

# Formulation and realisation of evaluation policy: inconcistencies and problematic issues

Guri Skedsmo

Received: 5 May 2010 / Accepted: 14 November 2010 /

Published online: 21 December 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** The introduction of the national evaluation system in the Norwegian education system can be described as a shift in the Norwegian educational policy from the use of input oriented policy instruments towards a more output oriented policy. The new tool-kit consists of different evaluation tools, such as standardised tests, diagnostic tests etc. which create new expectations to schools and principals. On the one hand, this article focuses on the policy purposes of the evaluation tools as they are formulated in policy documents. On the other hand, it investigates principals' perceptions of the use of the evaluation tools to improve student achievements. A text analytic approach focusing on language and discourse is used to analyse policy documents. The analysis of the principals' perceptions is based on a quantitative analysis of survey data and Structural Equation Modeling is applied as an analytic technique. Based on the analysis, questions can be raised whether all the policy purposes can be fulfilled or whether the tools might produce other effects than the objectives assigned to them. Problematic aspects related to the tools' modes of regulations and thereby the choice of tools are discussed, as well as inconsistencies between the policy purposes and the use of tools in practice.

**Keywords** Educational policy · Evaluation · Evaluation tools · Standardised tests · School leaders

## 1 Introduction

In 2005 a national evaluation system was introduced in the Norwegian education. This system includes mostly test based evaluation tools e.g. national tests, screening tests, local knowledge tests, and international comparative achievement tests, but also components such as School-leaving Examination and the Craft Certificate as

---

G. Skedsmo (✉)

Department of Teacher Education and School Research, Faculty of Education, University of Oslo, Oslo, Norway  
e-mail: guri.skedsmo@ils.uio.no

well as information material to key actors in the education system about how to use the test results to improve.

The introduction of the national evaluation system can be described as a shift in the Norwegian educational policy from the use of input oriented policy instruments towards a more output oriented policy. Traditionally, public schooling in Norway was regulated through the Education Act and the national curriculum. These defined the overall purposes of schooling, as well as for the individual subjects (Bachmann et al. 2008; Sivesind & Bachmann 2008). Furthermore, heavy investments in teacher education have also been an important strategy to ensure educational quality. There has, until recently, been no focus on testing student achievements and assessment of outcomes according to indicators of educational quality or standards. Instead, there has been a qualification system which has been based on the examination system and overall assessment grades (Hopmann 2003; Lundgren 2003; Sivesind & Bachmann 2008; Werler & Sivesind 2007).

The increased focus on educational outcomes in terms of student achievements implies concepts of educational quality which in form seems to be defined by expectations about specific outcomes. It also indicates a belief that any divergence between the expected outcomes and the level of achievements can be identified. As such, performance measurement becomes a key part of the evaluation processes. Along with this development, schools are increasingly being perceived as the unit of measurement and the need to make actors such as principals and teachers accountable is emerging as an important aspect of the evaluation processes (Skedsmo 2009). This is due to underlying ideas which presume no change or improvement of practice unless central actors are held accountable for the results achieved. However, compared to the ways of which accountability practices in relation to high stake testing are implemented in other countries, e.g. in England or in the US, there is little pressure put on key actors in the Norwegian education context. However, there is a stronger focus on the municipalities<sup>1</sup> and their responsibilities to follow up on and support schools.

On the one hand, this article investigates the policy purposes of the evaluation tools included in the national evaluation system as they are formulated in policy documents. On the other hand, it examines how Norwegian principals, who are defined in policy documents as one of the key actors responsible for improving educational quality, perceive the use of the evaluation tools. The following two empirical research questions have guided my analysis:

- What are the purposes of the evaluation tools introduced as key elements in the national evaluation system, and how can they be characterised?
- How do principals perceive, understand and respond to the use of evaluation tools and new expectations?

---

<sup>1</sup> In educational policy documents published after 2004, the municipalities are defined as “school owners”. Their main responsibilities within the area of education are defined in terms of adapting the national curriculum to local needs, running in-service training for teachers and school leaders, and ensuring the quality of schooling. It is required that the municipalities establish a system for evaluating and following up the results of the schools.

In the article I argue that the evaluation tools have their own logic which influence on their modes of regulation. Related to the analysis of the policy purposes of the evaluation tools included in the national evaluation system, there is a question whether all these purposes can be fulfilled or whether the tools might produce other effects than the objectives assigned to them formulated policy purposes. I would like to draw attention to some problematic aspects related to the tools' modes of regulations, but also to inconsistencies between the policy purposes and how the tools are perceived by the principals. These inconsistencies are related to the different contexts involved, namely the national education policy context, referred to as the *context of formulation*, and the context of practice, which is seen as the *context of realisation* (Lindensjö & Lundgren 1986, 2000). Since the conditions for realising educational policy are quite different compared to the conditions of formulating the national policy, there will always be inconsistencies between these contexts. The policy context and processes can be characterised by negotiations, renegotiations, and compromises. The principals are embedded in the local school context which influences on how they interpret national policy, make use of new policy instrument and tools, and how they respond to new expectations. However, too much inconsistency could lead to parallel evaluation structures and procedures and lack of coherence.

## 2 Theoretical framework and methodological approaches

The analysis in this article looks at evaluation policy from the angle of the tools. According to Lascoumes and Le Gales (2007), viewing educational policies through the instruments, the methods and concrete tools that structure these policies, has several advantages. Most important, it implies a stronger emphasis on the concrete procedures established to attain national aims, and makes it possible to study educational policies in a more material form. This is particularly useful in a complex education context, characterised by governing processes and interaction among multiple actors. Because of the instruments' properties, e.g. their underlying assumptions, they might produce effects independent of the aims and purposes ascribed to them. This means that they are capable of structuring policy according to their own logic.

Defining policy instruments as devices which “organize specific social relations between the state and those it is addressed to” (Lascoumes & Le Gales 2007, p. 4), has a number of implications. The definition takes into account that the effects the instruments produce depend on how the aims and purposes ascribed to them, and the meanings and representations they carry, are perceived, understood and responded to by key actors involved such as national authorities, municipal administrators, principals and teachers. As such, this approach to policy instruments puts the set of problems posed by the choice and use of instruments at the centre of attention.

Empirically, this study is restricted to recent changes in educational policy and school governing in Norway. The first research question relates to the analysis of policy texts and the information about the national evaluation system presented on the website of the Norwegian Directorate for Education and Training (2005a). A text analytic approach focusing on language and discourse is used to analyse the policy documents. The aim

of the analysis is not only to describe policy intentions and ideas, but to look for embedded meanings (cf. Fitzgerald 2007).

The second research question refers to the analysis of the data from the 2005 survey among Norwegian principals, conducted shortly after the national evaluation system was introduced and the first attempts with the national tests. In this way, the article aims to investigate how evaluation policy unfolds within the fields of leadership and related to the context of practice, as conceptualised by the principals. My analysis includes responses from 540 principals in compulsory education; in other words principals working in primary and lower secondary schools. The response rate was 67% with a national representative sample. Structural Equation modelling is used to investigate patterns in how the Norwegian principals perceive and understand the use of evaluation tools and procedures.

### 3 The comprehensive national evaluation system

The following overall aim is stated for the comprehensive national evaluation system:

...to contribute to quality development on all levels of the compulsory education with respect to adapted teaching,<sup>2</sup> and improved learning outcomes<sup>3</sup> for the individual student (The Norwegian Directorate for Education and Training 2005b. The author's translation).

In addition, the national evaluation system is supposed to provide information for the education sector about the national and local state of progress. This type of documentation is to form the basis for general decision making and for local work on evaluation and development. Furthermore, the national evaluation system is supposed to contribute to increased openness, transparency and dialogue about the school's practice (The Norwegian Directorate for Education and Training 2005b).

The comprehensive national evaluation system comprises various components or tools aimed towards fulfilling different purposes. It is based on the idea that the different tools are intended to provide more information about teaching and learning than a single tool is able to. The different components and their purposes are

---

<sup>2</sup> In several documents published by the Norwegian Directorate of Education and Training in English, the Norwegian concept of "tilpasset oppl ring" is translated as "adapted education" (The Norwegian Directorate for Education and Training, 2008c). I have chosen to use the term "adapted teaching" which is the term used by Peder Haug (2004). This term refers to a broad definition of the concept in terms of a foundation or ideology which is supposed to characterise the entire school system (Bachmann & Haug, 2006). While "education" is a more general term, the term "adapted teaching" always refers to the practice as well as to the intentions of this practice. The use of "adapted teaching" also takes into account that the practice takes place within a community in terms of a heterogenic class or group of students

<sup>3</sup> The Norwegian term "l ringsutbytte" is often translated as "learning outcomes". Both the Norwegian and the English term refer to a broader definition which comprises more than results in terms of student achievements on tests. These terms are also used in ways which take the perceptions of the students into account, and what they have gained from their learning experiences. In some documents published by the Norwegian Directorate for Education and Training, the term "learning dividend" is used (The Norwegian Directorate for Education and Training, 2007). This is, however, an economic term which seems to define learning outcomes as containing concrete and explicit results.

**Table 1** An overview of the various components included in the national evaluation system, as well as their purposes and functions

Tools	Purposes
Screening tests	To reveal the needs for support and adapted teaching related to the individual student as well as the schools
Local tests (summative and formative assessment)	To reveal the level of student achievement with respect to central subject areas as a basis to help the individual student to improve
Information material	Material provided by the National Directorate for Education and Training to help teachers, school leaders and municipalities using the system for learning and development purposes
National tests	To investigate the extent to which students' achievements are coherent with the aims of competencies in the national curriculum To inform students, teachers, parents, school leaders, municipalities, regional authorities and national authorities about the level of achievement as a foundation for improving and development
The school-leaving examination and the craft certificate	Summative evaluation intended to inform the society, employers and education institutions about the achieved level of competency
International comparative achievement studies	To evaluate and compare the level of achievement of Norwegian students to the level of achievement of students in other countries. To form the basis for policy formulation and to develop national quality indicators

presented in Table 1. A similar table is used on the website of the Norwegian Directorate for Education and Training<sup>4</sup> to present the different components and their purposes in Norwegian.

The use of screening tests, which is called “kartleggingsmateriell” in Norwegian, aims to discover what the students need in terms of individual support, adaptation and follow-up. Summative and formative assessment in terms of local tests, which in Norwegian is referred to as “karakter- og læringsstøttende prøver”, are intended to show the level of student achievement in order to improve learning for the individual student. At the same time, these local tests are part of the summative assessment in terms of the overall achievement grades (The Norwegian Directorate for Education and Training 2005a).

The Norwegian Directorate for Education and Training has developed various materials to help teachers, school leaders and municipalities use the evaluation system for learning and development purposes. These materials are also included as a component of the national evaluation system (ibid.).

On the one hand, the national tests aim to investigate the level of students' achievements and the degree to which the aims of competencies in the national curriculum are accomplished. On the other hand, the results of the national tests are expected to inform students, teachers, parents, school leaders, municipalities, and regional and national authorities about the level of student achievement. This information is regarded as vital for further improvement and development (ibid.).

Tools such as the School-leaving Examination and Craft Certificate represent traditional forms of assessment within the Norwegian education system. As a part of

<sup>4</sup> [http://udir.no/templates/udir/TM\\_Artikkel.aspx?id=1399](http://udir.no/templates/udir/TM_Artikkel.aspx?id=1399)

the national evaluation system, this component is supposed to inform the authorities and the public about the level of student achievements at an aggregated level. For the individual student, this component is still a sorting mechanism in terms of providing information to society, employers and educational institutions pertaining to the achieved level of competence (The Norwegian Directorate for Education and Training 2005a).

The purpose of participating in international comparative achievement studies, e.g. PISA, TIMSS, PIRLS, is to evaluate and compare the level of achievement of Norwegian students with those of students in other countries. This information is to provide a basis for formulating the national educational policy, and to develop indicators for the national quality measurement (*ibid.*).

All these tools are presented on the website of the Norwegian Directorate for Education and Training as components of the comprehensive system (The Norwegian Directorate for Education and Training 2005b). According to the formulated purposes, they are to serve different functions. However, very little information is provided about how these tools are supposed to work together in a coherent way.

Furthermore, as a support for the schools and municipalities, a web-based service called the School Portal, is provided by the Norwegian Directorate of Education and Training. This service is to inform municipalities and schools about their results<sup>5</sup>. Actually, the national tests and the School Portal are defined by the Norwegian Directorate for Education and Training as the most important components of the national evaluation system (The Norwegian Directorate for Education and Training 2005b).<sup>6</sup> When the national tests were attempted for the first time, it was possible to see the results of each school and compare different schools. Before the second round of national tests, which took place after a change of government from the conservative to the social democratic coalition government, it was decided that the school results were not to be publicly accessible. While it is now possible to access the results at the municipality and county levels, the results of the individual schools are only available to the municipality, the local school, students and their parents.

On the same website, it is also pointed out that the state supervision of schools, which is delegated to the regional Educational Offices, is an important part of the national comprehensive evaluation system (The Norwegian Directorate for Education and Training 2005b). In what way the state supervision of schools is linked or plays a part in the evaluation system is, however, not described.

#### 4 Purposes and characteristics of the evaluation tools

Included in the national evaluation system is a combination of so-called new and traditional tools. The School-leaving Examination and the Craft Certificate represent traditional tools within the Norwegian education system. The screening tests, the local

<sup>5</sup> The School Portal was launched as a separate web service system in 2004, but in 2007 the system was revised and re-launched as an integrated part of the website of the Norwegian Directorate of Education and Training: <http://skoleporten.utdanningsdirektoratet.no/english/Sider/WhatisSkoleporten.aspx>

<sup>6</sup> <http://udir.no/Artikler/Nasjonalt-kvalitetsvurderingssystem-for-grunnopplaringen/>

tests and the information material can also be referred to as traditional tools, but the ways in which they are used seem to have changed. National tests and international comparative achievement tests belong to the category of “new” tools.

With respect to the stated purposes for each of the tools, the evaluation system can be characterised by two main agendas. One agenda is characterised by aiming to *provide* information to gain oversight over the level of student achievements. The other agenda is linked to *using* the information to improving the overall system as well as the performance of the individual student. It could be argued that one way of looking at these two agendas is to say that one presumes the other. To provide data to gain oversight and then use the information as a basis for further improvement and development, can, theoretically be seen as purposes of evaluation working together rather than against each other. Indeed, the way the purposes are formulated, indicate that these agendas are intended to work together.

Although all the purposes stated involve both agendas, the first agenda seems to be under-communicated. Providing data is described in terms of “revealing needs for individual support”, “mapping the students’ knowledge” and “investigating the level of basic competencies” (The Norwegian Directorate for Education and Training 2005a). The use of the verbs “revealing”, “mapping” and “investigating” indicate neutral processes of providing information which make the results seem “objective”. The tools appear as data gathering methods, and the parts of the processes which include analysis and interpretation of the results as well as finding appropriate approaches to meet the needs are concealed. The choice of words such as “revealing needs” leads one’s attention towards the improvement agenda. The way the purposes are presented, it seems as if the first agenda providing information to gain oversight is just one step on the way to improvement and development.

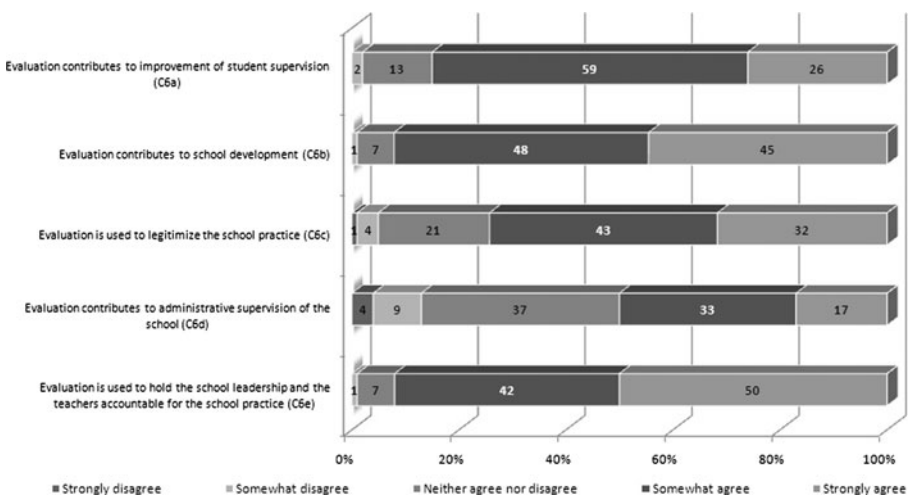
Examining the improvement agenda more closely, it basically implies providing information to all those involved, who then will use it as a basis for improvement. As such, the data concerning the level of student achievements are seen as the key to improving the overall system as well as the results for the individual student. It seems as if the evaluation system is intended to work reflexively in two ways. First, it presumes that achievement scores comprise the information which is needed to improve the overall system as well as individual performances. Second, in providing these data to the different actors; national authorities, municipalities, principals and teachers, it is presumed that the information automatically makes sense to them, and it is also taken for granted that they know how to use the data as well as what kind of actions they need to take in order to improve. As such, the system does not clearly differentiate between the needs of the individual student and the needs of the system, in terms of providing as well as using information for improvement purposes.

The fact that the international comparative achievement studies are included as a component in the national evaluation system brings in yet another aspect with respect to the oversight agenda. This shows that providing data and information to gain oversight includes an element of *comparability*, which is internationally oriented. It is an explicit aim to compare the level of achievement of Norwegian students with the levels of student achievement in other countries. In this way, the results of these studies are used as “benchmarks” or international “standards”. In Norway, these “standards” are in turn intended to be used to develop national indicators and to form the basis for developing the national education policy. This is

linked to the improvement agenda in terms of improving the quality measurement of the Norwegian education system, as well as to the goal of ensuring that the national policy is “on track”. As such, the comparability aspect represents a need to document the quality of education based on the level of achievement of 15-year olds and to use this to monitor the national progress (or lack of progress) compared to other countries. Linked to variables such as resources, these results also provide information about cost efficiency and the extent to which different education systems can be characterised as successful.

This comparability aspect is also emphasised within the national context. The stated purpose for the use of national tests pertains to evaluating the extent to which the schools succeed in developing basic competencies among their students according to national aims. The way the results are used indicates that comparisons seem to be rather strong driving motivational forces. For instance, when summing up the results from the national tests 2007, the Norwegian Directorate for Education and Training, emphasises comparing the results of boys and girls (The Norwegian Directorate of Education and Training 2008). Nothing is mentioned about the extent to which national aims have been reached. The gender differences in level of achievements within the national context are also compared to the gender differences on the PISA results. Even though the results of the national tests are not publicly accessible, at least not at the school or individual level, they are still used to compare counties, municipalities and to rank schools if the media gets hold of the results. To a certain extent, they are therefore used by the schools for comparison with other schools within the same municipality (cf. Elstad 2009).

Looking closely at some of the purposes linked to the way of gathering data and the type of information provided, questions can be raised about the extent to which all the purposes can be fulfilled. Questions can also be raised regarding the tools’ underlying assumptions, and whether they may produce effects other than the objectives assigned to them.



**Fig. 1** The principals’ perceptions of evaluation and its’ contribution ( $N=540$ ). 8-13 respondents are registered as missing on these questions



## 5 Norwegian principals' perceptions of functions of evaluation and evaluation tools

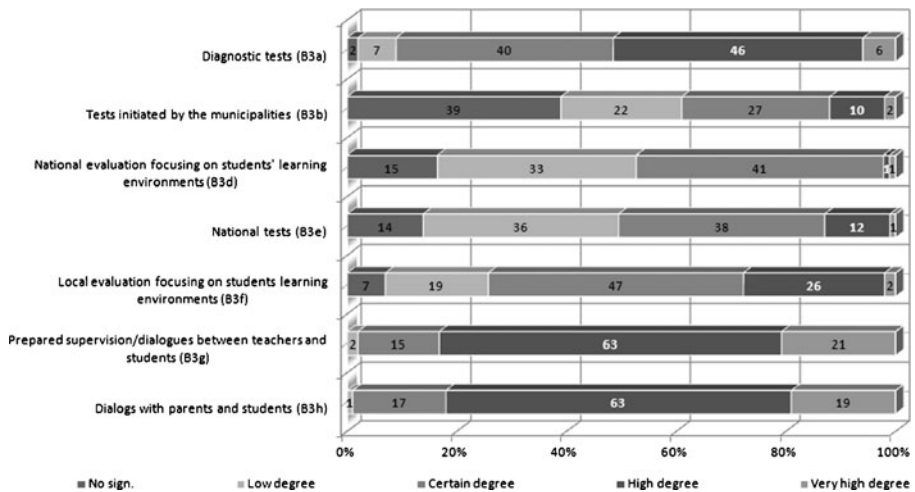
In order to map how the principals perceive the work on evaluation at their school, statements covering some various aspects of evaluation in terms of contribution were formulated. These describe functions of evaluation in terms of control, and development, whether the premises for evaluative work are centrally or locally defined, and the extent to which the premises are decided upon, or influenced by politicians or professionals (cf. Lundgren 1990, 2003). All these aspects can also be viewed as different tensions which mean that evaluation is not either control or development, but includes both elements. The principals' responses on different items about the function of evaluation are presented in Fig. 1.

In general, the table indicates that evaluation is an important part of school activities. Most of all, evaluation seems to be used to hold the school leadership and teachers accountable for school's practices, since most of the respondents have answered positively to this statement. Evaluation also seems to contribute to school development and to improving student supervision. More than two-thirds of the respondents said yes to the statement that evaluation is used to legitimise school's practices. However, 21% responded that they neither agree nor disagree to this statement.

The fact that approximately one-third of the respondents neither agree nor disagree, and that 13% partially disagree or strongly disagree to the statement about evaluation contributing to an administrative inspection of their school, is linked to the fact that there has not been a tradition of school inspection in Norway. While many municipalities have established systems to follow up schools, many have not. Indeed, when compared to the accountability systems in other countries, there is nothing at stake for Norwegian principals. The fact that so many of the respondents still find that evaluation contributes to holding principals and teachers accountable for school's practices could imply that there are different ways of being held accountable, which has to be seen in relation to the characteristics of the Norwegian education context.

As mentioned earlier, the policy purposes of the different tools included in the national evaluation system emphasise improving student achievement on the individual level as well as improving the quality of the overall education system. In the survey the principals were asked to assess the contribution of the different tools to improving student achievement. In addition to some of the components in the national evaluation system, questions about tools such as dialogues involving teachers, students and their parents were also included. These are concrete tools used by the teachers. According to the Education Act, all students are entitled to have such dialogues with their teachers (The Education Act 1998). Since this is part of the teachers' work on following up the individual students, and since these tools do not provide information which is reported to national authorities, they are not included as elements in the national evaluation system.

According to the principals, dialogue-based tools seem to contribute highly to improving student achievements (please see Fig. 2). More than 80% of the respondents answered that prepared supervision/dialogues between the teachers and the students,



**Fig. 2** Principals' perceptions of evaluation tools in terms of contribution to student achievement ( $N=540$ ). (There are many missing values registered on some of the variables. For instance, B3b Tests initiated by the municipalities have 79 missing values. This is probably due to the fact that many municipalities do not have established systems with local tests, and therefore the principals lack experience with this type of tool. In contrast, the two dialog tools have only five and seven respondents registered as missing)

and dialogues which include parents and students, contributed to a very high degree to improving student achievements. The respondents also seem to be positive towards the use of diagnostic tests. All in all, three-fourths of the respondents perceive that diagnostic tests contribute to improving student achievements.

Furthermore, locally based evaluation by the students, which focuses on their learning environments etc., seems to be perceived as somehow influential. Twenty-six percent answered to a high degree, 47% to a certain degree, while 19% think that to a low degree, with regard to this tool contributing in improving student achievement. The responses about national evaluations focusing on learning environment and national tests, as well as tests initiated by the municipalities show another pattern. Forty-one percent of the respondents perceive national evaluation by the students contributes to a certain degree in improving student achievement, 33% answered to a low degree, and 15% answered that this tool has no significance. Similarly, 38% of the respondents answered that national tests to a certain degree contribute to improving student achievement, 36% to a low degree, and 14% that national tests do not contribute at all. Tests initiated by the municipalities seem, according to the principals' answers and in relation to the other tools, to have less of an effect on improving student achievements. However, with regard to tools such as national evaluations focusing on learning environment and national tests, as well as tests initiated by the municipalities, it has to be taken into account that many of the respondents have limited experiences with these tools. For example, national tests had only been tried out once before the survey was conducted, and many of the municipalities have not established their own local system of testing. Nevertheless, the principals can still have an opinion about these tests as tools and to what extent they contribute to improving student achievement.

## 6 A model illustrating functions of evaluation and purposes of the evaluation tools as conceptualised by the principals

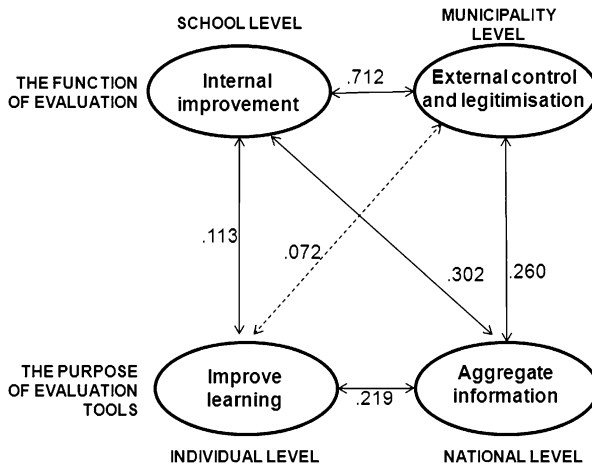
As mentioned earlier, SEM was used to analyse the quantitative data and to investigate the patterns and dimensions of how the principals responded to the different questions. Using the items, or the observed variables presented above, four latent variables were constructed; ‘improve learning’, ‘internal improvement’, ‘administrative control and legitimisation’, and ‘aggregate information’. These latent variables were then put into a joint model in order to investigate the interrelationship between the functions of evaluation and the purposes of the evaluation tools as conceptualised by the principals<sup>7</sup>. The different constructs also refer to the evaluation practice and to the use of tools linked to different administrative levels or areas of work, which in turn are related to improving learning for the individual student, improving the school, and the role of the municipality with respect to evaluation and the system as a whole.

All these areas are also included in the national evaluation system, and, as pointed out earlier, the system and the tools in use are intended to work in a reflexive way. For instance, national tests are intended to provide aggregated information for the overall system to determine the extent to which the level of student achievements are in line with the competency aims stated in the national curriculum. At the same time, the information provided by these tests is intended to be used by the municipalities to improve the quality, and also by the schools to help the individual student to improve. This intended reflexivity implies that the different areas of work are interconnected. The question is how strong the couplings between the different areas of work are, and if the information provided in one area of work is perceived as useful by others in order to change and improve their practice. Based on the perceptions of the principals, there seem to be discrepancies between policy intentions and evaluation practice. Figure 3 illustrates certain aspects of evaluation procedures as conceptualised by the principals.

The aspects are related to functions of evaluation and purposes of evaluation tools which represent different areas of work. ‘Improve learning’ relates to the individual student, internal improvement relates to the school level, ‘administrative control and legitimisation’ refer to the role of municipality with respect to evaluation, and ‘aggregate information’ refers to the overall system. The way these latent variables interrelate illustrates that different areas of work are linked together<sup>8</sup>. Providing aggregated information for the system is linked to evaluation in terms of external control and legitimisation, which points to the area of work of the municipality and internal improvement within the school. Based on the principals’ perceptions, it

<sup>7</sup> According to the Goodness-of-fit Indices, the overall fit between the model and the data is good. The model includes 82 parameters. The Chi-square is 61.39 with 70 degrees of freedom. The P-value is 0, which is highly significant and indicates that the model should be rejected as fitting the data, but the CFI has a value of .957 and the RMSEA is .035. Both the lower (.024) and the upper limit (.039) of the confidence interval of RMSEA are below .05. SRMR is .039. For detailed information about the analysis conducted, please see Skedsmo (2009).

<sup>8</sup> The relationship between the latent variables is generally characterised by having low covariances due to the fact that the models are based on cross sectional data. However, the stability of the latent variables constructed on the basis of these data can, to some extent, compensate for these low correlations. An essential part of the analysis is also how these models are made sense from and theoretically constructed and interpreted (cf. Gustafsson 2009).



**Fig. 3** The joint model showing the relations between the different latent variable models

seems as if the interplay between the different components of evaluation practice and the different functions of evaluation and evaluation tools involving different areas of work, is characterised by some degree of reflexivity.

However, there is a weak relationship between the evaluation of external control and legitimisation at the municipality level, and improved learning for the individual. This could indicate that this function of evaluation, which is related to municipality procedures, is more closely linked to hierarchical, administrative structures which include the school as part of an administrative system.

According to the purposes of evaluation tools, tools such as national tests aim towards improving outcomes for the individual as well as the overall education system. Due to certain problems with this tool, there is a question about the extent to which aggregated information for oversight purposes is actually linked to learning for the individual. Questions can be raised whether it is possible to use aggregated information, which of course is based on individual performances, to bring the information back to the individual in a way that helps the individual to improve. The individual improvement will probably depend on other factors, such as support from teachers and the work of the students. It can be argued that the interrelation between the constructs 'improve learning' and 'aggregate information' is problematic in theory as well as in practice.

With regard to reflexivity between the areas of work, it is a question whether, based on the principals' perceptions, the reflexivity is mainly reflected between the following elements; 'aggregate information', 'external control and legitimisation', and 'internal improvement'. These areas of work are also key elements of the principals' responsibility. This is because principals are not generally involved in classroom practices and activities on a daily basis, but focus more on the school level. Nevertheless, it is questionable as to why the link between 'internal improvement' within the school and 'improve learning' for the individual is not stronger. This could be a sign that activities to develop the schools and the evaluation of these activities are not necessarily connected to improving learning for the individual, and that this connection perhaps is more indirect. The interrelations

between these elements could reflect an administrative oriented evaluation system which has little to do with teaching practice, and not to mention actions taken to improve individual student achievements. The question is how are the linkages between this administrative oriented system and, for instance, the work of the teachers? If these couplings are weak, the consequence could be parallel systems in terms of teaching and learning activities and co-existing administrative systems.

## 7 Inconsistencies between the arenas of policy formulation and policy realisation

Related to the analysis of the policy purposes of the evaluation tools included in the national evaluation system, I raised questions whether all these purposes could be fulfilled, or whether the tools might produce other effects than the objectives assigned to them. In this last part of the article, I would like to draw attention to some problematic aspects of the formulated policy purposes related to the tools' modes of regulations, but also to inconsistencies between the purposes and how the tools are perceived by the principals. These inconsistencies are related to the different contexts involved, namely the national education policy context, referred to as the *context of formulation*, and the context of practice, which is seen as the *context of realisation* (Lindensjö & Lundgren 1986, 2000). As pointed out in the introduction, there will always be inconsistencies between these contexts due to the differences in the conditions for realising educational policy and formulating policy.

The first inconsistency I would like to draw attention to is linked to the policy agenda in terms of using the information provided by the national evaluation system to improve educational outcomes. As described earlier, the analysis of purposes of the tools revealed a dominating improvement agenda, while elements such as gaining oversight and monitoring educational outcomes were concealed. Looking at the principals' perceptions of the function of evaluation and the use of some of the tools included in the national evaluation system, these, however, reflect an administrative oriented system which includes oversight and control elements related to national and local authorities, as well as to the development of schools. While internal development seems to be the function of evaluation at the school level, external control and legitimisation are associated with municipalities' use of evaluation.

The second inconsistency is related to the differentiated needs of information to improve. The way the purposes of the evaluation tools are formulated, means that the national evaluation system does not differentiate clearly between the needs of the individual student and the needs of the system, in terms of providing as well as using information. Paradoxically, the necessity to differentiate between the need for information linked to different levels or actors in the education system have been emphasised by several policy texts which paved the way to establishing the national evaluation system, in particular White Paper No. 28 (1998–99) (cf. Skedsmo 2009). This was actually also mentioned by contributors in the EMIL-project<sup>9</sup> (cf. Kogan 1990).

<sup>9</sup> The EMIL-project was an initiative taken by the Royal Ministry of Education and Church Affairs, following the 1988 Report of the OECD examiners on educational policy in Norway. The aim of the project was to develop an evaluation model for the Norwegian education system.

The use of evaluation tools, as perceived by the principals, shows a more nuanced picture. It seems as if the principals conceptualise the purposes of different tools and functions of evaluation as linked to levels of the education system, also referred to as different areas of work. While some of the tools, which provide oversight, are directed towards improvement of system variables, other dialogue based forms are related to improving learning outcomes for the individual student. However, as mentioned earlier, this administrative oriented evaluation system is only loosely connected to tools used in classroom practices to improve student learning.

The third inconsistency is related to the use of the outcomes, and in particular for comparisons, related to the formulated purposes of the tools. For instance, the focus in the reporting about results of national tests is not consistent with the formulated purposes of the tools. The presentations of the results of the national tests on the website of the Norwegian Directorate for Education and Training focus on comparisons related to gender rather than the extent to which aims of competencies in the national curriculum have been reached, as is formulated as the main purpose of this tool (cf. The Norwegian Directorate for Education and Training 2008a, b). Several reports written by the different universities and centres responsible for the national tests, emphasise the linkages between the concrete tasks and items included in the tests and the aims of competencies in the national curriculum, the Knowledge Promotion (K06), and they discuss to some degree the validity of the tasks. Some of the reports also bring up difficulties related to interpreting complex aims in the national curriculum, which have consequences for developing tasks. However, the reports are “technical reports” and the extent of which national aims have been attained or not, is not discussed (cf. Moe 2009; Ravlo et al. 2008a, b; The Norwegian Reading Centre 2008; Vagle et al. 2008). This is obviously a complicated task, but it is still an explicit purpose formulated for national tests as a tool included in the evaluation system.

## 8 Implications for policy and research

Some implications can be drawn for policy and future research. I will here emphasise two points. First of all, I will draw attention to the choice of evaluation tools and that greater awareness towards the tools’ models of regulation is needed in order to fulfill the aims and purposes of establishing a national evaluation system. The choice of tools need to reflect different needs for information. The tools included in a *national* evaluation system should perhaps reflect the needs for information to inform policymaking. This implies monitoring of educational outcomes to gain oversight which can be used as a foundation for decision making. To improve individual outcomes requires other tools and sources of information, not at least processes of feedback and support by the teachers.

My second point is related to the relationship between policy and research. In our time research claims to inform policy by evidence. This claim relates to the interplay between the social realities of science, politics and education and where the methodology of evaluation seems to be of core significance for evidence based policy and also school governing. It is a question if statistical data on students’ performances provide sufficient information to form a basis for policy making as

well as improvement of educational outcomes. As pointed out in this article, evaluations do not provide objective data. Instead, they imply assigning value to something observed. There is a question about the role of professional perspectives in the processes of analysing and interpreting evaluation results as well as judgements related to implications for policy and practice. Advanced evaluation tools and techniques can never replace professional judgements!

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Bachmann, K., Sivesind, K., & Bergem, R. (2008). Evaluering og ansvarliggjøring i skolen. In G. Langfeldt, E. Elstad, & S. Hopmann (Eds.), *Ansvarlighet i skolen. Politiske spørsmål og pedagogiske svar* (pp. 94–122). Fagernes: Cappelen Akademisk Forlag.
- Bachmann, K. E., & Haug, P. (2006). *Forskning om tilpasset opplæring elektronisk ressurs*. Volda: Høgskulen i Volda.
- Elstad, E. (2009). Schools which are named, shamed and blamed by the media: School accountability in Norway. *Educational Assessment, Evaluation and Accountability*, 21(2), 173–189.
- Fitzgerald, T. (2007). Documents and documentary analysis: reading between the lines. In Briggs, A. R. J. & Coleman, M. (Eds.), *Research Methods in Educational Leadership and Management* (pp. 278–294). Los Angeles: Sage Publications.
- Gustafsson, J.-E. (2009). Strukturella ekvationsmodeller. In G. Djurfeldt & M. Barmark (Eds.), *Statistisk Verktøyslåda II*. Lund: Studentlitteratur.
- Haug, P. (2004). *Results from the evaluation of reform 97*. The Norwegian Research Council.
- Hopmann, S. (2003). On the evaluation of curriculum reforms. In P. Haug & T. A. Schwandt (Eds.), *Evaluating educational reforms. Scandinavian perspectives* (pp. 111–132). Greenwich: Information Age.
- Kogan, M. (1990). Fitting evaluation within governance of education. In M. Granheim, M. Kogan, & U. P. Lundgren (Eds.), *Evaluation as policymaking. Introducing evaluation into a national decentralised educational system* (pp. 234–245). London: Jessica Kingsley Publishers.
- Lascoumes, P., & Le Gales, P. (2007). Introduction: Understanding public policy through its instruments— from the nature of instruments to the sociology of public policy instrumentation. *Governance: An International Journal of Policy, Administration, and Institutions*, 20(1), 1–21.
- Lindensjö, B., & Lundgren, U. P. (1986). *Politisk styrning och utbildningsreformer. Gymnasieskola i utveckling (Vol. B 86:3)*. Stockholm: Skolöverstyrelsen och Liber Utbildningsförlaget.
- Lindensjö, B., & Lundgren, U. P. (2000). *Utbildningsreformer och politisk styrning*. Stockholm: HLS Förlag.
- Lundgren, U. P. (1990). Educational policy-making, decentralisation and evaluation. In M. Granheim, M. Kogan, & U. P. Lundgren (Eds.), *Evaluation as policymaking. Introducing evaluation into a national decentralised educational system* (pp. 23–41). London: Jessica Kingsley.
- Lundgren, U. P. (2003). The political governing (Governance) of Education and Evaluation. In Haug, P. & Schwandt, T. A. (Eds.), *Evaluating Educational Reforms: Scandinavian Reforms*: Information Age Publishing Inc.
- Moe, E. (2009). *Teknisk rapport og analyse av resultat frå nasjonale prøver i engelsk (Technical report and analysis of national tests in English)*: University of Bergen.
- Ravlo, G., Johansen, O. H., Andersen, T., & Vinje, B. (2008). *Nasjonal prøve i regning 8. trinn (National test in numeracy, 8th grade)*: The National Centre for Mathematics Education.
- Ravlo, G., Onsrud, G., Bondø, A., Berg, G., et al. (2008). *Nasjonal prøve i regning 5. trinn (National tests in numeracy, 5th grade)*: The National Centre for Mathematics Education.
- Sivesind, K., & Bachmann, K. (2008). Hva forandres med nye standarder? Krav og utfordringer med Kunnskapsløftets læreplaner. In G. Langfeldt, E. Elstad, & S. Hopmann (Eds.), *Ansvarlighet i skolen. Politiske spørsmål og pedagogiske svar* (pp. 62–93). Fagernes: Cappelen Akademisk Forlag.

- Skedsmo, G. (2009). *School governing in transition? Perspectives, purposes and perceptions of evaluation policy*. Dissertation. University of Oslo.
- The Education Act (1998): 20th of June 2008 No. 48).
- The Norwegian Directorate for Education and Training. (2005a). A coherent national test and evaluation system Retrieved February 5, 2005, from [http://udir.no/templates/udir/TM\\_Artikkel.aspx?id=1399](http://udir.no/templates/udir/TM_Artikkel.aspx?id=1399)
- The Norwegian Directorate for Education and Training. (2005b). Nasjonalt kvalitetsvurderingssystem for grunnsopplæringen (A national system for quality assessment in comprehensive education). February 14, 2008, from [http://udir.no/Artikler/\\_Vurdering/Nasjonalt-kvalitetsvurderingssystem-for-grunnsopplaringen/](http://udir.no/Artikler/_Vurdering/Nasjonalt-kvalitetsvurderingssystem-for-grunnsopplaringen/)
- The Norwegian Directorate for Education and Training. (2007). *The Education Mirror. Analysis of primary and lower and upper secondary education in Norway*.
- The Norwegian Directorate for Education and Training. (2008a). Analyse av resultater fra nasjonale prøver 2007 (Results of the national tests 2007). Retrieved March 15, 2009, from [http://udir.no/upload/Nasjonale\\_prover/Analyse\\_av\\_resultater\\_fra\\_nasjonale\\_prover\\_2007.pdf](http://udir.no/upload/Nasjonale_prover/Analyse_av_resultater_fra_nasjonale_prover_2007.pdf)
- The Norwegian Directorate for Education and Training. (2008b). Analyse av resultater fra nasjonale prøver 2008 (Analysis of the results of the national tests 2008). Retrieved March 15, 2009, from [http://udir.no/upload/Nasjonale\\_prover/2008/analysen/NP08\\_Analyse\\_sammendrag.pdf](http://udir.no/upload/Nasjonale_prover/2008/analysen/NP08_Analyse_sammendrag.pdf)
- The Norwegian Directorate for Education and Training. (2008c). *A Norwegian-English dictionary for the education sector*.
- The Norwegian Directorate of Education and Training. (2008). Analysis of results from the national tests. Retrieved February 21, 2009, from [http://www.udir.no/upload/Nasjonale\\_prover/Analyse\\_av\\_resultater\\_fra\\_nasjonale\\_prover\\_2007.pdf](http://www.udir.no/upload/Nasjonale_prover/Analyse_av_resultater_fra_nasjonale_prover_2007.pdf)
- The Norwegian Reading Centre. (2008). *Den nasjonale prøven i lesing på 5. trinn (The national test in Literacy at the 5th grade)*: University of Stavanger.
- Vagle, W., Roe, A., & Narvhus, E. (2008). *Den nasjonale prøven i lesing på 8. trinn (The national test in Literacy at the 8th grade)*: Department of Teacher Education and School Development, University of Oslo.
- Werler, T., & Sivesind, K. (2007). Norway. In H. Döbert, W. Hörner, B. von Kopp, & W. Mitter (Eds.), *The education system of Europe*. Dordrecht: Kluwer.