curriculum and large parts of the budget. Free school choice has been granted since the 1990s but in reality it is used to a very limited extent only. Less than 5% of the students attend private schools today.

Finland's education system is running without any monitoring or standard-setting at the national level (Vainikainen et al., 2017). After an unsuccessful attempt to introduce standardized testing in the 1970s, no population-based testing exists today. School inspection systems have always been weak and were abolished in the mid-1980s. Monitoring of the quality of the national education system happens since then through sample-based national testing and, in addition, on the local level through municipal assessments. Finland makes in addition extensive use of screening tests from early on and provides a lot of support of special needs students. There has been a strong trust in teachers' competence and a long tradition of teacher education at the master level since the 1970s. Grading starts in Year 6 and is the responsibility of the individual teachers.

Norway has the longest tradition of comprehensive schools of all four Nordic countries. They were introduced in the 1920s and go up to grade 10 today. Less than 5% of the students attend a private school. Also school choice is limited, school intake is in general based on residency in an area and only few municipalities such as Oslo grant free school choice.

In Norway, a national curriculum defines only broad overall objectives for each subject. As in Denmark and Finland, the local authorities are in control of the school curriculum and large parts of the budget. National tests are compulsory in school Years 5 and 8 in Norwegian, English and mathematics. These are population-based tests, and group means are reported by schools, municipalities and nationally. Grading starts in Year 8 only and is the responsibility of the individual teachers.

Sweden has been following a long tradition of standardized national assessments. They were introduced in the 1940s to support teachers in their grading. At that time, different school tracks existed in parallel on the lower-secondary level, and grades from primary school (*folkskola*) represented the core criterion for admission to secondary (grammar) schools (Lundahl & Tveit, 2014). Tracking was abolished in 1962 and a 9-yr compulsory school was introduced. A national curriculum described very detailed learning objectives, content and teaching methods, and grading happened from Year 8 onward.

This highly centralized school system changed substantially from the late-1980s onward. In 1994, a new curriculum for compulsory schooling reduced centralization. Only learning objectives were prescribed, but not content or teaching methods, and as in the other Nordic countries municipalities became responsible for organizing schooling and allocating resources. Free school choice was introduced in 1991 and a year later private (independent) schools were allowed (Imsen et al., 2017). Today, about 12% of the students in Sweden go to private schools.

Analytic Challenges

The main focus of the present paper is on within- and between-country change in educational achievement for the four Nordic countries Denmark, Finland, Norway and Sweden, the aim being to make inferences about determinants of the observed patterns of outcomes. Such inferences are of a causal nature and it is well known that numerous challenges and pitfalls threaten credible causal inferences from observational data. While methods for causal inference from observational data are being developed (e.g. Gustafsson, 2007, 2013), these methods typically require information from a large number of countries, which is not available here given our focus on four countries only. Case studies of long-term within- and between-country change in educational outcomes may be another approach that suits the present aims better. The four Nordic countries form a small set of culturally and linguistically homogeneous educational systems, but which also differ from one another because educational reforms have been implemented in different ways and at different points in time. If specific educational reforms or societal changes may be associated with large changes in achievement that occur within a short time frame, this may provide a basis for at least tentative causal inferences.

Another challenge in the present study is that participation in the ILSAs has been quite irregular over time and across the Nordic countries. Thus, the information provided by these studies alone is too sparse to provide a sufficiently strong empirical basis for the purposes of the present investigation. Therefore the set of studies analysed has been extended with the first study within the PIAAC, which investigates reading literacy, numeracy and problem solving skills in representative samples of the adult population (OECD, 2013). The PIAAC study is in many ways similar to PISA, but rather than sampling only 15-year-olds as is done in PISA, it covers the age range from 16 to 65 years.

One of the main questions that may be investigated with PIAAC data is which factors influence differences in level of performance between different age groups. In addition to factors related to age per se, differences may be due to cohort effects, such as influences of mass media, nutrition and amount and quality of education. By combining aggregated PISA and PIAAC data for 20 countries, Gustafsson (2016) showed that linear achievement trends estimated from five PISA rounds were strongly related to achievement differences between corresponding age groups in PIAAC. The results were interpreted to show that quality of schooling has a lasting impact on adult literacy and numeracy performance levels. This result implies that differences in level of achievement between different age cohorts may carry information about historical differences in quality of schooling. If that is the case, we may use comparisons between performance levels for different age cohorts for the Nordic countries participating in PIAAC to make inferences about the quality of schooling during different periods of time. However, given that the amount of evidence to support this kind of interpretation of the cross-sectional PIAAC data is limited, one of the aims of the study is to investigate if results from PIAAC agree with those from the ILSAs conducted with school-age samples.

Research Questions

Against the background of the considerations presented above the following three research questions will be investigated:

(1) To what extent do the achievement levels in numeracy and literacy for different PIAAC age groups agree with outcomes in the international comparative assessments for corresponding cohorts?
(2) Can we identify distinct within- and between-country changes in achievement?
(3) How can we explain such distinct within- and between-country changes, if there are any?

Methods

The ILSAs have evolved over half a century, successively developing solutions to numerous methodological and practical issues. The number and categories of participating countries have changed as has the time interval between studies investigating the same domain. The sampling and measurement designs have changed too, which has influenced the outcomes of the studies. Such changes are necessary to consider when comparing and interpreting results from the different studies.

It is useful to group the ILSAs into four main categories. The first consists of the early studies conducted by the IEA up until 1991; the second includes the TIMSS study conducted between 1995 and 2012, and PIRLS conducted between 2001 and 2016; the third category consists of the five rounds of the PISA study conducted between 2000 and 2012; and the fourth the PIAAC cross-sectional study of adults.

The Early IEA Studies

International Association for the Evaluation of Educational Achievement (IEA) was founded in 1958 and the period 1960–1990 may be regarded as a first phase in the development of the ILSAs (Gustafsson, 2008). The design and conduct of these studies was a pioneering effort, which required identification and resolution of a wide range of difficult methodological challenges.

A particularly challenging issue concerned the definition of populations of students to be included in the studies. Given that the typical age at which a child starts school varies between countries, students of a given age will attend different grades. For example, students in countries with a British school-system (e.g. England, Australia, Canada) typically start school at age four or five, while in the Nordic countries school start age is typically seven, so for these countries there may be differences of up to three grade levels for students with a certain age.

In the early IEA studies there was more emphasis on comparability with respect to age than with respect to grade. However, the IEA studies have successively implemented a grade-based population definition. The choice of principle for population definition may have considerable impact on the outcomes of country-level comparisons. The age-based principle benefits countries with early school-start, while the grade-based population definition benefits countries with late school-start.

The early IEA studies all used different scales for reporting the outcomes, which implies that it is difficult to compare results across studies. One way to solve this problem is to compute deviations from an international average for each country, for example in terms of standard deviation units (or *d*-values). The parameter of each country thus expresses how much better or worse its population does in a given study compared to the study's international average. Effects of general improvements in nutrition, mass media and so on are taken out so that it is easier to discover differential effects due to particular changes in one country.

However, when only a small number of countries participate, and particularly when some are outliers with exceptionally high or poor performance, the results may be of limited value. Nevertheless, given the lack of better solutions, this approach has been used to transform the results into the same scale, after removal of countries that do not belong to the EU/OECD or have participated only once. Tables A1 and A2 in the Appendix present the studies and countries included, and the achievement estimates for the participating Nordic countries.

The TIMSS and PIRLS Studies

While the early studies suffered from weaknesses with respect to sampling, measurement or analysis (Medrich & Griffith, 1992), methodological developments caused quality to improve. The TIMSS, which was conducted in 1995 (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996), took advantage of the developments in modern test theory, or item response theory (IRT: Mislevy, 1984; Rasch, 1960). This made it possible to put student results obtained on different subsets of items on one and the same scale. This allowed use of many more items with matrix-sampling designs in which students were assigned different booklets with partially overlapping item content. This technique also made it possible to equate measurements over time by keeping secret a subset of items from one occasion and reuse them at a later occasion. Since 1995, TIMSS has been repeated on a four-year cycle, so currently there are trend lines for mathematics and science between 1995 and 2015 for several countries.

The 1995 TIMSS was the largest study conducted thus far in terms of the number of participating countries and the number of grade levels covered. The results are reported on a scale with the mean set to 500 and the SD to 100. The actual scale was established in 1995 and its basic definition has not changed.

In 2001 the IEA conducted the first round of the PIRLS (Mullis, Martin, Gonzalez, & Kennedy, 2003) in 4th grade, which was based on similar principles as TIMSS. PIRLS may be seen as a continuation and development of the earlier IEA reading literacy studies, and it is repeated every fifth year.

Under the assumption that the set of participating countries was reasonably representative when the scale was established, the best comparability across study waves may be achieved by relying on the originally set scale with mean 500 and SD 100. Furthermore, for comparability across studies it is also in this case convenient to rely on the *d*-scale (effect size), so the results for all the countries have been transformed to *d*-values, using the 500, 100 metric.

The PISA Study

In 2000, OECD (2001) carried out the first round of the PISA study, with a similar methodology as that used in TIMSS. However, there also are differences. Thus, PISA uses age (15-year-olds) to define the population. Furthermore, while the IEA tests are developed on the basis of the curricula of the participating countries, the PISA tests are based on competences that are expected to be important for further education, working life and societal participation. The PISA is carried out every third year and includes tasks from the domains reading, mathematics and science. The OECD countries are required to participate in PISA, and in addition several other countries participate.

The PISA results too are reported on a scale with mean 500 and SD 100. As for TIMSS, the country results in PISA were transformed to *d*-values in the present study using the 500, 100 metric.

The PIAAC Study

The OECD (2013), established in 2012 the PIAAC study, which investigates literacy, numeracy and problem solving skills among adults in the age range 16–65 years of age. The first round of PIAAC was conducted in 2012, with 23 participating countries.

The PIAAC results are reported on a scale with a mean of 250 and SD 50. However, for the purposes of the present study a method used by Gustafsson (2016) was adopted. The IDB Analyzer (2015), which takes full advantage of the 10 so called plausible values reported (Rutkowski, Gonzalez, Joncas, & von Davier, 2010), was in a first step used to estimate literacy and numeracy means for each of the 10 age group defined by 5-year intervals (16–19, 20–24, 25–29, 30–34, 35–39. 40–44, 45–49. 50–54, 55–59, 60–65) in the four Nordic countries that participated in PIAAC. The international averages of numeracy and literacy for all the 20 countries that were included in the Gustafsson (2016) study were also computed. In this dataset participants who were not born in the country where the testing was done were removed to better support inferences about quality differences between countries in compulsory schooling. Some of the PIAAC age cohorts included large groups of immigrant test-takers who had not gone through the Nordic school systems as children because they only came as adults to the Nordic countries. This was in particular true for Sweden (e.g. labour immigration during the 1960s or refugee immigration in the first half of the 1990s: Pettersen & Østby, 2013).

For both literacy and numeracy the international averages increased up to the age cohort 25–29, after which there was an almost linear decline up to the oldest age cohort (Figure 1).

In the next step *d*-values for the difference between the international averages and the means for the four Nordic countries were computed for literacy and numeracy, respectively, for each of 10 age groups. In this way any systematic effect associated with the different age groups was taken out of the analysis. The *d*-value differences for the age groups were finally graphed for each of the four countries separately for literacy and numeracy.

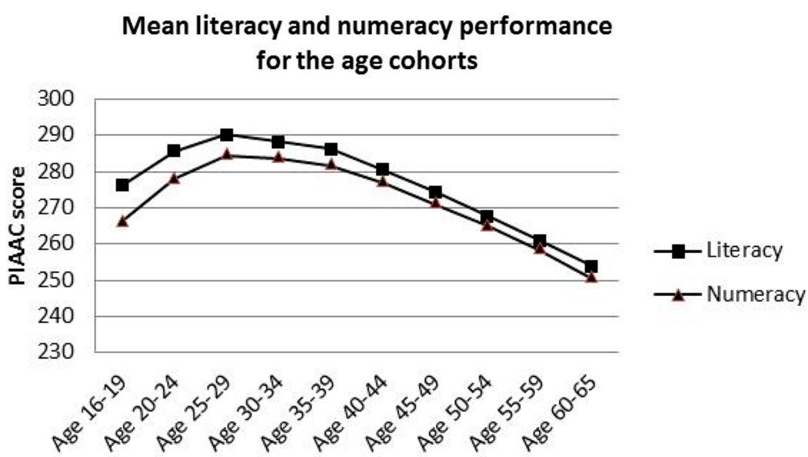

Figure 1. International averages of PIAAC literacy and numeracy performance across 20 countries (Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Great Britain/North Ireland, Ireland, Italy, Japan, Korea, the Netherlands, Norway, Poland, Russia, Spain, Sweden and the United States) for 10 age cohorts.

Results

The results are presented in several steps. First, comparisons are made between the PIAAC results and results from PISA, PIRLS, TIMSS and the early IEA studies in order to answer our first research question. Second, results concerning within- and between-country changes in achievement over time are presented to answer our second research question.

Agreement of PIAAC and the International Student Assessments: Reading Literacy

We start with results concerning reading literacy before we turn to numeracy.

PISA

Table 1 presents a compilation of the results from PIAAC and PISA, expressed in terms of *d*-values. In order to match the PIAAC 2012 age interval of 16–19 as closely as possible, results for PISA 2009 (18 years old in 2012) and PISA 2012 were combined, and to match the PIAAC 2012 age interval of 20–24 years old, results for PISA 2003 and PISA 2006 were combined. To match the PIAAC age group of 25–29 years the PISA 2000 study was used.

The superiority of Finland over the other countries is clearly demonstrated in both the PIAAC and PISA results, and the numerical estimates were quite close. For all four Nordic countries, the PIAAC results revealed a declining trend over time that is partly also reflected in the PISA results The PIAAC and the PISA results thus showed quite good agreement, even though PIAAC performance was higher than PISA performance for Sweden, and PIAAC performance of the age group 25–29 was for all four countries somewhat higher than the PISA 2000 performance.

PIRLS

The IEA PIRLS reading literacy assessment has been conducted three times (2001, 2006 and 2011) in 4th grade when students were about 10 years old. The participants in PIRLS 2011 did not leave compulsory school until 2016/2017, so for this group no PIAAC information is available to compare with. We therefore restrict our attention to PIRLS 2001 and PIRLS 2006. Participants in these studies typically left compulsory school 2006/2007 and 2011/2012, respectively. Norway and Sweden participated in both these studies.

As Table 2 reveals, for Sweden the PIRLS results were higher than the PIAAC results, while for Norway the PIAAC performance was close to the PIRLS performance. However, even though for Sweden the PIAAC estimates were lower than the PIRLS estimates in absolute terms, it is interesting to observe that the estimated performance change over the five-year period was about the same in both PIRLS and PIAAC.

Table 1. Reading literacy performance in PIAAC 2012 and PISA for corresponding age cohorts (*d*-values).

	Denmark	Finland	Norway	Sweden
PIAAC 2012				
Age 25–29	0.11	0.55	0.23	0.35
Age 20–24	−0.04	0.59	0.02	0.27
Age 16–19	−0.09	0.35	−0.12	0.13
PISA (15-year olds)				
PISA 2000	−0.04	0.45	0.04	0.15
PISA 2003/2006	−0.03	0.49	−0.04	0.15
PISA 2009/2012	−0.04	0.37	0.04	−0.02
Difference				
PIAAC Age 25–29 vs PISA 2000	0.15	0.10	0.18	0.20
PIAAC Age 20–24 vs PISA 2003/2006	−0.01	0.10	0.06	0.12
PIAAC Age 16–19 vs PISA 2009/2012	−0.05	−0.02	−0.16	0.15

Note: Comparisons between roughly equivalent age cohorts in PIAAC and PISA. Results for PISA 2003 and 2006 have been averaged, and results for PISA 2009 and 2012 have been averaged.

Table 2. Reading literacy in PIAAC and PIRLS (*d*-values).

	Norway	Sweden
PIAAC Literacy		
Age 20–24	0.02	0.27
Age 16–19	−0.12	0.13
PIRLS		
PIRLS 2001	−0.01	0.61
PIRLS 2006	−0.02	0.49
Difference		
PIAAC Age 20-24 vs PIRLS 2001	0.03	−0.34
PIAAC Age 16-19 vs PIRLS 2006	−0.10	−0.36

Note: Comparisons between roughly equivalent age cohorts in PIRLS and PIAAC.

Early Studies of Literacy

The second IEA reading literacy study was conducted in 1991 (Elley, 1994). This study comprised more than 30 countries in 8th and 3rd grade, and here too we investigate the deviations from the international averages for the Nordic countries. The results are presented in Table 3.

The IEA RL 1991 8th grade results of students roughly 15 years old may be compared with the results for the PIAAC 35–39 age group that in 1991 was about 14 to 18 years old. The agreement between the PIACC and IEA RL 1991 results was quite good, except that Norway performed higher in PIAAC than in IEA RL 1991.

The IEA RL 1991 3rd grade results of students roughly 10 years old may be compared with the results for the PIAAC age group 30–34 that was 9 to 13 years old in 1991. For Denmark there was little agreement, the PIACC result being at the international average, while the IEA RL 1991 3rd grade result was far below. For the other three countries the PIAAC and RL 1991 3rd grade results were quite similar.

The first IEA reading literacy study was conducted in 1970 (IEA RL 1970; Thorndike, 1973) and comprised samples from a mixture of two lower grades (3rd and 4th) and two upper grades (7th and 8th). Among the Nordic countries, only Finland and Sweden participated. We may compare the results for the upper grades with the results for the PIAAC age group 55–59 years, and the results for the lower grades with the PIAAC age group 50–54 years (see Table 4).

For Finland the IEA RL results for the upper grades were close to the international mean, which also was the case for the PIAAC results for the corresponding age group. However, for Sweden the PIAAC performance was much higher than the IEA RL performance. For the lower grades there was better agreement between the PIAAC and IEA RL results in the case of Sweden, but somewhat less agreement in the case of Finland.

Agreement between PIAAC and the International Student Assessments: Mathematics

We next present the results for mathematics.

Table 3. Reading literacy performance in PIAAC and IEA RL 1991 for corresponding age cohorts (*d*-values).

	Denmark	Finland	Norway	Sweden
PIAAC Literacy				
Age 35–39	0.08	0.60	0.26	0.45
Age 30–34	0.07	0.68	0.35	0.39
IEA RL 1991				
Grade 8	0.00	0.48	−0.12	0.29
Grade 3	−0.44	0.73	0.17	0.35
Difference				
PIAAC Age 35–39 vs Grade 8	0.08	0.12	0.38	0.16
PIAAC Age 30–34 vs Grade 3	0.51	−0.05	0.18	0.04

Table 4. Reading literacy in PIAAC 2012 and IEA RL 1970 (*d*-values).

	Finland	Sweden
PIAAC 2012		
Age 55–59 years	0.14	0.40
Age 50–54 years	0.34	0.29
IEA RL 1970		
Grade 7/8 (~14/15 years old)	0.05	−0.09
Grade 3/4 (~10/11 years old)	0.17	0.33
Difference		
PIAAC Age 55–59 vs Grade 7/8	0.09	0.49
PIAAC Age 50–54 vs Grade 3/4	0.17	−0.04

Note: Comparisons between roughly equivalent age cohorts in IEA RL 1970 and PIAAC.

Table 5. Mathematics performance in PIAAC and PISA for corresponding age cohorts (*d*-values).

	Denmark	Finland	Norway	Sweden
PIAAC 2012				
Age 25–29	0.31	0.49	0.18	0.42
Age 20–24	0.16	0.42	0.13	0.33
Age 16–19	0.01	0.27	0.04	0.23
PISA (15-years old)				
PISA 2000	0.14	0.36	−0.01	0.10
PISA 2003/2006	0.14	0.46	−0.07	0.06
PISA 2009/2012	0.02	0.30	−0.06	−0.14
Difference				
PIAAC Age 25–29 vs PISA 2000	0.17	0.13	0.19	0.32
PIAAC Age 20–24 vs PISA 2003/2006	0.02	−0.04	0.20	0.28
PIAAC Age 16–19 vs PISA 2009/2012	0.00	−0.02	0.10	0.36

Note: Comparisons between roughly equivalent age cohorts in PIAAC and PISA. Results for PISA 2003 and 2006 have been averaged, and results for PISA 2009 and 2012 have been averaged.

PISA

Comparisons between the PIAAC and PISA results can be made in the same way for numeracy as for literacy (see Table 5).

The comparison between PIAAC and PISA with respect to numeracy yielded similar results as those previously reported for literacy, with a relatively good agreement between the estimates of the two studies. However, for numeracy the difference in favour of PIAAC for Sweden was even more pronounced than for it was for literacy, as was also the tendency for the PIAAC performance of the age group 25–29 to be higher than the PISA 2000 performance for all countries.

TIMSS

For TIMSS 8th grade there is information from four TIMSS cycles for Norway and Sweden (TIMSS 1995, TIMSS 2003, TIMSS 2007 and TIMSS 2011). However, the students in TIMSS 2011 only were around 15 years when they were assessed, so there is no suitable comparison group in the PIAAC data. The data from TIMSS 2011 will therefore not be used for any comparisons with PIAAC. The data from TIMSS 2003 and TIMSS 2007 have been averaged, so as to form a suitable comparison with the 20–24 year age group. For Sweden, there was good agreement between the TIMSS 1995 performance and the PIAAC numeracy result for the age group that left compulsory school 1993–1997 (age 30–34). However, for TIMSS 2003/2007 performance was considerably lower than for the 20–24 year age group in PIAAC. For Norway too both comparisons showed substantially higher PIAAC performance than TIMSS performance (Table 6).

The Nordic participation in TIMSS 4th grade has been limited. Norway took part in 1995, 2003, 2007 and 2011, while Denmark and Sweden took part in 2007 and 2011. However, for the students who took part in TIMSS 2007 and TIMSS 2011, there is no corresponding PIAAC age group to compare with, so no results from TIMSS 4th grade 4 will be reported here.

Table 6. Mathematics performance in PIAAC and TIMSS for corresponding age cohorts (*d*-values).

	Norway	Sweden
PIAAC Numeracy		
Age 30–34	0.46	0.38
Age 20–24	0.13	0.33
TIMSS Grade 8 (~15 years old)		
TIMSS 1995	−0.02	0.4
TIMSS 2003/2007	−0.35	−0.05
Difference		
PIAAC Age 30–34 vs TIMSS 1995	0.48	−0.02
PIAAC Age 20–24 vs TIMSS 2003/2007	0.48	0.38

Note: Comparisons between roughly equivalent age cohorts in PIAAC and TIMSS. Results for TIMSS 2003 and 2007 have been averaged.

Table 7. Mathematics performance in PIAAC 2012 and older IEA studies for corresponding age cohorts (*d*-values).

	Finland	Sweden
PIAAC 2012		
Age 60–65	0.17	0.43
Age 45–49	0.35	0.44
IEA Studies (~14 year-olds)		
FIMS 1964	0.19	−0.63
SIMS 1980	−0.31	−0.92
Difference		
PIAAC Age 60–65 vs FIMS 1964	−0.02	1.06
PIAAC Age 45–49 vs SIMS 1980	0.66	1.36

Early Studies of Mathematics

Among the Nordic countries, only Sweden and Finland participated in the first mathematics study in 1964 (FIMS 1964; Husén, 1967a, 1967b), both with students in 7$^{\text{th}}$ grade. The results of this study may be compared with the PIAAC results for those who were 60–65 years old. The second IEA mathematics study (SIMS 1980) was conducted almost 20 years after the first one, in 1980–1982 (Robitaille & Garden, 1989). The results achieved in this study may be compared with the PIAAC results for those who were 45–49 years old. (see Table 7).

With the exception of Finland's results in FIMS 1964, the PIAAC results were very much higher in all comparisons, which mainly was due to the very low performance in the IEA studies.

Within- and between-country Changes in Achievement: Literacy

We now turn to descriptions of differences in levels of literacy performance for the four Nordic countries and how these vary across different age groups in the PIAAC study. Figure 2 presents the *d*-values for the difference between the international literacy average in PIAAC and the four country means for each of the 10 different age categories. In order to increase interpretability each age group is also labelled with the interval of years during which the participants left compulsory school, assuming that this generally was done at age 16.

For Denmark, literacy performance was relatively stable across the age groups, with a level close to the international average and mostly lower than in the other three Nordic countries. Finland had by far the highest level of literacy performance of all countries for the five youngest age groups, which were those who left compulsory school from around 1990 and onwards.

Except for the lower level of performance for Denmark there were relatively small performance differences among the countries for the five oldest age groups. Indeed, for the age-group

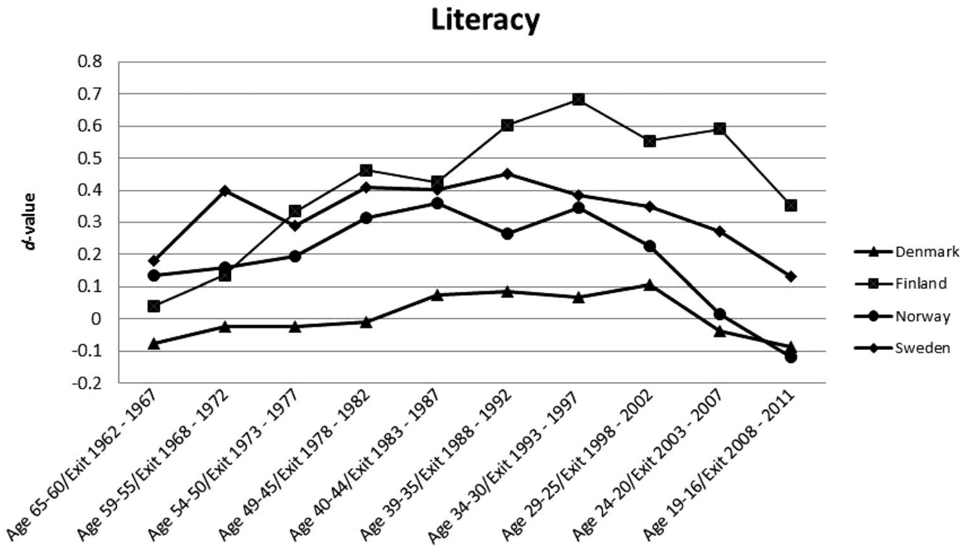

Figure 2. Literacy results (*d*-values) in PIAAC for 10 age cohorts in the Nordic countries. The difference has been computed between the mean for each age cohort and each country and the international average for the age cohort, after which the difference has been standardized with the average standard deviation across countries for the age cohort.

that left compulsory school 1983–1987 there were virtually no performance differences between participants from Finland, Norway and Sweden. For the age-group that left compulsory school 1988–1992 Finland had a performance advantage of at least $d = .20$ over all the other Nordic countries, and the performance advantage increased to at least $d = .30$ for the age group that left compulsory school 1993–1997. For the three youngest age groups there was a somewhat uneven declining trend for all countries, but Finland kept its performance advantage over the other countries. Compared to the international average, Finland's performance advantage for the age groups that left compulsory school after around 1990, except for the youngest one, amounted to around $d = .60$.

For Sweden and Norway literacy performance increased up to the age groups that left compulsory school 1990 and thereafter the trend has been decreasing. The performance levels peaked at around $d = .40$ for Sweden and $d = .30$ for Norway.

Within- and between-country Changes in Achievement: Mathematics

We next turn to the mathematics outcomes (Figure 3). Finland had the highest results among all countries for the younger age groups, the highest results being achieved for those who left compulsory school 1993–1997 ($d = 0.55$). For Finland, as for the other Nordic countries, numeracy performance declined for the age groups leaving compulsory school 1998 and onwards. The decline was steepest for Denmark and Norway.

Finland had the poorest numeracy performance (d around 0.15) among all the Nordic countries for the two oldest age groups (i.e., those being 16 years old 1962–1972). Among these age groups Sweden had the highest level of performance (d around .45), with Denmark and Norway in between. For the age group leaving compulsory school 1973–1992 there were relatively small differences in numeracy performance among the four countries, and a relatively high level of performance compared to the international average ($d = .25$–$.45$). Thus, the largest differences between the Nordic countries were observed among the oldest and the youngest age groups, and interestingly enough Finland had the poorest performance among the oldest groups and the best performance among the youngest age groups.

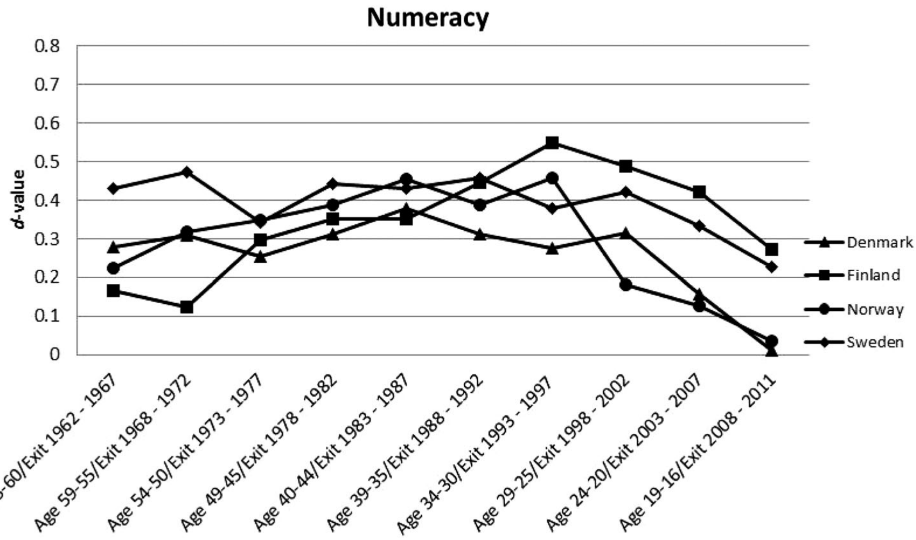

Figure 3. Numeracy results (*d*-values) in PIAAC for 10 age cohorts in the Nordic countries. The difference has been computed between the mean for each age cohort and each country and the international average for the age cohort, after which the difference has been standardized with the average standard deviation across countries for the age cohort.

Discussion and Conclusions

The current study set out to answer three research questions. The first was to what extent the achievement levels for different PIAAC age groups agree with the measured outcomes for corresponding cohorts in the international comparative assessments? The second was if we can identify distinct within- and between-country changes in performance, and the third question how can we explain any such distinct within- and between-country changes? The results pertaining to these research questions are discussed below.

Do Performance Levels for PIAAC Age Cohorts Agree with the Performance Levels According to the ILSAs?

The Results section includes a large number of comparisons between levels of performance observed in samples of school-aged participants in ILSAs on the one hand and the level of performance of PIAAC samples from corresponding age groups. If these performance estimates agree, this provides support for the idea that the performance levels of the age-cohorts in PIAAC reflect differences in level of quality of the basic schooling.

The empirical results vary across studies and subject-matter areas. In general, it seems that the agreement was better for literacy than for numeracy. Thus, for PISA literacy the agreement was good, with a few exceptions, and this was also the case for both the upper and the lower grades in the IEA RL 1991 study. For IEA RL 1970 there was some agreement, but also substantial differences. The comparisons based on PIRLS showed, in contrast, little agreement.

The pattern of results for PISA numeracy was similar to the pattern observed for PISA literacy but the PIAAC performance was higher than the PISA performance for Sweden in particular, but also for Norway. For TIMSS 8th grade there were substantial differences between PIAAC and PISA performance in almost all comparisons, and this was also the case for the two older IEA studies, SIMS 1980 and FIMS 1964.

There may, of course, be many explanations for why the PIAAC results differ from the ILSA results. One possibility is that the PIAAC performance has been influenced by, for example, upper secondary or tertiary levels of education, adult education and skill development in working-life (Gustafsson, 2016). While the effects of such factors are controlled for to the extent that they are common across countries, factors which are unique to smaller sets of countries are not. It also should be emphasized that the current study has a limited resolution because of the need to analyse the cross-sectional data in five-year intervals, which are not always well aligned with the time-points of data collection in the ILSAs. However, there also are other possible explanations for discrepancies, which are discussed below.

Differences Due to Measurement

Differences in the conceptualization, construction and administration of the PIAAC tests, and the tests used in the school-age ILSAs may cause lack of agreement. For example, Solheim and Lundetræ (2018) found that the tests of literacy in PIAAC, PISA and PIRLS gave rise to varying gender differences, and that this could be explained in terms of design characteristics of the tests. To the extent that age differences in performance are affected by such characteristics, this could also influence the outcomes of the comparisons between performance in PIAAC and PISA, and account for some of the relatively small discrepancies seen in Table 1. The TIMSS mathematics tests have a stronger emphasis on curriculum defined areas of mathematics, such as algebra and geometry, than have the numeracy tests in PISA and PIAAC. This may at least partly explain the better correspondence in the comparisons between PIAAC and PISA (Table 5) than in the comparisons between PIAAC and TIMSS (Table 6).

The PIAAC test was computer-administered individually, while the ILSAs considered here were paper-and-pencil tests administered in school-settings. These different modes of administration may also impact on the outcomes of the comparisons between performance in PIAAC and the school-age ILSAs.

Scale Equivalence

The composition of the countries participating in the different studies may also affect outcomes. In the comparisons between PIAAC and PISA this is not an issue, because the same set of countries were included, and the comparisons were made relative to the means of these countries. However, for comparisons with PIRLS and TIMSS the scores on the originally established scales were relied upon, under the assumption that the large number of countries participating in these two ILSAs would strengthen comparability. One possible explanation for the higher performance in PIRLS than in PIAAC for Sweden (Table 2) is that the set of countries which was used to establish the PIRLS scale in 2001 included a larger proportion of low-performing countries than PIAAC did. This seems indeed to be the case, and this is likely to be a major factor in explaining the discrepancy between PIRLS and PIAAC. It is, however, interesting to see that the performance change for Sweden between the two PIRLS cycles is of approximately the same size as the difference in performance between the two PIAAC age groups (−.14 vs. −.12). This suggests that even though the centre-points of the scales are misaligned, they can capture change over time in comparable ways.

For TIMSS 8th grade there also were substantial performance differences between TIMSS and PIAAC (Table 6), here in favour of PIAAC. This discrepancy too is likely to a large extent to be due to misalignment of the scales.

Comparability of Student Populations

Yet another source of differences in the outcomes of the comparisons has to do with which particular grades were included in the sample of participants in the ILSA. For example, Sweden participated in PIRLS 2001 with both 3rd and 4th grade samples. The 4th grade sample performed more than 40 score points higher than the 3rd grade sample ($d = .40$). This performance difference illustrates the dramatic effect that choice of participating grade has for the outcomes of the country comparisons, the 4th grade sample being at the very top of the ranking, and the 3rd grade sample being close to the international average.

Norway changed its school-start age from age seven to age six in 1997. This is one main reason why for Norway the level of performance in the ILSAs after 1997 tends to be lower than in PIAAC (Table 2 and Table 6). The change in school start age was implemented in such a way that students leaving 1st grade simply skipped 2nd grade, and went directly to 3rd grade. Students from this cohort participated in the TIMSS 2003 8th grade study (Grønmo & Gustafsson, 2010). Thus, the most likely explanation for the low level of performance in TIMSS 2003 compared to the PIAAC Numeracy performance is that Norway participated with students who actually only had seven years of schooling, even though they formally were in 8th grade.

In the early IEA studies the main emphasis was put on equality in the age of the students to achieve comparability across countries. However, this meant that students who started school early were in higher grades than students who started school at a higher age, and that the latter therefore had a performance disadvantage. This is one likely reason why performance in the older IEA studies tends to be lower than the PIAAC performance (Table 4, Table 7).

Study Failures

It does occasionally happen that results from an ILSA are not trust-worthy, because of procedural errors or other factors causing the results to be unreliable. In the IEA RL 1991 study good agreement was observed with PIAAC, except for the 3rd grade sample for Denmark, which had a very low level of performance, while the PIAAC estimate was close to the international mean. There is reason to believe, however, that the extreme IEA RL 1991 3rd grade result from Denmark is aberrant, possibly because of some procedural error in the conduct of this study. The result for Denmark was not only the lowest among all participating countries but the standard deviation was also by far the highest. In this situation it does seem more reasonable to accept the PIAAC estimate as the valid one, and reject the ILSA estimate.

Conclusion Regarding Comparability between PIAAC and ILSAs

Our first research question asks if the PIAAC achievement levels for different age groups agree with the measured outcomes for corresponding cohorts in the ILSAs. If we can answer this question affirmatively, we can rely on the PIAAC performance levels as indicators of the quality of basic schooling.

For literacy we find good agreement between PIAAC on the one hand and the results from PISA and the IEA RL 1991 study one the other hand. For PIRLS and IEA RL 1970 there is less clear agreement, but this may rather to be due to technical problems of establishing a proper scale for comparisons, or to a general lack of data in the case of the 1970 study, than to failure of PIAAC to reflect quality of schooling. However, it seems that we should be careful not to make strong statements about quality of schooling on the basis of PIAAC results for the oldest age groups.

For numeracy there was relatively good agreement between PIAAC and PISA, although not quite as good as for literacy. For TIMSS we encountered similar problems as for PIRLS, and there were large discrepancies between the PIAAC performance and the performance in the early IEA studies. Here too it may well be that the main reason for lack of agreement is that the scales used in the ILSAs are not comparable with the PIAAC scales. The problem of scale comparability needs to be addressed in further research.

Can we Identify Distinct Within- and Between Country Changes in Levels of Achievement?

The changes in levels of achievement across age groups for the four countries may be summarized as follows:

For Denmark only small differences between age groups were observed, but it may be concluded that Denmark throughout the studied period has had a relative strength in mathematics achievement and a relative weakness in literacy performance.

Finland's oldest age groups performed low in both literacy and numeracy, but for those leaving compulsory school from the mid-1970s achievement successively increased until it peaked in the mid-1990s for both literacy and numeracy. From around 1990 Finland's level of literacy achievement surpassed that of the other Nordic countries by a wide margin.

Norway's level of achievement increased according to PIAAC weakly from the 1960s until both literacy and numeracy started to decline from the mid-1990s to a level close to the international average at the end of the period.

Sweden had a relatively high level of literacy for those who left compulsory school in the 1960s and early 1970s, and possibly also a high level of numeracy. From the mid-1970s the literacy level successively increased until it peaked around 1990, after which it successively declined. The numeracy level kept quite constant between 1980 and 2000, after which it successively declined.

It may, furthermore, be observed that in spite of all the differences, there is a pattern which is common for most countries, with the possible exception of Denmark, namely a small but consistent increase in the level of achievement from the mid-1970s to around 1990 for both literacy and numeracy. Another common pattern is a decline in achievement levels from the late 1990s/early 2000s up to the end of the investigated period.

The typical pattern of change thus seems to be long-range trends, encompassing one or more decades, and they also seem to be common for several of the countries. Such trends are likely to be determined by multiple within-school and out-of-school factors, and this makes it difficult to isolate specific causal factors behind the change. However, there is one distinct pattern of change which concerns one country only, and this is the dramatic development of the level of reading literacy in Finland from the mid-1980s to the mid-1990s. In this case it is reasonable to expect that there is a single causal factor behind the development.

Explanations of the Development of Achievement in the Nordic Countries

Sahlgren (2015) argued that in Finland there was an accelerating increase of the level of achievement from 1965 until 2000, after which achievement decreased sharply to 2012. However, our results rather suggest that Finland's levels of achievement were low in both literacy and numeracy in the 1960s and early 1970s. From the mid-1970s achievement successively increased until it peaked in the mid-1990s for both literacy and numeracy. Sahlgren (2015) also rejected the hypothesis that the Finnish comprehensive school reform, which was successively introduced during the 1970s, was behind the high performance level in PISA. However, a closer look at the PIAAC results shows that there was an almost linear increase in the literacy performance levels of the age-groups leaving compulsory school between 1973 and 1997, and a similar pattern is seen for numeracy. The age-groups that went to school before the introduction of the comprehensive school performed about .10 *d* above the international average for both literacy and numeracy, while those who left school 1978–1982 performed .46 *d* and .35 *d* above the international average for literacy and numeracy, respectively. Ten years later these performance advantages had increased to .60 *d* and .45 *d*. These results suggest that the timing of the introduction of the comprehensive school was related to successively increased performance levels in both literacy and numeracy. The results also indicate that the Kerr, Pekkarinen, and Uusitalo (2013) study, which was conducted on data from the early phases of the implementation of the comprehensive school in Finland, grossly underestimated the effects of the reform.

It may also be observed that Finland's performance improvement between the mid-1970s and the mid-1990s can be divided into two periods. The first, which lasted to the mid-1980s was common with Norway and Sweden, and with respect to numeracy also with Denmark. During this period performance increased for all countries, and performance differences were relatively small between the countries. The second period comprised the years from the mid-1980s to the mid-1990s and during this period Finland's performance level continued to increase, and particularly so for literacy. For the other countries performance levels remained quite stable during this period. It seems reasonable that different explanations are required to account for the development during these two periods.

Sweden formally decided to implement comprehensive school in 1962, but as has already been mentioned there was an extended period of preparation in the 1950s. The implementation also involved successive revisions of the curriculum, and after a long period of planning a new teacher education designed to fulfil the needs of the comprehensive school was implemented in the late 1980s. Finland's comprehensive school reform formally started in 1972, but here too preparatory work was done in the 1960s. The reform was carefully planned and involved implementation of an elaborated system of special education, and a new five-year teacher education for primary grades, along with a large number of other changes (Sahlberg, 2011). Thus, in both Finland and Sweden implementation of comprehensive schooling not only involved abolishment of tracking, but also a large number of other changes, such as increased resources, increased quantity of schooling, and improved teacher education, just to mention a few examples.

Norway, which had abolished tracking already in the 1920s, took similar measures to improve its school system during this period of time. The most remarkable change was probably the extension of compulsory schooling from 7 years to 9 years during the 1960s, first as a pilot in a large number of municipalities, and from 1969 nation-wide (Tønnesen, 2004). A new type of school (*ungdomsskole*) and new requirements with respect to teacher education added to the massive expansion of quantity and quality in education. Finally, the same goes for Denmark, which did not abolish tracking until 1993 but extended compulsory school from seven to nine years in 1972.

It thus seems that all the Nordic countries during the 1960s and 1970s took measures to improve their educational systems. With the data available here it is impossible to sort out which specific factors were behind the performance improvements (see, however, & Kerr et al., 2013; Meghir & Palme, 2005), so we will have to leave the issue with the general hypothesis that the common performance improvement was due to the combined effect of different measures taken to improve the school system.

However, when it comes to the unique and dramatic improvement of literacy performance in Finland from the mid-1980s to the mid-1990s it should be possible to formulate a more specific hypothesis about the cause because both nine years of compulsory schooling and a teacher education on the master level was already in place. Indeed, such a hypothesis has been put forward by Kivirauma and Ruoho (2007; cf. Ihatsu & Savolainen, 1995) who suggested that the improved Finnish literacy achievement primarily was due to a part-time special education system that was successively developed in conjunction with the implementation of the comprehensive school. This system was mainly focused on reading- and writing-skills during the first years of schooling using a preventive pedagogical approach, and after being fully implemented around 1990 it involved around 20% of the Finnish students in grades 1 to 3 (Kivirauma & Ruoho, 2007). It has been argued that this unique form of special education accounts both for the high level of achievement in PISA literacy and for the low variance in scores (Kivirauma & Ruoho, 2007). The hypothesis that the unique improvement of literacy performance in Finland during the period mid-1980s to mid-1990s is supported by some simple statistics; in 1975 the system comprised 9% of the students, which had increased to 14% in 1985 (Ihatsu & Savolainen, 1995). The former group of students left compulsory school around 1985, while the latter group of students left compulsory school around 1995. While this result does not prove causality it certainly makes it worthwhile to conduct further research on the hypothesis.

Yet another common pattern which was recognizable for all countries was a decline in both numeracy and literacy from the late 1990s/early 2000s to 2011. However, this pattern must be seen in relation to the conclusion that the Nordic countries had an internationally high level of performance in the 1990s and it need not be the case that the same factors have caused the decline in all countries. Thus, for Sweden one main hypothesis is that the decentralization, deregulation and marketization reforms which were implemented in the early 1990s are the reasons for the decline, because they impacted negatively on the teaching profession and because they caused increasing school segregation. For Finland, Kivirauma and Ruoho (2007) expressed concerns that neo-liberal policies could cause the emphasis on preventive special education efforts for students at risk for reading and writing difficulties to be reduced, which may be expected to cause declining results.

For Norway, concerns have been expressed that the big changes in the curriculum expressed in the reform in 1996 with a lower focus on basic skills but a higher focus on individual student work may have caused the decline, in particular in mathematics where the traditionally high emphasis on applied mathematics was even more strengthened at the expense of pure mathematics necessary for a deeper understanding (Grønmo & Gustafsson, 2010). However, the fact that the participants from Norway in the TIMSS 2003 8th grade study only had seven years of schooling suggests that this may be an alternative explanation for the performance decline in Norway in the early 2000s.

General Observations and Conclusions

While all countries have participated in all rounds of PISA as a consequence of the requirement put on OECD-countries to do so, the participation in the IEA-studies has been uneven. Finland participated from 1964 to 1991, but thereafter the participation has been scattered. Sweden did so from 1964 to 1995, but less regularly thereafter, until 2007 when regular participation was resumed. Norway started to participate in the IEA studies in the 1980s and has since done so fairly regularly. Denmark's first participation was with the RL 1991 study but has since been uneven. Iceland has participated in TIMSS 1995 and in PISA. The uneven participation makes it difficult to extract trends in levels of educational achievement over longer periods of time for anything but a few countries, such as for Finland and Sweden.

Another threat to the valid use of these studies to establish trend lines is the great variation in principles for defining the populations, and the ensuing variation across countries in the choice of grade levels for participation. It has repeatedly been demonstrated how change of grade level one step up or down has had a dramatic influence on the outcomes, from a performance level at the international average to top level or to bottom level performance. The age-based selection principle which dominated the early studies is biased against countries with late school start, while the grade based selection principle is biased in favour of countries with late school start. Thus, even when there is no actual change, a change of principles of selection may give the impression of improvement or deterioration of achievement. It also seems quite clear that the age-based selection principles used in the early studies have had the consequence that the results for Finland and Sweden were substantially underestimated, and particularly so in mathematics.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Antikainen, A. (2006). In search of the Nordic model in education. *Scandinavian Journal of Educational Research, 50*, 229–243. doi:10.1080/00313830600743258

Baldersheim, H., & Ståhlberg, K. (2002). From guided democracy to multi-level governance: Trends in central-local relations in the Nordic countries. *Local Government Studies, 28*, 74–90. doi:10.1080/714004149

Beaton, A. E., Martin, M. O., Mullis, I. V., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1996). *Science achievement in the middle school years*. (IEA's Third International Mathematics and Science Study). Boston, MA: International Association for the Evaluation of Educational Achievement.

Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years*. (IEA's Third International Mathematics and Science Study). Boston, MA: International Association for the Evaluation of Educational Achievement.

Elley, W. B.1994). *The IEA study of reading literacy: Achievement and instruction in 32 school systems*. Oxford, UK: Pergamon.

Grønmo, L. S., Borge, I. Ch., & Hole, A. (2014). Mathematics in the Nordic countries: Trends and challenges in students' achievement in Norway, Sweden, Finland, and Denmark. In K. Y. Hansen et al. (Eds.), *Northern lights on TIMSS and PIRLS 2011: Differences and similarities in the Nordic countries* (pp. 107–136). Copenhagen, Denmark: Nordic Council of Ministers.

Grønmo, L. S., & Gustafsson, J. E. (2010). Student achievement in mathematics in Norway and Sweden as evidenced by TIMSS. Paper presented at the 4th IEA International Research Conference. Retrieved from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2010/Papers/IRC2010_Gronmo_Gustafsson.pdf

Grønmo, L. S., Onstad, T., Nilsen, T., Hole, A., Aslaksen, H., & Borge, I. C. (2012). *Framgang, men langt fram: norske elevers prestasjoner i matematikk og naturfag i TIMSS 2011* [Progress, but far ahead. Norwegian students' achievement in mathematics and science in TIMSS 2011]. Oslo: Akademika forlag.

Gustafsson, J.-E. (2007). Understanding causal influences on educational achievement through analysis of differences over time within countries. In T. Loveless (Ed.), *Lessons learned: What international assessments tell Us about math achievement* (pp. 37–63). Washington, DC: The Brookings Institution.

Gustafsson, J.-E. (2008). Effects of international comparative studies on educational quality on the quality of educational research. *European Educational Research Journal, 7*(1), 1–17. doi:10.2304/eerj.2008.7.1.1

Gustafsson, J. E. (2013). Causal inference in educational effectiveness research: A comparison of three methods to investigate effects of homework on student achievement. *School Effectiveness and School Improvement, 24*(3), 275–295. doi:10.1080/09243453.2013.806334

Gustafsson, J. E. (2016). Lasting effects of quality of schooling: Evidence from PISA and PIAAC. *Intelligence, 57*, 66–72. doi:10.1016/j.intell.2016.05.004

Haller, A. O., & Portes, A. (1973). Status attainment processes. *Sociology of Education, 46*, 51–91. doi:10.2307/2112205

Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. Cambridge, MA: MIT Press.

Holmlund, H., Hägglund, J., Lindahl, E., Martinson, S., Sjögren, A., Vikman, U., & Öckert, B. (2014). *Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola* [Decentralization, school choice and independent schools: Results and equity in Swedish schools]. IFAU report 2014:25.

Husén, T. (1967a). *International study of achievement in mathematics, Vol I*. New York, NY: John Wiley.

Husén, T. (1967b). *International study of achievement in mathematics, Vol II*. New York, NY: John Wiley.

IDB Analyzer (version 3.1) [computer software and manual]. (2015). Hamburg, Germany: International Association for the Evaluation of Educational Achievement.

Ihatsu, M., & Savolainen, H. (1995). Special education in Finland: Systems, history and current issues. *Australasian Journal of Special Education, 19*(1), 45–52. doi:10.1017/S1030011200023319

Imsen, G., Blossing, U., & Moos, L. (2017). Reshaping the Nordic education model in an era of efficiency: Changes in the comprehensive school project in Denmark, Norway, and Sweden since the millennium. *Scandinavian Journal of Educational Research, 61*, 568–583. doi:10.1080/00313831.2016.1172502

Kerr, S. P., Pekkarinen, T., & Uusitalo, R. (2013). School tracking and development of cognitive skills. *Journal of Labor Economics, 31*(3), 577–602. doi:10.1086/669493

Kivirauma, J., & Ruoho, K. (2007). Excellence through special education? Lessons from the Finnish school reform. *Review of Education, 53*(3), 283–302. doi:10.1007/s11159-007-9044-1

Laukkanen, R. (2008). Finnish strategy for high-level education for all. In N. C. Soguel & P. Jaccard (Eds.), *Governance and performance of education systems* (pp. 305–324). Dordrecht, The Netherlands: Springer.

Lundahl, C., & Tveit, S. (2014). Att legitimera nationella prov i Sverige och i Norge – en fråga om profession och tradition. [Legitimation of national tests in Sweden and in Norway: A question of profession and tradition]. *Pedagogisk Forskning i Sverige, 19*(4–5), 297–323.

Medrich, E. A., & Griffith, J. E. (1992). International mathematics and science assessment: What have we learned? US Department of Education, National Center for Educational Statistics, Research Report 92-011.

Meghir, C., & Palme, M. (2005). Educational reform, ability and family background. *American Economic Review, 95*(1), 414–424. doi:10.1257/0002828053828671

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49*(3), 359–381. doi:10.1007/BF02306026

Mullis, I. V. S., & Martin, M. O. (2007). TIMSS in perspective: Lessons learned from IEA's four decades of international mathematics assessments. In T. Loveless (Ed.), *Lessons learned: What international assessments tell Us about math achievement* (pp. 9–36). Washington, DC: Brookings Institution Press.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*. Chestnut Hill, MA: Boston College.

Organisation for Economic Co-operation and Development (OECD). (2001). *Skills for life: First results from PISA 2000: Education and skills*. Paris: OECD Programme for International Student Assessment.

Organisation for Economic Co-operation and Development (OECD). (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing. doi:10.1787/9789264204256-en

Organisation for Economic Co-operation and Development (OECD). (2014). *Education policy outlook: Denmark*. Paris: OECD Publishing.

Pettersen, S. V., & Østby, L. (2013). Immigrants in Norway, Sweden and Denmark. *Samfunnsspeilet, 5/2013*, 76–83.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Oxford, UK: Nielsen & Lydiche.
Robitaille, D., & Garden, R. A. (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. Oxford, UK: Pergamon Press.
Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International large-scale assessment data issues in secondary analysis and reporting. *Educational Researcher, 39*(2), 142–151. doi:10.3102/0013189X10363170
Sahlberg, P. (2011). *Finnish lessons*. New York, NY: Teachers College Press.
Sahlgren, G. H. (2015). *Real Finnish lessons: The true story of an education superpower*. London: Centre for Policy Studies.
Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274. doi:10.1037/0033-2909.124.2.262
Skolverket. (2004). *Internationella studier under 40 år. Svenska resultat och erfarenheter* [International studies during 40 years: Swedish results and experiences] Stockholm: Author.
Skolverket. (2009). *Vad påverkar resultaten i svensk grundskola? Kunskapsöversikt om betydelsen av olika faktorer* [What influences the results in Swedish compulsory school? Review on the importance of different factors] Stockholm: Author.
Skolverket. (2014). *Grundskolan i internationella kunskapsmätningar – kunskap, skolmiljö och attityder till lärande*. [The comprehensive school in international assessments: Knowledge, school environment and attitudes towards learning]. Rapport 407. Stockholm: Author.
Solheim, O. J., & Lundetræ, K. (2018). Can test construction account for varying gender differences in international reading achievement tests of children, adolescents and young adults? A study based on Nordic results in PIRLS, PISA and PIAAC. *Assessment in Education: Principles, Policy & Practice, 25*(1), 107–126. doi:10.1080/0969594X.2016.1239612
Thorndike, R. L. (1973). *Reading comprehension education in fifteen countries*. Stockholm: Almqvist & Wiksell.
Tønnesen, L. K. (2004). *Norsk utdanningshistorie: En innføring med fokus på grunnskolens utvikling* [Norwegian history of education: An introduction focusing on the development of comprehensive school]. Oslo: Fagbokforlaget.
Vainikainen, M. -P., Thuneberg, H., Marjanen, J., Hautamäki, J., Kupiainen, S., & Hotulainen, R. (2017). How do Finns know? Educational monitoring without inspection and standard setting. In S. Blömeke and J. E. Gustafsson (Eds.), *Standard setting in education. The Nordic countries in an international perspective* (pp. 243–258). Cham, Switzerland: Springer International.
Välijärvi, J., Kupari, P., Linnakylä, P., Reinikainen, P., Sulkunen, S., Törnroos, J., & Arffman, I. (2007). *The Finnish success in PISA – and some reasons behind it*. Jyväskylä, Finland: Finnish Institute for Educational Research: University of Jyväskylä.
Woessmann, L. (2016). The importance of school systems: Evidence from international differences in student achievement. *Journal of Economic Perspectives, 30*(3), 3–32. doi:10.1257/jep.30.3.3

Appendix

Table A1. Results for the Nordic countries in the early IEA studies of mathematics.

Study	Country	Age	Grade	Achievement
Mathematics				
FIMS 1964, upper grades	Finland	13.9	7	0.19
	Sweden	13.7	7	−0.63
	Int. Average (12)	13.7	8	0
SIMS 1980, upper grades	Finland	13.8	7	−0.31
	Sweden	13.9	7	−0.92
	Int. Average (16)	14	7.9	0

Note: An international average has been defined on the basis of a subset of the participating countries, including EU/OECD countries and countries participating more than once. For FIMS 1964 the comparison group includes Australia, Belgium, England, Finland, France, Germany, Israel, Japan, Scotland, Sweden, The Netherlands, and USA. For SIMS 1980 the comparison group includes Belgium (Fl), Belgium (Fr), Canada (BC), Canada (On), England and Wales, Finland, France, Hong Kong, Hungary, Israel, Japan, New Zealand, Scotland, Sweden, The Netherlands, and USA. Achievement is measured as deviation from the international average in standard deviation units (*d*).

Table A2. Results for the Nordic countries in the early IEA studies of reading literacy.

Study	Country	Age	Grade	Achievement
Reading				
RL 1970, lower grades	Finland	10.6	3.77	0.17
	Sweden	10.5	3.53	0.33
	Int. Average (10)	10.7	4.68	0
RL 1970, upper grades	Finland	14.6	7.56	0.05
	Sweden	14.6	7.51	−0.09
	Int. Average (11)	14.7	8.53	0
RL 1991, lower grades	Denmark	9.8	3	−0.44
	Finland	9.7	3	0.73
	Norway	9.8	3	0.17
	Sweden	9.8	3	0.35
	Int. Average (23)	9.68	3.48	0
RL 1991, upper grades	Denmark	14.8	8	0
	Norway	14.8	8	−0.12
	Finland	14.7	8	0.48
	Sweden	14.8	8	0.29
	Int. Average (24)	14.65	8.38	0

Note. An international average has been defined on the basis of a subset of the participating countries, including EU/OECD countries and countries participating more than once. For RL 1970 lower grades the comparison group includes Belgium (Fl). Belgium (Fr), England, Finland, Hungary, Italy, Scotland, Sweden, The Netherlands and USA. For upper grades in RL 1970 the same set of countries was used as comparison group with the addition of New Zealand. For RL 1991 lower grades the comparison group includes Belgium (Fr), Canada (BC), Cyprus, Denmark, Finland, France, Germany (East), Germany (West), Greece, Hong Kong, Hungary, Iceland, Ireland, New Zealand, Norway, Portugal, Singapore, Slovenia, Spain, Sweden, Switzerland, The Netherlands, and USA. For RL 1991 upper grades the same set of countries was used, with the addition of Italy. Achievement is measured as deviation from the international average in standard deviation units (*d*). Note that in the first IEA study from 1970 several grades (3 and 4 or 7 and 8) were included.