Schools, accountability and transparency—approaching the Nordic school evaluation practices through discursive institutionalism

Tommi Wallenius, Sara Juvonen, Petteri Hansen and Janne Varjo
Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

ABSTRACT
Over recent decades we have witnessed a growing emphasis on educational quality assurance and evaluation (QAE) around the globe. The trend, not only to intensify evaluative measurements, but also to publish school-specific indicators, has become visible also in the Nordic countries. In Sweden, Denmark and Norway, the governments have launched web-portals, in which various indicators can be observed and compared at the school level. However in Finland, the data is published only at a general level.

In this article we compare the discourses of educational experts on comprehensive school QAE policies and practices in four Nordic countries, Sweden, Denmark, Norway and Finland. Our aim is to clarify how the discursive practices reflect the current evaluation and publication policies and how the discourses construct the rationales of educational governance. We have approached our data (58 interviews) from the framework of discursive institutionalism, which sees both the underlying ideas and beliefs, and the discursive practices as the dynamic factors behind institutional change.

We argue, that in all the Nordic countries these discursive practices take place in a balancing discursive triad between global competence, neo-liberal accountability pressures and the traditions of the egalitarian Nordic comprehensive school—however with varying country-specific rationales on school accountability and transparency.

Introduction
Since the 1980s, there has been a steady growth of governing through evaluation and the use of data. The increase in the number of activities offering evidence about the performance of education in Europe and beyond is a signal of the development of quality assurance and evaluation (QAE) as a mode of governance within or across nation states Henry, Lingard, Rizvi, & Taylor, 2001; Ozga, Dahler-Larsen, Segerholm, & Simola, 2011). The overall restructuring of educational systems has been drawn from the principles of decentralization, devolution and deregulation (Whitty, Power, & Dalpin, 1998). Increased autonomy at the local level, however, in many countries has meant intensified reciprocal accountability measurements at the national level (Hudson, 2007; Smith, 2016).

The question whether the range of indicators, among them national testing, should be published at the school-specific level or not, has been a heavily debated and controversial issue in many countries and also among the academic community (Allen & Burgess, 2011; Ozga, 2009; Visscher, 2001). In the Nordic countries, the national school evaluation practices institutionalized during the 1990s and the 2000s into different trajectories (Eurydice, 2009a). In Sweden, Norway and Denmark, the governments have launched web portals which show manifold statistical indicators on a school level, while in Finland evaluation results are published only at a very general level. The differences lead one to think that there are different country-specific institutionalized ideas, rationales and discursive practices, not only on school evaluation, but also on school accountability or public information within the Nordic region.

Previous comparative country reports (e.g. Eurydice, 2009a; OECD, 2013) offer a good basis for detecting similarities and differences in countries’ policy practices but ‘static’ comparisons (e.g. test results are published/are not published) do not easily capture the institutional processes which constantly legitimate or challenge these decisions. In this article we point out the importance of discourses, how they both reflect and construct the understanding of what is held to be reasonable, needed, useful etc. in Nordic school evaluation contexts.

Thus, by comparing the discourses of the educational experts in four Nordic countries (Sweden, Denmark, Norway and Finland, in total 58 interviews) we aim to clarify how the discursive practices reflect the current evaluation and publication policies

CONTACT Tommi Wallenius, tommi.wallenius@helsinki.fi, Faculty of Educational Sciences, University of Helsinki, Siltavuorenpenger 1 B, PL 9, Helsinki, 00014, Finland
© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
and how the discourses construct the rationales of educational governance. We have approached our empirical data from the framework of discursive institutionalism (Schmidt, 2008) and problematize the experts’ views using two central concepts that label recent neo-liberal educational governance, namely accountability and growing demands for public information, transparency.

**National evaluation policies and the use of results in four Nordic countries**

In the Nordic countries, the foundations for the current QAE policies were formulated gradually during the 1990s and 2000s. Related to the ‘neo-liberal turn’ (e.g. Blossing, Imsen & Moos 2014), evaluation became a central mode of educational governance. The main features of the neo-liberal turn, decentralization, new public management and the introduction of market-logic actions such as school choice and privatization, have been described widely in the academic literature (e.g. Hudson, 2007; Lundahl, 2002; Ozga et al., 2011).

A brief overview shows the variety of the evaluation practices. Sweden was first to introduce a goal-oriented evaluation model with compulsory standardized testing for all pupils in 1994. The main aim was to enhance the comparability and fairness in pupils’ grading and to achieve systematic information of the decentralized school system by continuous evaluation (Eurydice, 2009b). Sweden was also the first to publish school-specific evaluation results. Since 2001, the results of national testing have been published on the National Agency for Education’s SIRIS internet database. Defining the audience as widely as possible was a clear principle from the start:

*The key social function of schools means that citizens have a democratic right to have access to this information. Child care and education affect almost everyone. In the Agency’s view, public access must therefore be as extensive as possible.* (National Agency for Education, 2018.)

Following Sweden in the 2000s, compulsory standardized tests for all pupils were also implemented in Norway and Denmark (Eurydice, 2009a; OECD, 2011a, 2011b). However, the decision about how to publish the results was not as straightforward as in Sweden. In Denmark, the Folkeskole Act in 2006 increased elements of evaluation and accountability in the Danish school system by requiring various kinds of documentation on schools and students’ individual learning progress and making it mandatory to publish results of the school-leaving examinations of compulsory education for schools and municipalities (OECD, 2011b, p. 28). Finally, in 2016 the Ministry of Education launched a web portal which enables the comparison of the schools across several indicators.

Also in Norway, the emergence of a national testing system was ulterior and more delicate. The question of publication policy went through a heavy political debate, initially leaving an undecided outcome, one in which the results for individual schools in the national tests were not made available by the responsible agency itself, but by the media, since the agency was required by law to provide the information to the press on request. A few years later, the national testing results were added in the official web portal ‘Skoleporten’, on which several indicators can be observed at the national, local and school levels (Norwegian Directorate for Education and Training, 2011; pp. 16–18; Hovdaugen, Vibe, & Seland, 2017).

Finally, in contrast to the other Nordic countries, the Finnish national testing policy and its practices institutionalized strikingly different. Despite many similar policy processes in the 1990s (decentralization, school choice), it was decided that the Finnish national testing was to be sample-based, covering roughly 5 to 15 per cent of the age cohort. A central argument for this even internationally rare practice (see Eurydice, 2009a, p. 27) was to avoid public school rankings, which were perceived as being detrimental and likely to accelerate school segregation (Jakku-Sihvonen, 2013). It is stated in the Basic Education Act (628/1998) that only ‘the salient results shall be published’, meaning in practice that results are published at the regional level or according to certain population factors.

The differences in the official policy guidelines on how school performance is evaluated at the national level, and especially for whom and for what uses the results are aimed at indicate that there are varying country-specific ideas, rationales and discourses that underlie these policies. In order to comprehend this diversity, we will next discuss two essential concepts, accountability and transparency.

**Accountability and transparency in the governance of education**

*Accountability* has become a cornerstone of public sector reforms in many countries. In general, the underlying rationale in accountability is that the producers are held accountable for the outcomes they generate. In the educational context, teachers and schools—who are trusted with the imperative task of teaching and instructing children—are the ‘producers’, while pupils’ test results function as the measurable ‘outcomes’ (Rosenkvist, 2010). Different country-specific emphases in respect of accountability have led to defining countries as representing either ‘hard’ or ‘soft’ or ‘high’ and ‘moderate’ QAE techniques (OECD, 2013).
Features of ‘hard’ QAE countries consist typically of external evaluation measures, inspections, standardized high-stake testing and ranking lists, commonly combined with different outcome-related incentives, e.g. sanctions or rewards. Emphasis on self-evaluation, sample-based or thematic evaluations and the absence of school rankings feature in ‘softer’ QAE policies (OECD, 2013; also Ozga et al., 2011; Smith, 2016).

The concept of accountability is tied to its counterpart, autonomy. In the Nordic context, the autonomy of the local actors, municipalities and schools, was increased considerably in the 1990s decentralization process. Decentralization and autonomy, however, are usually approached as a trade-off that includes reciprocal accountability measures:

In a context of school autonomy, greater policy attention is given to areas such as school leadership, capacity for schools to self-manage (including self-evaluation and the monitoring of the quality of teaching and learning) and ability to implement improvement processes. In addition, the greater responsibilities assumed by schools imply greater accountability requirements such as external school evaluation and public reporting of student performance. (OECD, 2013, p. 45.)

Thus, a central question which follows, is accountable for whom, and consecutively for what reason, why? In other words, which audiences or stakeholders are seen as being entitled to have access to evaluation results and how is this decision legitimized in society?

To conceptualize the question for whom, we utilize the idea of accountability as a relational and hierarchical web (see Bracci, 2009) composed of different layers with specific stakeholders. In this view, the accountability relationships originate from below always to a higher level. The test results are produced in a class led by a teacher and in a school run by the principal. In a decentralized system the evaluation results may be used by the local authorities but because of the idea of reciprocity between autonomy (local) and accountability (state), the national level actors (ministry and policymakers) may be seen as the primary users of the test results. However, the highest level of the hierarchy can be thought of as representing society, as in a democracy, the elected political representatives are ultimately in a political accountability relationship with the members of society for their decisions and governance practices (Darling-Hammond, 2004).

Furthermore, a more direct form of accountability supports the idea of society as the primary results user. Here, we emphasize the concept of transparency as a growing ideology that affects the accountability relationships shown above. As Tero Erkkilä (2012) argues, the principle of public information has always been a core feature of the Nordic states’ governance, but during the last few decades, the connotation of public sector information has shifted globally, including in the Nordic countries, to emphasize public sector efficiency and market accountability.

In the educational sector, this trend has become visible through several standardized measurements and the production of public data and rankings, aimed not only at the decision makers, but more directly at the pupils’ parents. At least three distinctive, yet overlapping accountability functions or rationales can be identified at the basic education level: accountability to taxpayers—a right of those paying for services to know the results; accountability to parents as school choosers—in order to make a rational school choice, the parents must be entitled to have access to relevant and comparable information on the quality of the service; and accountability to citizens—a more general governance principle, which highlights the undisputed democratic right of the citizens to have access to governance information.

The interwoven concepts of accountability and transparency seem to be the dynamic factors which keep pushing governments towards intensified evaluation and to publish results more openly. As presented above, these four Nordic countries have all approached the questions of how to collect school performance data optimally and the extent to which the results ought to be published, in different manners, based on their own historical, cultural and political trajectories. The official legitimations were described briefly earlier, but if we want to turn our focus to more recent policies or anticipate future paths, we have to look at the discursive practices which constantly interpret, renew or challenge the legitimacy of the current evaluation practices in the society.

From ideas to discourses—discursive institutionalism (DI) as a theoretical framework

As the theoretical framework to analyse our empirical data, we have followed the idea of Vivien Schmidt (2008) on discursive institutionalism (DI). Whereas the other neo-institutionalist ‘schools’ (see Hall & Taylor, 1996) have been more prominent in explaining the continuity of institutions, Schmidt (2008, 2010) sees the ideas and discourses as the fundamental and dynamic factors behind institutional change.

In this article, we understand the national testing practices and the publication policy as institutional forms representing the QAE culture in each society. We accept the notion of sociological institutionalism that in order to sustain their legitimacy, the institutions have to deal with a certain type of ‘logic of
appropriateness’ (March & Olsen, 1989), connected to the cultural norms and values in society. However, new ideas and discourses constantly shape the balance of the legitimacy.

In DI, ideas are seen as the substantive content of discourse. The ideas occur at three levels of generality: 1) as policy solutions providing suitable means for solving a specific problem or achieving the objectives set; 2) at a more general policy level as programmatic beliefs (Berman, 1998), which are more like paradigms or underlying principles to define for example the problems to be solved, the goals to be achieved and the norms, methods, and instruments to be applied; and 3) a philosophical level which works as the fundamental core, often remaining implicit or taken for granted (Schmidt, 2008).

The ideas become conveyed, challenged and exchanged in the discursive interaction processes between different actors and audiences. The coordinative discourse consists of actors at the centre of the political decision-making (politicians, civil servants, experts etc.), whereas the communicative discourse takes place in a wider political sphere through mass media and nowadays more through social media (Schmidt, 2008).

Despite the fact that discourses can originate from and be modified from the civil society to the administrative level, the direction of discursive interaction is often top down. Political elites tend to interweave the coordinative and the communicative discourses into a master discourse which sets out not only the visions for policymaking—what is, and what ought to be—but also defines the terms and the frames for the public discussion—what is rationalist, appropriate, how and why (Schmidt, 2008).

Aim, method and data

Thus, by comparing the coordinative discourses of the educational experts in each country, we aim to find new clarification in the discursive practices and in the legitimation processes to explain the similarities and the differences in the current comprehensive school QAE practices in the Nordic region. Our research questions are as follows: 1) how is the recent national testing policy and especially the use of the results legitimated/challenged by the educational experts in the Nordic region? 2) which underlying ideas or programmatic beliefs seem to guide these discursive practices? A special attention has been given to the Finnish case in contrast with the other Nordic countries. We assume that both the ideas and thereby also the discourses in the Finnish data differ from the other Nordic countries in respect of QAE-related rationales but also how the question of transparency is approached.

Our data comprise 58 interviews with well-versed experts in the educational field (senior politicians, civil servants, academics etc.) in each country: 17 from Sweden, 14 from Denmark, 18 from Norway and nine from Finland. The interviewees were identified through a ‘snowball sample’ method (also ‘chain sample’), in which the informants were asked to suggest other relevant informants (Noy, 2008). The method was found useful to map the central actors in an international and comparative research setting. To begin the discussions, one main question was addressed to the informants, ‘what is the story of your country’s basic education?’ The open interview structure provided freedom to the informants to bring up the elements of the story they found most important, evaluation being indisputably one central topic discussed throughout the data when reaching more recent events.

After a careful reading of the interview transcriptions, three discursive approaches emerged in each country’s data, but with important country-specific emphases: firstly, national evaluation in relation to global competence; secondly, a (neo)-liberal pressure to increase market accountability of the education system; and thirdly, national QAE for serving the traditions of the Nordic egalitarian school systems.

We will next move to the country analyses and summarize our findings in the following chapter. In order to protect the privacy of the interviewees, all the quotations are presented anonymously.

Sweden—politicized quality and strict accountability

Overall, the question of national assessment was not a particularly contested issue in the Swedish interviews—the need for evaluation was rarely questioned nor much defended in our data. The methods and functionality of the current national QAE system were discussed in practical terms, compared with the other Nordic countries, and the question was more how evaluation measures should be implemented and what their consequences might be, than why.

In the PISA aftermath, the role of faring in international competition became more prominent in the QAE discussion, and in the Swedish data, even when talking about national evaluation, PISA tended to come up. Swedish pupils did quite well in the first PISA study, but the results steadily declined in the subsequent evaluations until more recent years. Having previously perceived themselves as ‘the big country in the Nordic region’ (SWE12), the declining results opened the system up to more scrutiny, and finally in 2008 an independent body of school inspections was established. The quality of education both in terms of national performance and international competition became a political issue, and as one
interviewee described it, education became a more important issue in the 2006 general elections than ever before. In the narratives, the perceived decline in the school results was not only a concern about providing a quality education for new generations, but also about poor performance in comparison to other countries, which, again, strengthens the need for national performance measures. Simultaneously, there were several critical notions about PISA, questioning whether it actually benefits national educational policies.

[If you look at the grades in Sweden, grade point average has gone up all the time. In the national tests the results are more or less the same each year. So, we have not seen anything new, there is no national investigation that would show that Sweden is doing worse, it is only in the international measurements. (SWE13)]

The rationale behind the QAE system is that it will safeguard equality of opportunities, an idea which is closely tied to the welfare-state model’s educational equality. Since pupils’ grades affect advancement to further education, they should be comparable and not depend on their teacher, school, or classmates’ results. Moreover, supervising the quality of schools is also central in ensuring that each student receives instruction of equal quality.

The market-driven system had a distinctive role compared to the other Nordic countries in terms of steering quality. On the one hand, the quasi-market system was based on the idea that only good schools will survive in the market, and the responsibility for monitoring the quality of schools was delegated to the customers, i.e. families. On the other hand, the presence of private actors in the market created a new need for monitoring schools, to ensure all schools, independent or public, meet the national standards—the need to re-centralize. In both aspects, transparency works as a tool for accountability: publishing school-specific results is an attempt to keep them accountable—not only to the state but also towards parents, who are then thought of as making rational and fully informed decisions on the education of their offspring. In practical terms, schools and teachers were also made accountable to parents via an appeal system, through which an unsatisfied parent-customer can report their concerns to the school authorities. Parents’ increasing role was not seen as being unproblematic in the discussions, yet publication of school-specific results was not commonly considered to be a ‘hot political topic’, indicating that it is already a normal practice, not worth questioning.

And parents moving in [the school], which I spoke about before, for which the door was opened in the mid-90s, it has led to parents having a possibility to report. Anything that doesn’t seem to work well one can report to the Schools Inspectorate. So, schools work quite a lot today answering the Inspectorate about reports they have had. (SWE13)

The Swedish QAE discourse had a clear vein of concern: questions about having too much control and providing too few tools for development were raised. According to many informants, the need to have the basic education system and its quality under control has led to keeping teachers and schools under a microscope. In the most critical notions, the current accountability system was compared to ‘a Soviet-type control system, only implemented with market-based mechanisms’ (SWE06). Overall, there was a lot of concern about teachers and the fact that they are supervised too much, which in turn, has affected the popularity of the teaching profession—which, again, is a threat to the quality of teaching. Inspecting teachers’ work diminishes trust and, as one informant put it, takes away their professional ownership (SWE05). Some of the informants worried that too many reforms had been undertaken in the wake of the declining PISA evaluations in the 2000s, and the focus of developing the education system now lies too heavily on test results. Another question was, if evaluation data is used too much to control and too little to develop, and whether schools and teachers get enough tools for improving their work, rather than focusing on meeting national standards.

As we got all these deregulation, decentralisation, marketisation, the state had to take new responsibilities. Be much more of a control institution or apparatus. And so today Swedish teachers they, there are more grades, more tests, more inspections. […] There were maybe too many reforms and too much documentation because it’s also about having individual plans and documentation, teachers have to sit down and write novels almost. […] So the teacher-, teaching profession has gone down in popularity immensely over the years. (SWE03)

Denmark—gradual steps towards holistic accountability

Similar to in other countries, the Danish interviews were framed heavily by international comparisons, mainly the PISA assessment. Denmark was described as underachieving in the first PISA and eventually this ‘shock’, due to its high attention value, fostered the political imperative to reform the Danish school system by introducing a goal-oriented model with common goals and by strengthening the national quality assurance and evaluation system:

Over time, we stopped discussing the relevance of PISA measurement as a philosophical and methodological enterprise. And instead we discussed, what needs to be done with the Danish school. So in a sense, the power of PISA, regardless of all the
methodological qualms, the political and pragmatic side of PISA and its implication, so to speak, won. Because every time there is a PISA discussion in Denmark, the discussion is, 'how should we reform the school next'. It is no longer 'what is wrong with PISA'. (DEN04)

A major change in the Danish evaluative culture was mentioned as having taken place during the 2000s as the focus of the evaluation shifted gradually from the diagnostic and formative pupil assessment to emphasize school level evaluation and accountability. The OECD was seen as the ‘main driver’ behind the quality reforms but the province of Ontario in Canada was also mentioned as ‘a heavy source for inspiration’ (DEN03). The incongruence of high costs and low results was to be tackled by a more systematic and standardized evaluation despite the criticism by the teachers and their union:

So we had a visit from the OECD and they said you need to do something about the way you follow up on your students and how you evaluate. You need to have a better evaluation system. . . . And we came with our report in June 2006 saying that we want to have national tests, we want to have quality reports. We want it to stand in the aim that folkeskole should prepare for further education. It’s not enough in itself because we are in a globalised country. And each municipality should prepare quality reports about some statistics of how schools are doing. And the Teachers Union hated this. They really hated it. (DEN05)

Fostered market accountability, publication policy and especially the quite recently launched web portal were presented in the interviews as highly debated and controversial issues. The discussions included doubts about its unintended consequences, but at the same time, publishing school results and quality reports were described as a factual—even if very limited—source of information, to which the citizens, that is parents, are entitled either in their role as a school chooser or as a controller of the flow of tax money:

For [the government] it’s, they think, that by giving this access to a quite simple web-page where you can compare schools in your district and things like that, it helps parents to do school choice, and benefit from the idea of free choice of schools. That’s the idea. . . . And some think that it’s misleading. I don’t think it’s misleading. Of course it’s, grades are grades. Test results are test results. So they are of course an information, I think. . . . But, it can be misleading if you only look to those. (DEN03)

Methodological issues were also raised when discussing the publishing policy. Value-added evaluation results which control the socio-economic factors of the pupil population at each school were promoted, be it with reserve. In general, many Danish interviewees had mixed feelings: on the one hand, there was mistrust of the competition logic and school rankings, but on the other hand, an opportunity for better school development and better follow-up, and especially the role of public information, was approached as a fundamental and inescapable principle in a modern society:

So, I’m very sceptical about rankings. But I’m also extremely sceptical about the view that rankings should be prohibited. Because in a knowledge society, people who have a little bit of skills can publish whatever they like. It is no longer the idea that we have this one government that controls all the information. So, I would rather have that we undermine these rankings because we made them ridiculous, and we discuss them, and we compare them over the years. I would rather go in that direction, than having this idea that said ‘you know, the people don’t deserve to know because they don’t have the competence to interpret these things’[. . .] I would say, publish it if you want to. Let’s have the debate. And it’s fairly ridiculous in the long run in a way. (DEN04)

Despite all the problematizing, the near future trend was presumed to continue towards a higher level of accountability than changing the policy direction. The recent Danish QAE policy was even described as a self-piloting process, immune to domestic power relationships.

I think the next step we will see, is introduction of incentives. Our accountability system has been introduced slowly, step by step over the last ten fifteen years. And it doesn’t matter, whether it is a right-wing or a left-wing oriented government, there’s no big differences when we change government. (DEN03)

Norway—quality assurance for the ‘child’s best’

Most of the Norwegian educational experts described the 1990s as the starting point for a new and emerging assessment culture. As in many other countries, the 1990s meant a new kind of orientation to public education as a national economic investment. While previously Norwegians believed that they had ‘the best education system in the world due to huge investments’, now the focus was on cost-efficiency. Evaluation was seen as a tool to observe and clarify the ‘paradox’ of educational investments.

Yet, in the 1990s education politics, they didn’t conclude by saying that you have to introduce a national test system. But they asked whether they can know about the quality if they don’t have any kind of research or evaluation of some kind. So they asked how can you know, if you don’t check it? (NOR01)

Findings from several domestic and international evaluations in the late 1990s, together with the Norwegian version of the ‘PISA shock’, legitimized extending the QAE processes towards a more systematic and comprehensive testing apparatus, to get research-based information in order to reveal the
‘black box’ (NOR17) of the Norwegian education system. As a result, national tests were introduced in 2004.

And then the PISA results came in 2001, which kind of pushed a lot of ongoing development forward. People tend to point towards the PISA results but I believe that, a lot of different elements that were pushed forward in the education reforms of the 2000s, like the evaluations showing a lot, a lack of disciplinary issues. [...] And people started to question what people learn in schools. What do people learn? We cannot only trust that we put a lot of money in there and we focus on the processes. We have to know something about the processes as well. (NOR06)

The public availability and use of the evaluation results was initially described, debated and found to be complex. Still, the publication policy did not show up as the main question in the Norwegian data. For instance, many of the interviewees couldn’t describe the current publication policy. ‘The confusion might be explained by the many gradual phases during the challenging implementation process of the testing framework described earlier (see Hovdhaugen et al., 2017).

However, even though parents and pupils are entitled to school choice, school-specific indicators were not considered to guide parents’ choices excessively, even in the big cities. Instead of concealing the results, providers of education were seen as having emphasized the complex nature of the evaluation information—as something that cannot be used directly to draw simplified conclusion on ‘good’ or ‘bad’ school:

And we have spent a lot of time, and also the researchers to explain, that this is just, this tells very little. So if this is your only result, you should, you have to look into more. And you have to analyse this, in connection with what you know about the students from your class and from your other observations. So this is just one piece of information. And we have explained for many years that, it’s no use in publishing the rankings. Because, the differences are so small. (NOR09)

Of all the Nordic countries, perhaps in Norway the QAE policy was most clearly discussed from the pupils’ perspective, primarily to serve the pupils. Both the external assessments (national tests) and self-evaluations were described as having been developed in order to monitor education providers (schools & municipals) in doing their best for the child. This ‘child-centric assessment’ discourse, the historical roots of which date back to the early 1930s, works as an important legitimation mechanism between the early starting point of ‘soft’ and ‘humanistic’ assessment and the recently intensified quality assurance policy in Norway. All the evaluation practices are carried out to help and benefit the pupil:

Our education system is very different today than before. Because now we know more, we have developed, more research, more indicators. It’s more about the content, it’s more about the student learning. It’s more focused on student learning. So I think the story from that time is... it’s more on quality, quality assessment. How to make sure that all students really learn, and really fulfill their potential. (NOR09)

Finland—non-accountability through appropriate trust

The Finnish comprehensive school system got world-wide recognition as the first PISA results were published in 2001. Reaching the top positions among all participants was somewhat unexpected, a sort of positive PISA-shock compared to many other countries in which results were lower than anticipated. However, since 2006 the Finnish PISA results have indicated a gradual decline in the pupils’ learning and more recently a growing variance between female and male pupils. These trends were widely discussed in the Finnish interviews as alarming signals for the Finnish school system. Interestingly, the conclusions drawn were not linked to any needs to increase control aspects or accountability on Finnish schools, but merely raised concern on budgeting or teachers’ career training (FIN08). The present leading PISA performers from South-East Asia were described in an unfavourable tone as ‘intensive production units’ with ‘teaching-to-the-test culture and long school hours’ (FIN02), as if playing totally another game with different rules.

And now we have more these PISA assessments and the recent results indicate a slight decline. But we are still, according to these, the best in Europe. Not bad at all. Then there are these others, Shanghai, Korea and the rest. This competition league is a bit different, even if not knowing all the details. (FIN03)

The Finnish informants beheld a strong and a shared collective understanding of the mechanisms related to school system evaluation. The unintended consequences of high-stake testing practices and school accountability seemed to justify, even if implicitly, the reasoning for promoting continuity than change. Notably, no potential benefits of improving the Finnish school system through changing the present national evaluation practices were identified. This can be understood as one central programmatic belief (Berman, 1998) constructing the Finnish evaluation discourse, which is distinctive to other Nordic countries. On the contrary, introducing a more systematic evaluation system with standardized testing is simply not seen as being of benefit to the system.

Unlike in many Anglo-American countries, in which they have been seeking for better results through measurement and testing and sort of tighter school-level control, I don’t believe it is our way. I don’t
believe it will bring results. The question is more of how to build the school culture which will support the learning from the early childhood on. And now this segregation process we have is worrying. It is not of what the school does alone but also of how learning is valued in the society and I think in this we have gone backwards now. (FIN10)

A general concern about the direction of the Finnish comprehensive school in the near future was present in the interviews. However, the concern did not apply only to the basic education or the PISA results, but to a more general attitude towards learning and education in society. This is noteworthy because the education system and especially the comprehensive school system have been traditionally valued highly, as cornerstones for a small country to succeed in global competition. Changes in the societal environment could be expected to open new discourses in the QAE field. Here, the potential change towards increased evaluation or accountability is still articulated as not being needed because of a high level of trust among the central actors, namely the teacher, the principal and the local/national governance officials.

Our governance system is now based on the trust that the local actors follow the core curriculum for basic education. And if someone does not, what does it follow – nothing. At least never has, even it is a norm. And we like to think that in the municipalities they trust the principals and the teachers for doing their job, which they also have done. In the background we have of course our teacher education, not as good as advertised, but still fairly good so that the teachers are able to get their pupils far. So, we don’t need this kind of testing system, which I think would only lead to teaching to the tests. One must remember that education is for life, learning is still an adventure. (FIN01)

The pressure to change some core elements of the Finnish QAE policy was recognized and outlined by some of the interviewees. The identity of a Finnish teacher as an autonomous actor and societally-respected professional was seen as a key factor in buffering the pressure. Only when discussing the fairness of the grading system was autonomy considered to be problematic but still subsidiary. Without any compulsory standardized national testing or examination, the equivalence of the grades given by the teachers was considered as a sort of blind spot in the Finnish school system.

And the core competences, let’s say mathematics, are declining and soon it may be so low that I presume it is EK [Confederation of Finnish Industries] first who states that the comprehensive school system is collapsing and we must have a better clarity of our learning. And then we will have a testing unit waiting already for doing this task. And it is easy to see how it leads to a claim for having final exams like in the US. But for now we don’t have that. The Finnish teacher is so autonomous it is unthinkable. And it may have even gone too far, no one really knows what is going on inside the classroom. The Finnish principal does not listen out for the teacher because it just isn’t right, because of the trust. The new curriculum is very loose, we don’t have any inspection system like in the old days, no control for textbooks so you can just sit there and do your group work of what you wish. And then that we don’t have any tests and control for comparability, we know now from various researches that the grades given for the same level of competence vary by schools, I think this problem will arise much louder, especially if we fail in the PISA. (FIN04)

In contrast to other Nordic countries, the publication policy was not discussed in the Finnish data in terms of either accountability or transparency. The absence was not surprising, as the issue has been rarely touched and the current policy seems to enjoy a wide common understanding among the decisionmakers as well as in society. This can be understood as another underlying truth or paradigm in the Finnish evaluation culture, where different actors, also parents are expected to follow a certain logic of appropriate behaviour (March & Olsen, 1989), not to challenge the autonomy of the schools. Moreover, excessive consumer activity was considered to be detrimental to the core of the Finnish comprehensive school system, to break ‘the idea’.

I guess it was in 2009 when this one private tutoring company started. These provide tutoring for certain areas that have enough potential customers, say well-off families. And this is market-based, we cannot prevent the supply. But it will break the idea of the comprehensive school if the parents start to estimate that their child won’t get enough impulses or guidance and start to pay for it, then it will crack. (FIN06)

The discursive triad of school evaluation practices in the Nordic

Based on our empirical findings, we argue that the discussion of school evaluation practices in the Nordic region is formed in the interplay of a discursive triad, presented in Figure 1. The three main discourses are related to three forms of accountability, specific to the Nordic context.

A global competence discourse concerns the overall quality of the basic education system in the context of economic competitiveness and international comparisons. This is mostly discussed when referring to the PISA results, linking lower rankings to the imperative increase of testing and accountability measures (political accountability). The importance of PISA is noted in each country, if in varying manners: in Denmark, Norway and especially in Sweden, the national discourse is more about reforming the system to perform better in international measures. Since it attracts such wide-
spread attention, PISA acts as a final legitimating force for national reforms. In Finland, the effects have been almost the opposite per the initial success, and despite the recent decline in the latest PISAs, the Finnish system is still seen as being competitive.

A (neo)-liberal discourse articulates the increased testing as a follow-up service for individual schools and pupils. The publication of school-specific evaluation results is primarily expressed as a fundamental principle of governance transparency and as subsidiary to promoting consumer behaviour and school choice (market accountability). The market accountability is strongest in Sweden, where the public nature of schools’ results is built in to the quasi-market system, whereas in Finland, this discourse is most clearly absent. Overall, the Nordic countries are at different stages of the institutionalization of evaluation practices, not only structurally but also discursively. In Sweden, intensified evaluation and school-specific indicators are approached in practical terms, and in Denmark and in Norway, as a dynamic process that is still seeking its shape. In Finland, the discursive practices aim to defend the status quo from external challenges.

A Nordic comprehensive school discourse highlights the traditions of Nordic egalitarianism in order to prevent the increase of market-logic in education. This discourse is manifested in our data as a worry for the teachers about their status and autonomy, which external evaluation or publishing results challenges. It derives its legitimacy from the traditional idea of a common and equal Nordic comprehensive school (egalitarian accountability), working as a counterforce against the neo-liberal discourse. This discourse is strong especially in Finland and often becomes articulated by referring to warning examples of high-stakes testing.

How these three main discourses are manifested is relational to the programmatic beliefs and rationalities of school evaluation in society. For example, in Finland the whole idea of raising educational quality through intensified evaluation seems highly questionable. This rationalization has undisputedly become stronger in the 2000s because of the PISA success. Furthermore, the idea that the members of civil society construct the highest layer in the school accountability hierarchy is not questioned or even recognized as a political issue in Finland. On the contrary, the negative effects of testing and public results are many times transmitted as taken for granted. The balance of the three discourses combined with the underlying ideas and rationales specific to each society explain why public school results in one context are seen as supporting equality and efficiency, but in another as the ultimate source of inequality.

Discussion

In this article, we have touched on a topic of growing importance, school performance and its publicity in the Nordic countries. By comparing the coordinative discourse of the educational experts we were able to detect three main discourses, which reflect but also construct the present school evaluation practices in each four country. Our analysis showed that the balance of the Finnish discourses differs substantially from the other Nordic countries, not only because of the PISA success, but also in the rationales that underlie the discursive practices on pupils testing, school accountability or transparency.

Our notions are well in line with a common understanding, Finland being an upstream case in the global trend of intensified QAE practices. By contrast, the governments in Sweden, Denmark and Norway have launched web-based data portals that include various comparable indicators on pupils’ learning results, teacher-pupil ratios, school resources etc. This information is aimed at different stakeholders, not least to the parents in order to promote school comparisons and to choose a school (see e.g. www.valjaskola.se). Despite the attempts to produce a broad picture of the school environment consisting of multiple factors, school-specific data have often been compiled in the media into simplified listings and rank orders, accompanied by headlines about ‘the best and the worst’ performers.

The Finnish decision to carry out sample-based national testing has turned out tenable. It has served its purpose of evaluating the school system in general, but not to promote or highlight school differences, which comparisons always tend to create. If data
existed for all schools, the interest in its use comparatively would very likely grow. The only comparable pupil performance test in the Finnish education system, the matriculation exam at the end of upper secondary education, is widely reported in the media by average scores of each school every year. However, it is reasonable to assume that the demands for greater transparency, whether concerning public institutions and good governance or the private sector (e.g. tax havens, corruption etc.), will increase rather than decrease in future societies. The educational sector is not a separate part of that trajectory. Therefore, it is important to continue scrutinizing the formation of the coordinative discourse, since it sets the frames, the possibilities and the limits—what is ‘appropriate’ and how—also for the future policymaking.

The change in the governance of education through increased accountability and transparency has not yet reached Finland. According to our analysis, these pressures simply do not resonate with the Finnish core beliefs on school evaluation and its benefits. Despite the gradual decline in more recent PISA assessments, the coordinative discourse produced by the Finnish elite has been consistent in setting the boundaries for public debate. Even though the Finnish school system has been subjected to critical observation several times in the news this year, neither the guidelines for a national testing system nor publicity issues have been seriously challenged on any front. The master discourse of Nordic egalitarianism, articulated through the autonomy of the Finnish teacher, trust in the system and the detriments of high-stakes testing, accompanied by the PISA success, has so far effectively controlled the faintest neoliberal opinions.

Notes
1. https://siris.skolverket.se.
2. https://www.udannelsesstatistik.dk/grundskolen/overblick. Note: the national testing results are still ordered for restricted use only (Ministry of Education, 2016).
3. https://skoleporten.udir.no/.

Disclosure statement
No potential conflict of interest was reported by the authors.

ORCID
Tommi Wallenius http://orcid.org/0000-0003-0956-5254
Sara Juvonen http://orcid.org/0000-0003-2798-9376
Petteri Hansen http://orcid.org/0000-0003-2209-3618
Janne Varjo http://orcid.org/0000-0003-2528-4241

References
Allen, R., & Burgess, S. (2011). Can school league tables help parents choose schools? Fiscal Studies, 32(2), 245–261.
Berman, S. (1998). The social democratic moment. Cambridge: Harvard University Press.
Blossing, U., Imsen, G., & Moos, L. (Eds.). (2014). The Nordic education model. In ‘A school for all’ encounters Neo-liberal policy. Dordrecht: Springer.
Bracci, E. (2009). Autonomy, responsibility and accountability in the Italian school system. Critical Perspectives on Accounting, 20, 293–312.
Darling-Hammond, L. (2004). Standards, accountability, and school reform. Teachers College Record, 106(6), 1047–1085.
Erkkilä, T. (2012). Government transparency: Impacts and unintended consequences. Basingstoke: Palgrave.
Eurydice. (2009a). National testing of pupils in Europe: Objectives, organisation and use of results. Brussels: Education, Audiovisual and Culture Executive Agency.
Eurydice. (2009b). National Testing of pupils in Europe: Objectives, organisation and use of results. Sweden, Brussels: Education, Audiovisual and Culture Executive Agency.
Hall, P., & Taylor, R. (1996). Political science and the three new institutionalisms. Political Studies, 44(5), 936–957.
Henry, M., Lingard, B., Rizvi, F., & Taylor, S. (2001). The OECD, globalisation and education policy. In International Association of Universities: Issues in higher education, Amsterdam: IAU Press.
Hovdhaugen, E., Vibe, N., & Seland, I. (2017). National test results: Representation and misrepresentation. Challenges for municipal and local school administration in Norway, Nordic Journal of Studies in Educational Policy, 3(1), 95–105.
Hudson, C. (2007). Governing the governance of education – The state strikes back? European Educational Research Journal, 6(3), 266–282.
Jakkusihvonen, R. (2013). Oppimistulosten arvointijärjestelmistä ja niiden kohdetutkimuksista [On evaluation systems of educational outcomes and their challenges]. In A. Räisänen (Ed.), Oppimisen arvioinnin kontekstit ja käytännöt (Vol. 2013, pp. 3). Opetushallitus: Kouluutuksen seurantaporaporti.
Lundahl, L. (2002). Sweden: Decentralization, deregulation, quasi-markets - and then what? Journal of Education Policy, 17(6), 687–697.
March, J., & Olsen, J. (1989). Rediscovering institutions: The organizational basis of politics. New York: Free Press.
Ministry of Education. (2016). Retningslinjer for bruk af data fra de nationale test til statistiske og videnskabelige undersøgelser. [Guidelines for the use of national test data in statistical and scientific research]. ISBN: 978-87-603-3075-9. Ministeriet for Børn, Undervisning og Ligestilling, 2016.
National Agency for Education (2018). SIRIS a source of information for understanding of the world of education. Retrieved 15 Jan 2018, from, <https://siris.skolverket.se/siris/?f?p=SIRIS:33:0>
Norwegian Directorate for Education and Training (2011). OECD review on evaluation and assessment. Frameworks for Improving School Outcome. Country Background Report for Norway.
Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. International Journal of Social Research Methodology, 11(4), 327–344.
OECD. (2011a). OECD Reviews of Evaluation and Assessment in Education: Denmark. Claire Shewbridge, Eunice Jang, Peter Matthews and Paulo Santiago.
OECD. (2011b). OECD Reviews of Evaluation and Assessment in Education: Norway. Deborah Nusche, Lorna Earl, William Maxwell and Claire Shewbridge.

OECD. (2013). Synergies for better learning. In An international perspective on evaluation and assessment. OECD Reviews of Evaluation and Assessment in Education. DOI: http://dx.doi.org/10.1787/9789264190658-en.

Ozga, J. (2009). Governing education through data in England: From regulation to self-evaluation. Journal of Education Policy, 24(2), 149–162.

Ozga, J., Dahler-Larsen, P., Segerholm, C., & Simola, H. (Eds.). (2011). Fabricating quality in education. In Data and governance in Europe. London: Routledge.

Rosenkvist, M. (2010). Using student test results for accountability and improvement: A literature review. (OECD Education Working Papers). Paris: OECD.

Schmidt, V. A. (2008). Discursive institutionalism: The explanatory power of ideas and discourse. Annual Review of Political Science, 11, 303–326.

Schmidt, V. A. (2010). Taking ideas and discourse seriously: Explaining change through discursive institutionalism as the fourth ‘new institutionalism’. European Political Science Review, 2(1), 1–25.

Smith, W. C. (Ed.). (2016). The global testing culture: Shaping education policy, perceptions, and practice. Oxford: Symposium books.

Visscher, A. J. (2001). Public school performance indicators: Problems and recommendations. Studies in Educational Evaluation, 27(2001), 199–214.

Whitty, G., Power, S., & Dalpin, D. (1998). Devolution and choice in education: The school, the state and the market. Melbourne: Acer.