

## Chapter 9

# Improving Equity Through National-Level Assessment Initiatives



G. A. Nortvedt, K. B. Bratting, O. Kovpanets, A. Pettersen, and A. Rohatgi

**Abstract** This chapter investigates how a national-level assessment initiative may improve equity in early years numeracy education, taking the Norwegian mapping tests for primary grades 1–3 as an example. Three assessments, one test for each grade level, were launched in the 2013–2014 school year and have been used every year since. In accordance with Nordic model principles, the test content is available to teachers to ensure familiarity with the test content and the formative use of the assessment outcomes to improve teaching and learning for students identified as at risk of lagging behind. Analysis of student data reveals that, 6 years after the first implementation, no inflation can be seen in test scores. Thus, an exposed assessment may remain robust within an educational system that aspires to transparency, such as the Norwegian one. However, analyses of interview data and achievement data reveal that teachers often struggle to use the assessment outcomes to improve teaching. These results suggest that the initiative to improve equity in primary school numeracy education depends on teachers' assessment literacy. In accordance with Nordic model principles, schools have significant autonomy and are responsible for identifying professional development needs for their teachers. This research confirms the dilemmas in the Nordic model between national-level and local initiatives and responsibilities.

**Keywords** At-risk students · Numeracy · Mapping tests · Assessment for learning · Teachers' assessment literacy · Equity in education

---

G. A. Nortvedt (✉) · K. B. Bratting · O. Kovpanets · A. Pettersen · A. Rohatgi  
Department of Teacher Education and School Research, University of Oslo, Oslo, Norway  
e-mail: [g.a.nortvedt@ils.uio.no](mailto:g.a.nortvedt@ils.uio.no)

© The Author(s) 2020  
T. S. Frønes et al. (eds.), *Equity, Equality and Diversity in the Nordic Model of Education*, [https://doi.org/10.1007/978-3-030-61648-9\\_9](https://doi.org/10.1007/978-3-030-61648-9_9)

## 9.1 Introduction

The Nordic Education Model is grounded in a social democratic ideology and an egalitarian philosophy. Its core values are equity and equal opportunities, inclusion and social justice, embedded in national school laws and curriculum documents to define ‘A School for All’ that ensures all students are given opportunities to reach their maximum potential (Imsen, Blossing, & Moos, 2017; Telhaug, Mediås, & Aasen, 2006). As such, the educational authorities in the Nordic countries implement policies and tools that not only describe educational equity in a Nordic context but also aims to assist schools in striving for equity. In Norway, for instance, national mapping tests in numeracy are available at the primary school level as part of the Norwegian quality assessment system (NQAS). This is not a unique situation; national governments often implement assessment strategies or policies to enhance students’ opportunities to learn (Nortvedt & Buchholtz, 2018). The three mapping tests, one for each of the grade levels 1–3, are designed to identify students at risk of lagging behind who would benefit from more targeted teaching. Therefore, the tests are conducted with the aim of offering all students the opportunity to be successful in learning, and as such, improving equity in learning opportunities. Each test is accompanied by support material<sup>1</sup> for the teachers and schools to enhance the schools’ efforts as they strive to improve mathematics education for all. The mapping tests differ from many other national-level assessments in some important aspects. For instance, the test data are owned by the local school and not reported in national league tables, and test results should be used formatively (Blömeke & Olsen, 2018).

After a period of test development, piloting and standardisation, the same tests remain in use for at least 5 years consecutively. Moreover, the tests have a high ceiling effect by design, ensuring that targeted students can solve many of the tasks in the test. This means that, unlike typical screening tests, the Norwegian mapping tests provide teachers with information about what identified students know and can do (Nortvedt, 2018). Over time, test content is expected to become highly familiar to teachers. Such transparency connected to national-level initiatives is within the Nordic model principles (Telhaug et al., 2006) and supposed to foster equity. Moreover, transparency enables teachers to further develop their assessment literacy due to opportunities to work with the test content and results.

In this chapter, we relate equity to the policy level and policy-level initiatives. In particular, we address whether national policy initiatives and assessment tools can contribute to equity in schools. As it is the teachers who administer the tests and interpret and use the test outcomes to inform their teaching, their work with the

---

<sup>1</sup>Teacher guides and a national website hosted by the Directorate for Education and Training. While the Grade 2 test is compulsory, grade 1 and 3 tests are voluntary. Still, almost all schools use all three assessments. Taken together, in this chapter we refer to the tests and supplementary material as the assessment

mapping tests is an important part of the national-level initiative. Indeed, trust in teachers to take on this responsibility is embedded in the Norwegian initiative.

We are aware that, with the implementation of such national assessment tools as mapping tests, there is a question about the extent to which they contribute to equity. Both the quality of the assessment and their use may be an issue (Stobart, 2008). Moreover, previous research has shown that teachers' assessment literacy is a critical aspect of their use of assessment data (Popham, 2009). As such, our aim with this chapter is to discuss how, through such national-level initiatives as the Norwegian mapping tests, an education system can enhance equity regarding student learning opportunities. For this purpose, we draw on analysis of student assessment data and teacher interviews.

## 9.2 Theoretical Framework

This section presents an overview of previous research that serves as a framework for our study. Key aspects of equity, assessment for learning and assessment literacy are discussed before presenting previous research on national-level assessment initiatives in the Norwegian context.

### 9.2.1 *Equity, Equality and Inclusion in Education*

The term *equity* is frequently used in both educational research in general and in mathematics education in particular, but often, no clear definition is provided, and the term is used in relation to different issues (Buchholtz et al., 2020; Espinoza, 2007; Roos, 2018). Moreover, equity is often used interchangeably with *equality*, causing confusion and ambiguity in the research literature (e.g. Espinoza, 2007; Zhu, 2018). We follow Rousseau and Tate (2003), who state that equity is associated with fairness or justice in terms of provision of education, while equality is related to sameness, non-discrimination or the state of being equal. Samoff (1996) highlights how equitable education necessitates structural inequalities, for example, to offer adapted education and differentiation.

Some teachers may consider equity in terms of inclusion (Nortvedt & Wiese, 2020). In mathematics education research, the concept of inclusion can refer to both inclusion in society (taking part in the classroom) and inclusion in the form of adapted teaching (Roos, 2018). This is in line with Espinoza's (2007) argument that a set of definitions and conceptualisations should be used that address different dimensions and stages of the educational process rather than striving for a unique understanding of equity and equality. Further, the National Council for Teachers of Mathematics research team argues that equity includes components related to both conditions of learning and outcomes. Their main concern is 'how mathematics education research can contribute to understanding the causes and effects of inequity, as

well as strategies that effectively reduce undesirable inequities of experience and achievement in mathematics education’ (Gutstein et al., 2005, p. 94). According to Zhu (2018), individualised approaches are necessary to achieve equity in mathematics education, taking into account differences in students’ individual needs and providing differentiated treatments rather than regarding and treating all students equally.

### 9.2.2 *Assessment for Learning*

Assessment for learning (AfL) is an important tool to adapt teaching and learning activities to the needs of the individual student. As defined, AfL constitutes ‘all those activities undertaken by teachers and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged’ (Black & Wiliam, 1998, pp. 7–8). Further, Wiliam (2011) argue that the most important purpose of educational assessment is to serve and support learning. Previous studies have shown that good assessment practices can lead to improved learning (Hattie, 2009; Hattie & Timperley, 2007), including improved achievement and understanding in mathematics (Wiliam, 2007). As such, many educational systems have attempted to implement such assessment practices as AfL, but research shows that learning how to practice AfL is challenging for teachers (Hopfenbeck et al., 2017; Nortvedt, Santos, & Pinto, 2016). AfL is strongly connected to ideas of equity in education. Formative use of assessment data should result in targeted interventions and ensure that all students are engaged in challenging mathematics learning (Heritage & Wylie, 2018).

The term *mapping tests* traditionally denotes assessments that are used to identify (map) what students can do (Ginsburg, 2016), with mathematics mapping tests often focussing on student misconceptions (Burkhardt & Schoenfeld, 2018), for instance, in relation to understanding numbers and number operations (e.g. Wiliam, 2007). As such, mapping tests have traditionally been used in the Nordic countries to provide tools for teachers that can be used to inform teaching (Räsänen et al., 2019). However, Gersten et al. (2009) claim that mapping tests only have an effect when followed up with targeted interventions. In other words, implementing national-level assessments alone is not sufficient to improve equity.

In addition to mapping tests, screening tests have been used in special needs education to identify students at risk of learning difficulties or lagging behind (Gersten et al., 2009). The main aim of screening tests is to divide students into groups, not provide information about students that can inform teaching. This aim influences the assessment design and screening tests are usually designed to provide information mainly around the cut-off score to avoid erroneous classification of individual students. As such, it is challenging to use screening tests formatively.

The focus on AfL could be disrupted if teachers and schools perceive national-level assessment initiatives, such as mapping tests, as high-stakes tests. Internationally, researchers have raised a concern that, when test content is known

to teachers, it could lead to increases in test scores rather than increased student achievement (e.g. Harlen, 2007). Moreover, increases in scores could represent test inflation due to teachers practicing the test content with their students (e.g. Stobart, 2008). Prior research has repeatedly found that teachers who administer what they perceive as high-stakes tests focus on the content of the tests, administer repeated practice tests, train students how to respond to specific types of questions and adopt transmission styles of teaching (Stobart, 2008). Such behaviours stand in the way of using assessment outcomes formatively to support the learning process (Brookhart, 2011; Burkhardt & Schoenfeld, 2018; Popham, 2009; Reay & Wiliam, 1999). Therefore, teachers' assessment literacy is fundamentally important for their understanding of the purpose of the assessment and their ability to use the assessment outcomes formatively (Popham, 2009).

### ***9.2.3 Teachers' Assessment Literacy***

Teacher assessment literacy can be defined as their understanding of the principles of sound assessment (Popham, 2009; Stiggins, 2005). This includes knowledge about tests, interpretation of test results, and most importantly, understanding how to apply these results to improve student learning. These elements are key aspects of assessment literacy because adjusting instruction and knowing what to teach next are critical components of AfL from an equity perspective (Heritage, Kim, Vendlinski, & Herman, 2009). According to Brookhart (2011), teachers need to be able to analyse tests to determine what knowledge and thinking skills are required for students to solve the test items. Such analytical skills can assist teachers in using assessment results to plan their future instruction and adapt it to all students. As part of this, teachers should be able to administer external assessments and interpret their results to form decisions regarding students, classrooms and schools (Brookhart, 2011; Campbell & Collins, 2007).

A positive attitude towards the use of assessment data to assist any student lagging behind is an important aspect of teachers' assessment literacy. Importantly, teachers need to be able to cooperate with school leaders and teaching colleagues in interpreting and using assessment data to the best advantage of their students. This is an important contribution to equity because it fulfils the fundamental principle of adapted education that is a core value in the Nordic educational systems (Telhaug et al., 2006). Assessment literacy is closely related to understanding diversity and adaption of instruction. For instance, research has shown that teachers often believe classroom tests provide more cognitive diagnostic information than national-level tests do regarding students' learning processes, consequences for meaningful learning and use of learning strategies (Leighton, Gokiert, Cor, & Heffernan, 2010). Such beliefs could indicate a gap in the teacher's assessment literacy that might influence the extent to which the teacher will be able to use assessment data from external tests to enhance student learning.

### 9.2.4 *National-Level Assessments from an Equity Perspective*

The Nordic model emphasises education for all, and early intervention and AfL are implemented through national policies to ensure equitable education (Imsen et al., 2017; Telhaug et al., 2006). Within the Nordic countries, children have the right not only to go to their neighbourhood school but also to receive education that will help them fulfil their potential (Buchholtz et al., 2020). This is implemented in the Norwegian Educational Act, for instance, by means of the principles of inclusion, AfL and adapted teaching for all students (Forskrift til opplæringslova, 2006). Policy-level initiatives to steer and strengthen learning in schools through national-level efforts focus mainly on the curriculum; however, they also consider assessment practices. In an international context, research has shown that national-level efforts often prioritise the use of summative assessment for accountability and monitoring purposes, rather than formative-oriented assessment formats (Stobart, 2008). This is somewhat different in Norway, where only formative assessment is implemented in primary education (Forskrift til opplæringslova, 2006).

The NQAS differs from many other systems in that it includes national-level assessments to be used formatively (Andreasen & Hjørne, 2014; Blömeke & Olsen, 2018; Elstad, Nortvedt, & Turmo, 2009). Regarding primary school, Sweden has national tests in mathematics in grade 3 and Denmark in grades 2–6. In both countries, teachers should use the national tests to determine the extent to which students have reached curriculum goals (Skolverket, n.d.; Børne- og undervisningsministeriet, n.d.).<sup>2</sup> According to Andreasen and Hjørne (2014), these assessments function primarily as external summative assessments in contrast to the formatively oriented Norwegian mapping tests. However, both Denmark and Sweden have national policies highlighting that teachers should use test outcomes as part of their on-going assessment of their students. In this respect, the Swedish and Danish primary schools could function formatively.

For a national-level effort to contribute to equity, it should be used to adapt teaching and assessment to the needs of individual students. According to Nordenbo et al. (2009), it is crucial for teachers to find that they can use the national-level assessments outcomes in their work and feel ownership over the assessment data, as well as to perceive that they can influence matters regarding implementation of the assessment; these factors all influence teachers' intentions to use the assessments.

In our opinion, it is not sufficient that assessments are formatively oriented; test outcomes also need to be used formatively to improve instruction. If schools and teachers simply use the test score for comparison, this will lead to a mere summative use of the test, which will stand in the way of School for All (Andreasen

---

<sup>2</sup>Finland do not have external assessments at the primary level aimed at individual students at the primary level (<https://www.infonline.fi/sv/livet-i-finland/utbildning/det-finlandska-utbildnings-systemet>).

& Hjørne, 2014). This may indicate that the formative use of the test is necessary for the assessment to contribute to equity.

### ***9.2.5 The Norwegian Context***

In 2006, the Norwegian Ministry of Education and Research released the white paper titled ‘Early Intervention for Lifelong Learning’, presenting a national policy for how the education system may contribute to social equalisation (Kunnskapsdepartementet, 2006). This white paper refers to the Organisation for Economic Co-ordination and Development (OECD) evaluation of assessment practices in Norway, which pointed to Norwegian schools having weak strategies for following up students lagging behind due to a lack of information on student progression. Unclear descriptions of expected learning outcomes and a lack of mapping tools for identifying students in need of extra teaching were also highlighted. National-level research also demonstrated that Norwegian teachers tended to ‘wait and see’ when students demonstrated difficulties (Nordahl & Hausstätter, 2009; Solli, 2005).

Following advice given in the policy, the first primary school mapping tests were introduced in 2008. The second generation of mapping tests was implemented in 2014 and is still in use (Utdanningsdirektoratet, 2018). The mapping tests were intended as a tool that could support teachers and schools in identifying students at risk at an early stage and help teachers adapt their teaching to these students’ needs (Nortvedt, 2018). In other words, although the tests are taken by all students, they mainly provide information about the identified students.

## **9.3 The Present Study**

The aim of this paper is to investigate how national-level assessments might contribute to equity in schools using the Norwegian mapping tests introduced in 2014 as our case. As such, we aim to answer three research questions (RQs):

- RQ1: What happens to the test quality when an assessment is exposed over time in terms of its psychometric properties?
- RQ2: To what extent are students identified as at risk at one grade level still at risk at the next? In other words, to what extent does the assessment contribute to improved learning for the students identified as at risk?
- RQ3: To what extent do teachers understand and use the outcomes of the mapping tests to improve learning for students identified as part of their assessment literacy?

## 9.4 Method

To answer the three research questions, this chapter draws on quantitative and qualitative data related to different aspects of the implementation and use of the mapping tests in Norwegian mathematics classrooms as follows: quantitative data at the student level from the mapping test implementations in 2014–2019 and qualitative data at the teacher level from semi-structured interviews conducted in 2016. By combining the strengths of both quantitative and qualitative aspects of data analysis and large datasets, we aim to provide complementary and deeper knowledge that can contribute to educational research on equity as it is understood in a Nordic context.

### 9.4.1 Design

Addressing RQ1, student-level data from the mapping test implementation in 2014–2017 were used to investigate the test quality of each of the three mapping tests. The main aim was to investigate whether the assessments retain their psychometric properties over the period of 4 years. Data from the test implementation in 2015–2017 were linked to data from the first implementation in 2014, applying a concurrent calibration using Xcalibre to investigate whether students of a given ability level had the same probability of getting a certain total score on a test across implementations.

To address RQ2, we drew on data from 11 schools that were invited to participate in a three-year project, providing item-level data for their students for each year (2018–2020). Data from 2018 and 2019 were used to investigate what happened over time with students who were identified as at-risk students in grades 1 or 2.

Finally, addressing RQ3, data on the teacher level ( $N = 7$ ) from semi-structured interviews were used to investigate how teachers conceive, implement and follow up on the mapping tests. Teachers' engagement with the mapping tests provides insights into how the mapping tests are used and the extent to which they might contribute to enhancing equity.

### 9.4.2 Samples and Recruitment

Sample 1 comprises data on the item and student levels (grades 1–3) for each mapping test implementation from 2014 to 2017. A new sample was selected each year, meaning that sample 1 is suitable for investigating test quality (Table 9.1).

Sample 2 is a convenience sample consisting of item-level data from grade 1–3 students in 11 schools. The total sample is presented in Table 9.2. It should be noted



**Table 9.1** Sample for each test implementation for grades 1, 2 and 3 in 2014–2017

	2014	2015	2016	2017	Total
Grade 1	2281	989	969	906	5145
Grade 2	2483	1102	1073	886	5544
Grade 3	2286	967	947	888	5088

**Table 9.2** Sample for each test implementation for grades 1, 2 and 3 in 2018–2019

	2018	2019	Total
Grade 1	425	268	693
Grade 2	354	363	717
Grade 3	316	213	529

that, due to students changing schools or a lack of parental consent<sup>3</sup> to participate in the study, this sample is limited to only a part of the total sample for 2018 and 2019. This means that the combined sample participating in both grades 1 (2018) and 2 (2019) includes 259 students, while the combined sample participating in grades 2 (2018) and 3 (2019) includes 150 students. As these samples are small, both quantitative and qualitative analyses are necessary to analyse the data.

For both samples 1 and 2, the school principal was first approached and asked if the school could participate in the data collection. For sample 1, one school class at each grade level was invited to participate. For sample 2, all classes/students in grades 1–3 were invited to participate.

Sample 3 consists of seven teachers from four schools across two school districts (see Table 9.3). Six of the teachers were recruited through the school principal to participate in the study. The seventh teacher (David) was purposefully selected for the study due to his previous interest in the mapping tests and the lack of male teachers in the sample. All seven teachers provided informed consent to participate in the study.

### 9.4.3 Data Collection

To collect data on the item level for all students, the schools were asked to provide student booklets for each student. Data were coded and registered for later analysis, and one database was constructed for each assessment for each year. In addition, a combined database for each grade level comprising data from 2014–2017 was made, and two linked databases were constructed from sample 2 students who had participated in two consecutive years.

---

<sup>3</sup>Sample 2 data are collected with student identities and require parental consent. As such, the project has been reported to the Norwegian Centre for Research Data (NSD; project number 58107). Sample 1 data are collected without school and student identity and therefore do not require parental consent.

**Table 9.3** Background information for participating grade 1, 2 and 3 teachers

Teacher	School	Grade level	Experience with implementing the mapping test
Anna	A	2	Experienced
Anita	A	1	Experienced in special needs but first year as a regular teacher
Bente	B	1	Experienced
Brita	B	1	Experienced
Camilla	C	1	Administered the mapping tests for the first time in 2016
Carina	C	3	Much experience at higher levels with other assessments; first time administering mapping tests
David	D	2	Experienced

The first author of this chapter conducted semi-structured interviews with seven teachers after the mapping test implementation in 2016. Each interview took place in a secluded room in the participant's school and lasted 60 min on average. All interviews were audiotaped and later transcribed. Two grade 1 teachers working closely together (Bente and Brita) were interviewed together. All other interviews were individual. The teachers were asked how they prepared for and implemented the mapping test with their students, analysed the test outcomes and followed up the mapping test results with identified students. The tests were taken in late March or early April, and the interviews were conducted in late June.

#### 9.4.4 Data Analysis

Regarding RQ1, item response theory (IRT)-based test-equating procedures in the form of concurrent calibration were performed to investigate the extent to which item characteristics were maintained over time or whether test inflation occurred. Concurrent calibration was the preferred test-equating procedure because it allows pairwise comparison of test characteristics across two timepoints. The assumption here is that the test measures the same construct at both administrations.

As the tests were not changed between 2014 and 2019, and because the same tests were implemented at each timepoint, all test items have been treated as anchor items. Thus, the ability estimates ( $\theta$ ) from the different test administrations (at the same grade level) resulting from such calibration will be on the same scale as one another, making the scores from two tests comparable because both the  $a$  and  $b$  parameters are invariant across the population.

To investigate how the mapping tests affect students over time (RQ2), a small subsample comprised data on two timepoints for students moving from grade 1 to grade 2 and for students moving from grade 2 to grade 3. These data were used to investigate how student results typically develops across the two timepoints. In this

analysis, we primarily used descriptive statistics, such as averages, cross-tables, the analysis of variance (ANOVA) test and chi-square analysis.

Regarding RQ3, the interviews were analysed using meaning condensation following Kvale and Brinkmann (2009) to identify the teachers' conceptions of both the mapping tests and the identified students. This analysis aimed to uncover teachers' experiences and their reflections on test administration and data analysis in addition to following up on students.

In the first stage, three of the authors analysed the data separately (Kvale & Brinkmann, 2009). In the next stage, the authors alternated between working individually and collaboratively to enable meaning condensation and interpretation of the interview data.

Table 9.4 illustrates meaning condensation of natural units of teacher statements. During the interviews, teachers provided rich descriptions of their work and reflections, enabling their talk to be broken down into natural 'meaning units' that were analysed using meaning condensation. Finally, derived meanings were interpreted. All quotes used in the results section have been translated from Norwegian to English by the authors. Rather than translating them word by word, the translations focus on representing the core ideas and understandings expressed by the teachers to better align with the applied analytical process.

**Table 9.4** Illustration of meaning condensation and interpretation

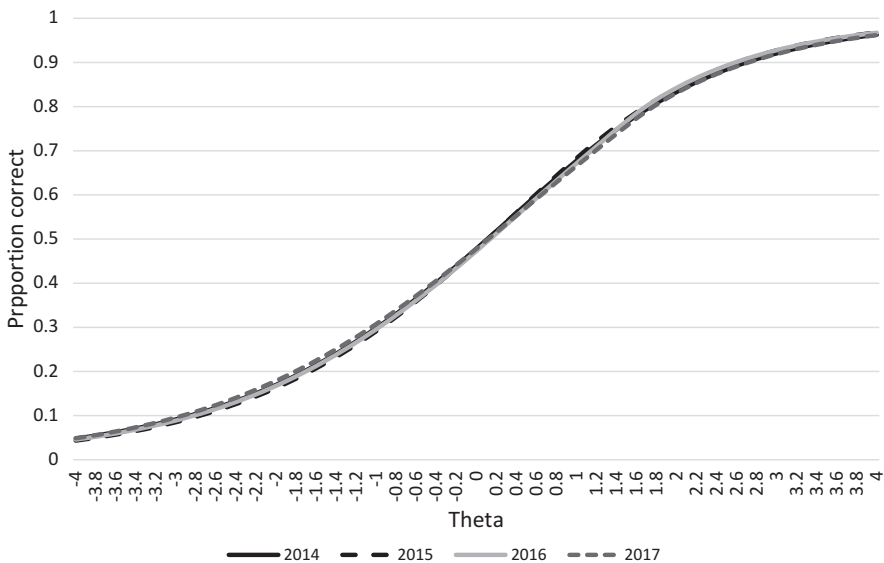
Natural unit (teacher statement)	Central themes	Interpretation
<i>Student follow-up—We have sometimes made course groups,</i>	Establish groups to organise follow-up	Making groups organise the follow-up
<i>We score, I do think that mapping is for finding out where on the map students are and what they need to practice.</i>	Mapping test will help teachers identify what students can do and need to work on	Mapping test is a tool for teachers
<i>So, we have made course groups, for instance, if we see that many share the same challenges in relation to a topic, so, we have made them across classrooms, because we have three classes, so I think the courses are OK.</i>	Following up groups of students with similar challenges across classes makes courses work well	Following up in groups is appropriate
<i>[Oppfølgingen av elevene; vi har kjørt noen ganger litt kursgrupper, vi retter, jeg tenker jo at kartlegging er for å se hvor elevene er i terrenget og hva man trenger å øve på. Så har man kjørt litt kursgruppe for eksempel hvis man ser så, så mange som har utfordringer hvis det gjelder ett emne så har man kjørt, smelt dem sammen. og det er på tvers av trinnene da (hmmh) for det er tre klasser hos oss. så synes jeg går helt kurant.]</i>		

## 9.5 Results

In this section, we present the results following the order of the research questions that guided our investigations. The insights gained from the data analysis related to the three RQs, as well as the relationship between the three outcomes, are further elaborated on in the discussion section.

### 9.5.1 What Happened to the Mapping Test Quality After Five Test Administrations?

RQ1 focussed on what happens to the quality of the tests when the assessments are exposed over time. Specifically, do the assessments retain their psychometric properties even after four test administrations? Figure 9.1 shows the test response function (TRF) for the grade 2 test for the 2014–2017 test administrations. The curves more or less overlap, revealing that a student with a certain ability level in 2015–2017 had more or less the same probability of providing the same proportion of correct responses as a student with the same ability level in 2014. This means that the expected test performance is the same across years, and the examinees show the same expected distribution of performance in the four test administrations. The cut-off score calculated in 2014 is 41 points ( $\theta = -1.366$ ). This is close to where the test has the maximum information and the measurement error is very small (0.20).



**Fig. 9.1** Test response function for the grade 2 test for the 2014–2017 test administrations

Previous research in other countries has often found test inflation in exposed assessments. Test inflation typically happens for two different reasons, which are as follows: (1) teachers practise with students so they know how to respond to the test questions in advance and (2) teachers use their familiarity with the test and the test outcomes to improve their teaching. In Fig. 9.1, this would have been the case if the TRF graphs representing 2015–2017 test administrations rose above the line representing the 2014 administration. However, as shown in this figure, no test inflation was observed for the grade 2 mapping test.

Similar outcomes were obtained for the grades 1 and 3 mapping tests. Taken together, these outcomes lead to two likely interpretations, which are as follows: (1) there is no inflation in test scores due to test robustness, and (2) schools seemingly do not succeed in utilising the assessment data to improve mathematics instruction in primary grades 1–3. While the first interpretation points to test quality, the second points to potentially low assessment literacy or interest in using the assessment data. Neither interpretation can be excluded based on the current analysis.

### 9.5.2 *What Happens over Time to Students Identified as ‘At Risk’ in Grade 1 or 2?*

Data from the linked database, comprising data from the 2018 and 2019 samples, were used to investigate our second research question on what happened to students identified as being at risk in grades 1 or 2 in 2018: Were these students still below the cut-off score in 2019?

Table 9.5 shows the outcome patterns for the students ( $N = 259$ ) who attended grade 1 in 2018 and grade 2 in 2019, while Table 9.6 shows the outcome patterns for the students ( $N = 150$ ) who attended grade 2 in 2018 and grade 3 in 2019. Table 9.5 reveals that approximately 20% of the 259 students going from grade 1 to grade 2 were below the cut-off score in grade 1, grade 2 or both years. In this sample, nearly 1 in 10 students was identified as at risk in the two consecutive school years. While 5% of the grade 1 students were no longer identified as at risk in the following year, there was also a relatively large group of students (7%) who were not identified in grade 1 but fell below the cut-off score in grade 2 and were identified as at risk.

Similar patterns were observed for the transitions from grade 2 to grade 3 (Table 9.6). Nearly 20% of the students were identified as at risk in one or both school years. Fewer students (5%) were below the cut-off in both years. The same

**Table 9.5** Achievement levels of students in grade 1 (2018) and grade 2 (2019)

		Grade 2, 2019		
		Identified as at risk	Not identified as at risk	Total
Grade 1, 2018	Identified as at risk	9%	5%	14%
	Not identified as at risk	7%	79%	86%
Total		16%	84%	100%

**Table 9.6** Achievement levels of students in grade 2 (2018) and grade 3 (2019)

		Grade 3, 2019		
		Identified at risk	Not identified as at risk	Total
Grade 2, 2018	Identified as at risk	5%	6%	11%
	Not identified as at risk	10%	79%	89%
Total		16%	84%	100%

**Table 9.7** Average scores in grades 1 and 2 for groups of students identified as at risk in both years, increasing from at-risk status, falling to at-risk status or scoring above the cut-off score in both years

	%	Average score in grade 1 (SD)	Average score in grade 2 (SD)
At risk both years	9%	32.6 (4.562)	28.5 (8.523)
Increasing from at risk	5%	36.1 (2.397)	46.1 (2.290)
Falling to at risk	7%	43.1 (2.198)	37.4 (3.789)
Above both years	79%	47 (2.648)	50.9 (3.231)

Note: Maximum score (cut-off) for grade 1 is 50 (39) and for grade 2 is 55 (41)

number of students moved from below to above the cut-off score. In this sample, 1 in 10 students was not identified as at risk in grade 2 but was identified as at risk in grade 3.

The outcomes may indicate that some teachers succeed in using the test results to improve student learning for their students. At the same time, they also show that some students identified as at risk in 2018 were still identified as at risk in 2019. This may indicate that the second school year did not provide students with sufficient opportunities for learning numeracy.

Table 9.7 shows the average scores for the groups of students that scored below the cut-off score in grade 1, grade 2 or both years. The tests have a ceiling effect, which affects the average scores for the group scoring above the cut-off score in both years. However, for the other three groups, average scores can be calculated, and an ANOVA test demonstrates significant differences between the four groups in both years [ $F(3,255) = 221.615, p < .001$  and  $F(3,255) = 264.286, p < .001$ ], with one exception: The students who have improved their results from below to above the cut-off score, in grade 3, does not score significantly lower than the group of student who scored about both years.

The group average scores presented in Table 9.7 indicate that the students who were identified as at risk in both years scored significantly below the cut-off in grade 1, with an average of 32.6 points (cut-off 39 points), but they scored even further from the cut-off score in grade 2, when the average scores was 28.5 points (cut-off 41 points). This indicates that the teachers did not succeed in increasing the at-risk students' attainment, and the increased standard deviation supports this interpretation. The students who transitioned from below to above the cut-off score, on

**Table 9.8** Average scores in grades 2 and 3 for groups of students identified as at risk in both years, increasing from at-risk status, falling to at-risk status or scoring above the cut-off score in both years

	%	Average score in grade 2 (SD)	Average score in grade 3 (SD)
At risk both years	6%	32.1 (8.880)	46 (12.052)
Increasing from at risk	5%	40 (1.414)	63.88 (3.137)
Falling to at risk	10%	47.7 (3.826)	53.27 (4.605)
Above both years	79%	51.9 (2.6435)	68.4 (3.123)

Note: Maximum score (cut-off) for grade 2 is 55 (41) and for grade 3 is 72 (59)

average, were closer to the cut-off values, and at the same time, they scored well above the cut-off score in grade 2. In addition, the standard deviation was smaller for the second year, indicating that the students were more similar regarding achievement levels in 2019 compared to 2018. There was also a group of students who, on average, scored well above the cut-off in grade 1 but below the cut-off in grade 2. Judging by the increased standard deviation, more variation is visible in student achievement in grade 2 for the latter group.

Table 9.8 shows the group average scores for students in grades 2 and 3, showing similar patterns to those revealed for the grade 1–grade 2 transition. Table 9.8 indicates that, at this level, the students who were identified as at risk in both years also scored significantly below the cut-off in both years. This indicates that the teachers did not succeed in increasing the at-risk students’ attainment, and again, the larger standard deviation supports this interpretation. Judging by the larger standard deviation for the students identified as at risk both years, more variation is visible in student achievement for this group, something that may make it more challenging for teachers to interpret the test outcomes and response patterns of these students.

### 9.5.3 *To What Extent Does the Mapping Test Function as a Tool for Teachers to Support Student Learning?*

Teachers’ assessment literacy is a determinant of their work with mapping tests. For this reason, the teachers were asked about how they prepare for, administer and follow up the mapping test with their students. During the interviews, the teachers also shared their views and experiences about the mapping test and their work with the students identified as being at risk.

In the responses, four of the interviewed teachers express that in their view the mapping test could work as a tool for teachers and help them identify topics to address with their students. David’s metaphor about placing students on a map is related to AFL:

... Mapping students is done to see where on the map the students are and what we need to practice more. [... kartlegging er for å se hvor elevene er i terrenget og hva man trenger å øve på.]

At the same time, our analysis revealed that the purpose of the mapping test may be somewhat unclear to some of the interviewed teachers. Anna's reflection below illustrates this and shows that, although she also highlights the formative aspect of the assessment, she is uncertain whether this is an external assessment or a tool for teachers. This exemplifies how teachers might struggle with understanding what distinguishes one test from another:

Anna: But I do not really know what the purpose [of the assessment] is. Is it like a national test where you should give feedback immediately? Or is it more like a tool for us teachers, you know? [Men jeg vet egentlig ikke helt hva som er målet. Er det som en nasjonal prøve som man skal gi tilbakemelding med en gang? Eller er det et verktøy for oss lærere ikke sant?]

Although emphasising the formative aspect, David also indicates that the mapping test provided insight into his teaching, suggesting that teachers may see alternative uses for mapping test data.

Bente and Brita, the two grade 1 teachers, report that they administered the test according to set guidelines, and they devote considerable time to analysing the test results. Even so, they express scepticism towards the test, partly because they believe the students are too young, and there is a risk of the testing being an uncomfortable experience for some students. They clearly express that conducting the mapping test is something they are obligated to do, and they are somehow sceptical of the test results. However, they view the assessment as a tool they can use to improve their instruction.

To prepare for the test, teachers need to go through tutorial materials that include instructions for how to administer the test. All the interviewed teachers state that it is important to create standardised conditions for all students in the testing. Internal school guidelines, in addition to the national guidelines, help the teachers create equal conditions when adapting the test situation to individual students as well. At the same time, however, the teachers sometimes feel the guidelines contribute to inequity. The test is timed so that students with naïve or rigid strategies will not have time to finish calculation tasks using these strategies. In particular, the time restraints are viewed as frustrating by the teachers, who find them unfair for low-achieving students:

Anita: We got a little frustrated with the time restriction because some first-graders would have done much better if the test wasn't timed. Because then I think everyone could have shown what they knew, not how much they could accomplish in a certain amount of time. [Vi ble litt frustrerte av det med tida fordi noen førsteklassinger hadde gjort det bra hvis det ikke var på tid. Fordi da tenker jeg da at alle hadde fått vist hva de kunne, ikke hvor mye de kunne prestere på et visst tidsrom.]

As Anita's statement exemplifies, the teachers feel their students would be able to show more of their competence if they had more time to respond to the test items. Thus, the teachers sometimes express that the test results do not reflect the



perceived level of competence of their students. The teachers also mention other factors, for example, the scoring procedures or unfamiliar item formats, which they feel affect student results.

The interviewed teachers show an awareness of other factors influencing test outcomes related to the student or the student's background, including learning difficulties, difficult situations at home, lower attention levels, misconceptions, linguistic challenges or careless mistakes. This is expressed by David in the following quotation, in which he points to factors outside school that influenced the test taking of two of his students, and consequently, wrongly identified them as at risk:

David: I think a lot of things in her life aren't so easy for her in general (...) and if then in a way her life outside of school has taken hold at an unlucky time, it might explain, right (...) So two of the students that are at risk, it is not mathematics interventions but other interventions that are needed. [Her for den ene sin del så tenker jeg at hun ikke har det så lett generelt {...} Og hvis da på en måte livet hennes utenfor skolen har gjort seg gjeldende på et uheldig tidspunkt så kan det forklare, ikke sant {...} Så to av de elevene de har under bekymringsgrensa så er det ikke matematikkfaglig tiltak, men andre tiltak som er nødvendig.]

All seven teachers indicate that they spend considerable time preparing for, administering and attempting to understand the outcomes of the mapping test. None of the interviewed teachers report any difficulties scoring the tests, but analysing the data is challenging for many of them. Judging by his statement above, David connects difficulties with analysing data to a lack of classroom-level teaching initiatives.

Teachers may struggle to interpret the test results if they do not trust them. Moreover, analysis of the teacher interviews indicated that the teachers prioritise identifying student errors, misconceptions and mistakes the students might make if they misunderstand the task instructions. This could explain why it is difficult to plan interventions, as AfL builds on what students know and can do. Still, some interviewed teachers show awareness of their instruction and how this might influence student learning, as well as how it might influence the mapping test results and response patterns.

Overall, the seven teachers list many kinds of teaching innovations aimed at individual students or groups of students, in small-group or classroom teaching, including the following: engaging in learning conversations with students, setting up learning goals for individual students, using extra time when available with identified students, using more manipulatives and concrete materials when teaching, station teaching and grouping identified students with similar difficulties to work on specific topics. Moreover, differentiating task or activities during whole-class instruction, introducing peer assessment and learning partners, making courses for groups of students and focussing on mathematical concepts are also highlighted. However, most of the teachers state that they lack time to follow up on the students after the test, and thus, their main efforts have to wait until after the summer holiday. To facilitate more teaching interventions, they need time to plan (independently and in cooperation with colleagues) to identify necessary resources (time and teaching materials) and how teaching students in cooperation with colleagues could target

identified student needs. Moreover, they indicate that, in this process, they need help from the leaders in their school.

## 9.6 Discussion

Our analysis revealed that the mapping tests are robust; the item and test characteristics have not changed significantly over time (RQ1). Some students improved their results over time, while some did not, and some students even showed a decline in their understanding of numbers and calculation skills (RQ2). Moreover, although the teachers took care to administer the test following national and school guidelines, they struggled to interpret the test outcomes, and although a wide variety of interventions were listed, they were sometimes delayed until the fall (RQ3).

To frame our discussion, we draw on prior research on how assessment initiatives can be used to enhance equity in schools. In addition, prior research on equity (Espinoza, 2007; Zhu, 2018), AfL (Heritage et al., 2009; Wiliam, 2007), teachers' assessment literacy (Brookhart, 2011; Popham, 2009) and what teachers need to be able to do to use assessment data to improve students' opportunities to learn is used to discuss possible lessons learned from the Norwegian mapping test implementation.

### 9.6.1 *National-Level Initiatives Such as the Mapping Tests May Contribute to Equity in Schools*

For national-level assessments, such as mapping tests to contribute to equity, they need to be robust and identify students at risk of lagging behind (Brookhart, 2011; Stobart, 2008). In addition, teachers need to be able to administer the test in the same way and use the test outcomes to improve their teaching (Stobart, 2008). The IRT analysis demonstrated that the Norwegian mapping tests are robust, and judging by the interviews, the teachers managed to implement the assessment according to the national guidelines. As such, mapping tests may contribute to equity.

The analysis of the test data indicates that Norwegian teachers likely do not 'teach to the test' because the mapping tests functioned in the same way after several years of exposure. An alternative interpretation is that what the teachers practice with the students did not influence students' ability to respond to the test items. This outcome is contrary to the test score inflation that has been observed in other countries (e.g. Stobart, 2008), and it may be related to the school's ownership of test outcomes. We argue that, in situations with low-stakes national assessments, no test inflation and locally owned data, external assessments may contribute to equity because the teachers can feel more ownership to the data and influence over the use

of the assessment. Taken together this may provide more reliable measures for the identified students.

A third explanation, and a slightly less positive one, is that Norwegian teachers have not improved their instructional practices sufficiently, and over time, they have not offered better opportunities for learning for students identified as being at risk by the mapping tests. The analysis of what happens to identified students over time supports this interpretation: Some identified students (8% in total) were still at risk in the following school year. However, at the same time approximately one in two students identified as at risk in 2018 (or 7% of the total sample) scored above the cut-off score the following year. As such, we take these outcomes to mean that mapping tests can contribute to equity. Still, to improving equity classroom instruction needs to offer identified students possibilities to develop better conceptual understanding and calculation skills related to the key aspects of the mapping tests. The analysis of the interview data supports this interpretation because follow-ups were often delayed.

Previous research indicates that teachers often lack necessary assessment literacy to follow up on assessment outcomes (Heritage et al., 2009; Leighton et al., 2010). Some statements from the interviews may indicate that this is the case for some—but not all—of the interviewed teachers. As such, understanding how teachers' conceptions and beliefs about the mapping test interact with AfL initiatives is crucial.

### ***9.6.2 Teachers' Assessment Literacy and Assessment for Learning Practices Conditions How Mapping Tests Might Contribute to Equity***

According to Gersten et al. (2009) and Brookhart (2011), mapping tests need to be followed up with targeted instructions to improve learning. The tests are administered in the spring, and most of the interviewed teachers stated they experienced a lack of time to follow up with the students in the spring semester. Instead, they planned to do so after the summer break. Perhaps this notion of the mapping tests as end-of-year tests causes the teachers to view them as summative rather than as part of the on-going formative assessment they conduct during the school year. Stiggins (2005) argues that assessment that takes place during the learning process can contribute to the formative use of tests, and thus, promote student learning. The teachers' statements about following up during the autumn semester support this summative conception of the tests' purpose. In addition, teacher statements about already knowing who struggles prior to the mapping test supports the interpretation of viewing the tests as summative. We argue that teachers need to view and use the tests as formative for them to contribute to equity (e.g. Heritage et al., 2009). Still, the seven teachers had already implemented some teaching interventions in the late spring and early summer, and such activities as peer assessment, setting learning

goals and involving students in learning conversations can be viewed as AfL activities. We argue that whether teachers view the mapping tests as summative or formative depends on their assessment literacy.

The mapping test data are locally owned. The intention with the mapping tests is that schools and teachers will feel ownership, and the primary goal is teaching interventions rather than reporting. As such, the tests can function as a support tool and not an accountability measure. However, the interviews showed that we might question whether every interviewed teacher view the tests as a tool for improving teaching and learning.

The analysis of the interview data revealed that the teachers held different perceptions about classroom and external tests. According to Brookhart (2011), this could influence their assessment literacy. Leighton et al. (2010) noted that many studies have shown that teachers have somewhat negative attitudes toward large-scale national assessment. Our study may support this finding, as some interviewed teachers saw the mapping tests as an external assessment that evaluates students rather than a tool they could use to improve teaching and learning. At the same time, we found that the interviewed teachers sometimes did not trust the test results of identified students, and it may be inferred that they believed that students' test performance reflected test-taking strategies rather than numeracy skills.

Using assessment outputs to inform teaching is fundamental to formative assessment (Brookhart, 2011). The support material that accompanies the mapping tests supposedly helps teachers do this. It provides information about how the test works, how to interpret the results, what it means when students are identified as at risk and suggestions for further instruction. However, based on the interview data, it is questionable whether all teachers are actually being provided with adequate support material.

## 9.7 Concluding Remarks—Linking Equity, National-Level Initiatives and Assessment Literacy

The overall question in this chapter related to whether a national-level initiative—in this case, the Norwegian mapping tests—can improve equity in schools. That there is no inflation in test scores supports using the same tests over time and trusting teachers to use them as intended. Further, a large proportion of the students who were below the cut-off score one year were above it the next. This could be due to factors not included in this research, but it may also be an outcome of using the information from the mapping test, and thus, contributing to equity. Overall, these observations support the idea that mapping tests can improve equity.

In accordance with Nordic model principles for transparency (Telhaug et al., 2006) and school autonomy, the mapping test content is available to teachers and schools, and test results are locally owned. Moreover, schools are trusted to use the mapping test outcomes in accordance with national guidelines. As a result,

familiarity with the test content helps ensure formative use of assessment outcomes to improve teaching and learning. For instance, Norwegian schools are responsible for identifying professional development needs for their teachers (Imsen et al., 2017). Imsen et al. (2017) discuss a dilemma in that schools simultaneously have to deal with national-level assessment and regulations while having autonomy to interpret the curriculum and plan instruction. Our research can be seen as confirming this dilemma between national-level and local initiatives and responsibilities.

Going forward would mean developing national-level initiatives that allow for local adaption that can assist schools in further developing teachers' assessment literacy. In addition, future endeavours should provide educators with the means to develop more knowledge and strategies for targeting their teaching to students at risk, thereby enhancing equity by being better prepared to adapt teaching. We propose a three-part strategy in line with the traditions and values in the Nordic model to ensure that national-level assessments contribute to equity in primary school as follows: (1) offering high-quality assessments, (2) offering helpful and useful tutorials and support material and (3) implementing national and local initiatives that can assist teachers in further developing their assessment literacy. Norway has implemented the first two of these. However, to take full advantage of these two parts of the strategy, we argue that the third is necessary because this will help all schools and teachers improve their assessment literacy, leading to more equitable mathematics education. At the same time, to be aligned with the Nordic model principles, transparency and school autonomy must be maintained.

Each of the elements identified above seems like sound advice, but we argue that it is only when they come together that we will see development. First, quality assessments are more than mere psychometric sound assessments. They are accompanied by documents that provide teachers and school leaders with insights into how the assessments are developed, what they measure and how they should be implemented. Second, the tutorial and support material should assist teachers and schools in analysing assessment data to understand what students know and can do. Moreover, it should also help teachers to translate this knowledge into an understanding of what students should learn next and how to achieve this. Only when this is in place will the assessment operate as AfL and contribute to equity (e.g. Heritage et al., 2009). Finally, to assist teachers and schools in using the mapping tests and tutorial materials and to ensure that this initiative fosters teachers' assessment literacy, we need to offer local and national support focussed on teachers' conceptions of students and assessment.

Teachers' positive attitudes toward mapping tests are instrumental to using the assessment outcomes to improve equity in school. Based on this argument, if teachers do not believe that the mapping tests are a helpful tool for improving their instruction, and if they do not have the necessary assessment literacy, the tests are not likely to contribute to improved teaching and learning opportunities for identified students or to equity. However, at the same time, we emphasise that, as researchers, we have a primary responsibility to conduct research that can inform all three aspects of the above-mentioned strategy to promote equity in primary school mathematics instruction.

## References

- Andreasen, K., & Hjørne, E. (2014). Assessing children in the Nordic countries: Framing diversity and matters of inclusion and exclusion in a School for All. In U. Blossing, L. Moos, & G. Insen (Eds.), *The Nordic education model. 'A School for All' encounters neo-liberal policy* (pp. 155–172). Dordrecht, The Netherlands: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7–74. <https://doi.org/10.1080/0969595980050102>
- Blömeke, S., & Olsen, R. V. (2018). På vei mot et sammenhengende nasjonalt kvalitetsvurderings-system. *Acta Didactica Norge*, 12(4), 1–21. <https://doi.org/10.5617/adno.6278>
- Børne- og undervisningsministeriet [Ministry of children and education]. (n.d.). *Om de nationale test* [About the national tests]. <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test%2D%2Dtrivselmaaling-og-sprogproever/nationale-test/om-de-nationale-test>. Accessed 18 Aug 2020.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Buchholtz, N., Stuart, A., & Frønes, T. S. (2020). Equity, equality and diversity—Putting educational justice in the Nordic model to a test. In T. S. Frønes, A. Pettersen, J. Radišić, N. Buchholtz (Eds.), *Equity, equality and diversity in the Nordic model of education* (pp. 13–41). Cham, Switzerland: Springer.
- Burkhardt, H., & Schoenfeld, A. (2018). Assessment in the service of learning: Challenges and opportunities or Plus ça Change, Plus c'est la même Chose. *ZDM*, 50(4), 571–585. <https://doi.org/10.1007/s11858-018-0937-1>
- Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education. *Educational Measurement: Issues and Practice*, 26(1), 9–18.
- Elstad, E., Nortvedt, G. A., & Turmo, A. (2009). The Norwegian assessment system: An accountability perspective. *Cadmo*, 17(2), 89–103.
- Espinoza, O. (2007). Solving the equity–equality conceptual dilemma: A new model for analysis of the educational process. *Educational Research*, 49(4), 343–363. <https://doi.org/10.1080/00131880701717198>
- Forskrift til opplæringslova [Regulation to the Education Act]. (2006). *Chapter II Undervisningsvurdering [Day-to-day assessment]*, § 3–1.
- Gersten, R., Beckmann, S., Clarke, B., Foeggen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools (NCEE 2009–4060)*. <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>
- Ginsburg, H. P. (2016). Helping early childhood educators to understand and assess young children's mathematical minds. *ZDM*, 48(7), 941–946. <https://doi.org/10.1007/s11858-016-0807-7>
- Gutstein, E., Fey, J. T., Heid, M. K., DeLoach-Johnson, I., Middleton, J. A., Larson, M., et al. (2005). Equity in school mathematics education: How can research contribute? *Journal for Research in Mathematics Education*, 36(2), 92–100.
- Harlen, W. (2007). *Assessment for learning*. Los Angeles: Sage.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24–31.
- Heritage, M., & Wylie, C. (2018). Reaping the benefits of assessment for learning: Achievement, identity, and equity. *ZDM*, 50(4), 729–741. <https://doi.org/10.1007/s11858-018-0943-3>
- Hopfenbeck, T. H., Lenkeit, J., El Mastri, Y., Cantrell, K., Ryan, J., & Baird, J.-A. (2017). Lessons learned from PISA: A systemic review of peer-reviewed articles on the programme for international student achievement. *Scandinavian Journal of Educational Research*, 62(3), 333–353.
- Insen, G., Blossing, U., & Moos, L. (2017). Reshaping the Nordic education model in an era of efficiency. Changes in the comprehensive school project in Denmark, Norway, and Sweden

- since the millennium. *Scandinavian Journal of Educational Research*, 61(5), 568–583. <https://doi.org/10.1080/00313831.2016.1172502>
- InfoFinland. (n.d.). *Det finländska utbildningssystemet* [The Finnish educational system]. <https://www.infofinland.fi/sv/livet-i-finland/utbildning/det-finlandska-utbildningssystemet>. Accessed 18 Aug 2020.
- Kunnskapsdepartementet [Ministry of Education and Research]. (2006). *... og ingen stod igjen. Tidlig innsats for livslang læring (St. meld. Nr. 16(2006–2007) [... And no one was left behind. Early intervention for lifelong learning (White paper no. 16(2006–2007))]*. <https://www.regjeringen.no/>
- Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Los Angeles: Sage.
- Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy & Practice*, 17(1), 7–21. <https://doi.org/10.1080/09695940903565362>
- Nordahl, T., & Hausstätter, R. S. (2009). *Spesialundervisningens forutsetninger, innsats og resultater* [Conditions, efforts and outcomes of special education]. Høgskolen i Lillehammer.
- Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Sjøgaard Larsen, M., et al. (2009). Pædagogisk bruk av test – et systematisk review [Using tests in education – a systematic review]. In *Evidensbasen*. Copenhagen, Denmark: Danish Clearinghouse for Uddannelsesforskning, DPU, Aarhus University.
- Nortvedt, G. A. (2018). «Det er et verktøy, ikke sant, for oss» – Erfaringer fra fire gjennomføringer med kartleggingsprøver i regning 2014–2017. *Acta Didactica Norge – tidsskrift for fagdidaktisk forsknings- og utviklingsarbeid i Norge*, 12(4), 1–22.
- Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, 50(4), 555–570. <https://doi.org/10.1007/s11858-018-0963-z>
- Nortvedt, G. A., Santos, L., & Pinto, J. (2016). Assessment for learning in Norway and Portugal: The case of primary school mathematics teaching. *Assessment in Education: Principles, Policy & Practice*, 23(3), 377–395. <https://doi.org/10.1080/0969594X.2015.1108900>
- Nortvedt, G. A., & Wiese, E. (2020). Numeracy and migrant students: A case study of secondary level mathematics education in Norway. *ZDM: Mathematics Education*, 52(3), 527–539. <https://doi.org/10.1007/s11858-020-01143-z>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Räsänen, P., Daland, E., Dalvang, T., Engström, A., Korhonen, J., Kristinsdóttir, J. V., et al. (2019). Mathematical learning and its difficulties: The case of Nordic countries. In A. Fritz, V. G. Haase, & P. Räsänen (Eds.), *International handbook of mathematical learning difficulties: From the laboratory to the classroom* (pp. 107–125). Cham, Switzerland: Springer.
- Reay, D., & Wiliam, D. (1999). ‘I’ll be a nothing’: Structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343–354.
- Roos, H. (2018). Inclusion in mathematics education: An ideology, a way of teaching, or both? *Educational Studies in Mathematics*, 2019(100), 25–41. <https://doi.org/10.1007/s10649-018-9854-z>
- Rousseau, C., & Tate, W. F. (2003). No time like the present: Reflecting on equity in school mathematics. *Theory Into Practice*, 42(3), 210–216.
- Samoff, J. (1996). Which priorities and strategies for education? *International Journal of Educational Development*, 16(3), 249–271.
- Skolverket [Swedish National Agency for Education]. (n.d.). *Nationella prov i grundskola* [National tests in compulsory education]. <https://www.skolverket.se/undervisning/grundskolan/nationella-prov-i-grundskolan>. Accessed 18 Aug 2020.
- Solli, K. A. (2005). *Kunnskapsstatus om spesialundervisning i Norge* [Special education in Norway, status report]. Utdanningsdirektoratet.

- Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324–328.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.
- Telhaug, A. O., Mediås, O. A., & Aasen, P. (2006). The Nordic model in education: Education as part of the political system in the last 50 years. *Scandinavian Journal for Educational Research*, 50(3), 245–283.
- Utdanningsdirektoratet. (2018). *Hva er kartleggingsprøver* [What are mapping tests]? <https://www.udir.no/eksamen-og-prover/prover/hva-er-kartleggingsprover/>
- Wiliam, D. (2007). Keeping learning on track. In F. K. J. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053–1098). Charlotte, NC: Information Age.
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–15.
- Zhu, Y. (2018). Equity in mathematics education: What did TIMSS and PISA tell us in the last two decades? In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited lectures from the 13th International Congress on Mathematical Education. ICME-13 Monographs* (pp. 769–786). Cham, Switzerland: Springer. [https://doi.org/10.1007/978-3-319-72170-5\\_43](https://doi.org/10.1007/978-3-319-72170-5_43)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

