

En litteraturgjennomgang på det norske prøvefeltet:

Bruk av prøver for å forbedre praksis, prøver som styringsverktøy og prøver i politikkutforming (2005-2018)

Gunstein Egeberg og Ove E. Hatlevik

Oslo: februar 2020.

Oppsummering

I løpet av 15 år har det vi kjenner som «prøvefeltet» utviklet seg fra et uferdig og lite robust område til et mer modent felt der vi i dag kjenner bedre til mange av forholdene som prøvene berører. Denne rapporten er et forsøkt på å identifisere og kategorisere publikasjoner som oppfyller ett av følgende kriterier: Publikasjoner om prøvene som styringsverktøy, publikasjoner om bruk av prøvene for å forbedre praksis, og prøvenes funksjon i politikkutforming og til andre policyformål.

Etter gjennomgang av et stort antall abstrakt fant vi at det var 81 tilgjengelige publikasjoner som oppfylte kriterier for inkludering. En gjennomgang av disse 81 publikasjonene viser at forskningsinteressen har vært nokså stabil siden starten, men med variasjoner mellom år. Det er nasjonale prøver som får mest oppmerksomhet med prøvene i lesing som de best undersøkte. Regning er også inkludert i en del arbeider. Andre prøvetyper og fag er mindre berørt. Selv om rammeverkene for prøvene og selve prøve kvaliteten er bedre nå, gjenstår det en del arbeid for å hjelpe lærere og skoler til å nyttiggjøre seg informasjonen som de ulike prøvene gir.

Nøkkelord

Utdanningsdirektoratet; Nasjonale prøver; Kartleggingsprøver; Læringsstøttende prøver; Forbedre praksis; Prøver som styringsverktøy; Policy; Fag; Måling;

Innholdsfortegnelse

Oppsummering	1
Innledning	5
Prøveutvikling i regi av Utdanningsdirektoratet: 2002 - 2019	5
Prinsipper for litteraturgjennomgang	6
Metode	6
Funn	7
Publikasjoner relatert til tema å forbedre praksis i skolene	10
<i>Administrasjon, tilrettelegging og forbedring av praksis</i>	<i>10</i>
<i>Kort analyse</i>	<i>12</i>
<i>Undersøkelser om hvordan prøvene fungerer i praksis</i>	<i>12</i>
<i>Kort analyse</i>	<i>14</i>
<i>Vurderingspraksis</i>	<i>14</i>
<i>Kort analyse</i>	<i>18</i>
<i>Mestring</i>	<i>18</i>
<i>Kort analyse</i>	<i>20</i>
Publikasjoner under tema styring	20
<i>Kompetanseutvikling</i>	<i>20</i>
<i>Skolen som organisasjon</i>	<i>21</i>
<i>Kommune/skole</i>	<i>21</i>
<i>Ansvar for styring</i>	<i>22</i>
<i>Kort analyse</i>	<i>25</i>
Publikasjoner med tema policy	25
<i>Nasjonalt kvalitetsvurderingssystem (NKVS)</i>	<i>25</i>
<i>Nasjonale prøver</i>	<i>27</i>
<i>Andre temaer</i>	<i>29</i>
<i>Kort analyse</i>	<i>31</i>
Publikasjoner med andre metode/fag	31
<i>Kort analyse</i>	<i>33</i>
Oppsummering og avslutning	34

<i>Om publikasjonene</i>	34
<i>Policy, styring og utvikling av undervisning</i>	34
Litteraturliste	36
Vedlegg	41

Innledning

Denne rapporten tar for seg bruken av Utdanningsdirektoratets ulike prøver utviklet for Skole-Norge. Oppdraget som ble gitt var å undersøke hva vi vet om *bruken* av de ulike prøvene. Vi har valgt å tolke «bruk» i vid forstand og ser på rollen ulike prøver spiller i lærerens jobb og for skolelederne. Vi ser imidlertid også på andre områder, slik som skoleeier, politikktutforming og forskning. For å svare på bestillingen, har vi sett gjennom en rekke arbeider av ulik typer, men med hovedvekt på vitenskapelige arbeider. Vi har prioritert arbeider som tydelig kobler prøvene til de ulike bruksområdene. I arbeidet er det gjort avgrensninger mot mer teknisk orienterte arbeider, slik som valideringsrapporter og metodologiske arbeider. Det er også gjort avgrensninger mot arbeider som bruker prøvene på sekundært vis, for eksempel der prøvene brukes som utgangspunkt for utvalg av kandidater til studier der temaet i seg selv ikke er relevant for denne rapporten. Selv om rapporten ikke bygger fullt ut på metodikken til en systematisk litteraturreview, gir rapporten et inntrykk av hvilke temaer, fagområder, prøvetyper og bruksområder som er bedre dekket og hvilke som virker å være kun i mindre grad undersøkt.

I denne rapporten er det tre prøvetyper som blir berørt: nasjonale prøver, kartleggingsprøver og læringsstøttende prøver. Vi har ikke funnet arbeider som berører karakterstøttende prøver direkte. Rapporten har et begrenset omfang og er gjort innenfor en ramme på omkring 260 timer. Vi har primært lagt vekt på å identifisere og kategorisere aktuelle publikasjoner. Det har ikke vært mulighet for å gjøre mer dypgående analyser av publikasjonene og på tvers av publikasjonene.

Bakgrunn for arbeid med denne rapporten er en forespørsel fra Utdanningsdirektoratet om å beskrive vitenskapelige arbeider på prøvefeltet innenfor følgende tre områder:

- 1) Bruk av prøvene i skolen for å forbedre praksis
- 2) Bruk av prøvene som styringsverktøy
- 3) Prøvenes funksjon i politikktutforming og til andre policyformål

Prøveutvikling i regi av Utdanningsdirektoratet: 2002 - 2019

I 1989 påpekte OECD at det manglet et nasjonalt system for å overvåke kvaliteten i det norske utdanningssystemet (Olsen, Hopfenbeck & Lillejord, 2013). I Stortingsmelding Kultur for læring (2004) kom det forslag om å utvikle indikatorar på resultat-, prosess- og strukturkvalitet (Allerup, Kovac, Kvåle, Langfeldt, & Skov, 2008; Lie, Hopfenbeck, Ibsen & Turmo, 2005). Det ble formulert et mål å bruke skolenes resultater på nasjonale prøver som indikator på resultatkvalitet (Blömeke og Olsen, 2018; Tveit, 2018). Nasjonale tester ble opprettet. Det var Utdanningsdirektoratet som fikk ansvar for dette.

I 2004 gjennomførte Utdanningsdirektoratet for første gang nasjonale prøver på 4. trinn (lesing og regning) og på 10. trinn (engelsk, lesing og skriving) (Lie mfl., 2005). Etter en vurdering av prøvene fra 2004-2005 (Lie mfl., 2005) ble det vedtatt å ta en pause fra prøvene for å kvalitetssikre de nasjonale

prøvene. Det ble startet opp igjen med nasjonale prøver i 2007. Siden 2007 har prøvefeltet blitt videreutviklet med mer sofistikerte nasjonale prøver, flere kartleggingsprøver og andre prøvetyper.

Prinsipper for litteraturgjennomgang

Når det gjelder gjennomgang og oppsummering av forskning, så finnes det ulike tradisjoner og forskjellige metoder. Grant & Booth (2009) trekker frem 14 ulike varianter av forskningsgjennomganger, f.eks. meta-analyse, systematiske kunnskapsoversikter, mapping review, scoping review og litteraturgjennomgang. De mener at litteraturgjennomgang er den mest generelle formen for identifisering og oppsummering av tidligere forskning. Ved en litteraturgjennomgang er det viktig å klargjøre hva som er utgangspunktet og hvilke val som er tatt ved planlegging og gjennomføring (Krumsvik, 2016). Et sentralt premiss ved en litteraturgjennomgang er at det kommer tydelig fram hvordan søke- og vurderingsprosessen er gjennomført.

Identifisering av sentrale tema og begrep

Denne rapporten bygger på et rammeverk for å identifisere sentrale tema, begreper, metoder og funn på følgende tre områder. Disse tre områdene er brukt for å inkludere vitenskapelige arbeider i rapporten:

- 1) Bruk av prøvene i skolen for å forbedre praksis
- 2) Bruk av prøvene som styringsverktøy
- 3) Prøvenes funksjon i politikktutforming og til andre policyformål

Metode

Vi har valgt en praktisk og effektiv innfallsvinkel i denne rapporten på grunn av begrensninger i rammene som ble gitt. I stedet for et svært omfattende litteratursøk i ulike databaser i henhold til typiske innretninger for litteraturreview, så har vi tatt utgangspunkt i fagmiljøene som har utviklet prøvene. Vi tok kontakt med de ulike fagmiljøene, også med sentrale personer som Julius Björnsson og Rolf-Vegar Olsen, og jobbet oss ut fra dette. I tillegg gjorde vi noen avgrensede søk i databaser og så gjennom noen av basene for mastergradsarbeider. Vi har også sett litt på litteraturlister i de ulike arbeidene, men kun i begrenset grad lett aktivt i disse. I hovedsak har vi sett etter arbeider i disse listene som har virket spesielt sentrale.

Vi søkte og fant en rekke abstrakt. Alle abstraktene ble lest. Vi vurderte hvilke abstrakter som omhandlet bruk av prøvene som styringsverktøy og bruk av prøvene for å forbedre praksis, samt prøvenes funksjon i politikktutforming og til andre policyformål (se Prinsipper for litteraturgjennomgang). Vi identifiserte 120 abstrakt som var i tråd med disse tre inkluderingskriteriene. Vi valgt derfor å forsøke å hente inn i fulltekster på disse 120 publikasjonene. Men viste seg å være vanskelig å få tak i fulltekst på alle arbeidene. Det var også tilfeller hvor to abstrakt omhandlet samme publikasjon, f.eks. ved samarbeid om mastergrad. Vi fikk tak i fulltekst på

110 publikasjoner, og etter gjennomlesning av disse valgte vi å utelate 29 arbeider (se vedlegg 1) fordi vi oppfattet at de var på siden av de tre kriteriene for inkludering (se Prinsipper for litteraturgjennomgang). Vi endte dermed opp med 81 arbeider som ble inkludert i rapporten. Disse ble gjennomlest og kodet etter *forfatter, tittel, årstall, type arbeid, prøvetype, prøveår, fagområde, trinn, metode, nivå, pedagogisk orientering, psykologiske egenskaper, organisasjonstema*. Det ble også skrevet en kort kommentar og et litt lengre abstrakt. Arbeidene i rapporten er svært ulike og rammeverket for kategorisering passer noen arbeider bedre enn andre og ideelt sett skulle nok rammeverket vært litt mer raffinert utviklet. Likevel, kategoriseringen gir oss for eksempel mulighet til å vurdere hvordan interessen for å undersøke de ulike prøvene har endret seg over tid, hvilke temaer som har vært av interesse og hvilke typer prøver som virker mest interessant for forfatterne.

Tilnærmingen som er valgt innebærer at viktige temaer kan ha blitt utelatt, at relativt sentrale miljøer har blitt oversett og at kategorisering og oppsummering dermed kan være skjev. Å gjennomføre en mer fullstendig litteraturreview for hele prøvefeltet vil være krevende, men opplagt nyttig. Med tanke på den store bredden i prøvetyper, faglige og metodiske temaer, ulike aktører (som skole, skoleeier, policy, forskning), kan det være lurt å dele et slikt arbeide noe opp.

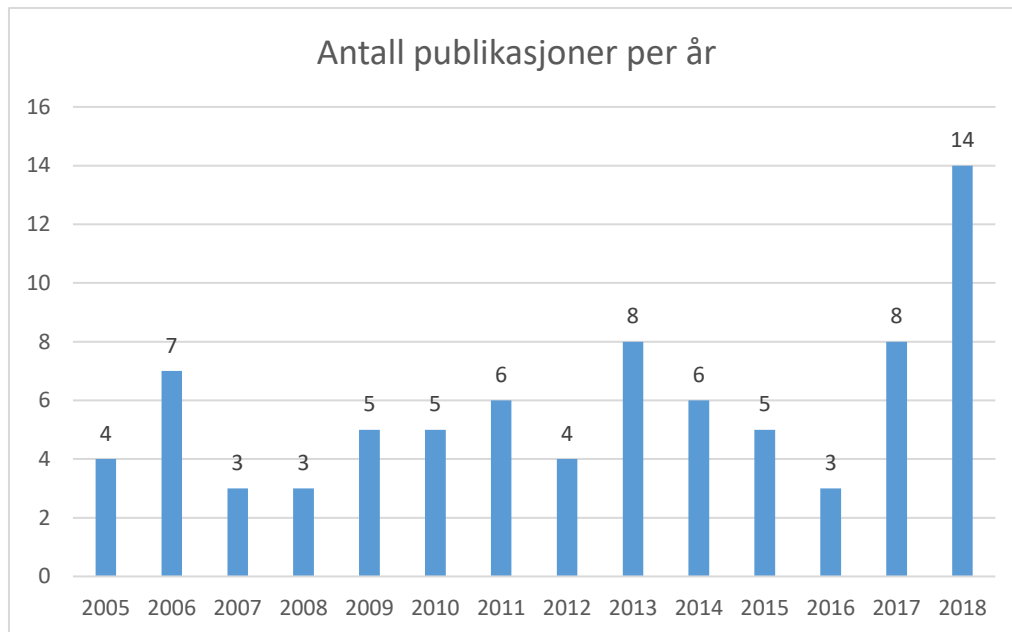
Som sagt har vi skrevet et kortfattet abstrakt fra hvert arbeide. Vi har benyttet dette som hjelp i analysearbeidet og ved kategorisering, men det er viktig å peke på at abstraktet ikke må leses som en oppsummering av arbeidet, til det er det altfor kortfattet. Mange av arbeidene dekker temaer som ikke er direkte relevante for denne rapporten, noen har sågar hovedvekt på slike temaer. I slike tilfeller vil abstraktet vi har inkludert kun referere til de delene vi mener er sentrale for temaet i denne rapporten. Vi har inkludert abstraktene i rapporten og systematisert disse etter tema. De korte inntrykkene av arbeidende er ment å gi et inntrykk av temaer og hjelpe leseren til å velge ut relevante arbeider han eller hun ønsker å lese mer om. Fremfor noe, så viser abstraktene den store bredden i temaer og problemstillinger som forskere og andre har berørt i sine arbeider. Denne bredden gjør det krevende å systematisere arbeidene. Vi har gjort et forsøk i denne rapporten, men det er klart at siden mange av arbeidene dekker mer enn et område, så vil kategoriene og plasseringen av arbeider i disse kunne diskuteres videre. For eksempel er mange arbeidene naturlig knyttet til metodiske og faglige forskningstemaer, men for å unngå at alt for mange av arbeidene havner i denne kategorien, har vi forsøkt å plassere flest mulig av arbeidene i andre relevante kategorier.

Det er som sagt bruken av de ulike prøvene som står sentralt i denne rapporten. De ulike aktørene i skolen benytter prøvene på ulikt vis og til ulike formål. Det er primært lærere og skoleledere som har direkte mulighet til å bruke prøvene i skolen for å forbedre praksis. Videre er det primært skoleledelsen og skoleeier som tar i bruk prøvene som styringsverktøy. Dessuten er det trolig skoleeier og myndigheter som kan anvende prøvene i politikktutforming og til andre policyformålet. Forskere er på sin involvert i alle disse nivåene, men har også egne forskningsområder som kommer frem i mange av arbeidene.

Funn

Gjennom vårt litteratursøk endte vi til slutt opp med å gjennomgå 81 publikasjoner i fulltekst fra årene 2005-2018. I løpet av disse 14 årene er det i gjennomsnitt på 5,8 publikasjoner per år med

spredning fra 3 publikasjoner (2007, 2008 og 2016) og 2016 til 14 bidrag i 2018. En forklaring på det siste kan være at Acta Didaktika hadde et temanummer om prøver i skolen i 2018.



Bakgrunn om publikasjoner

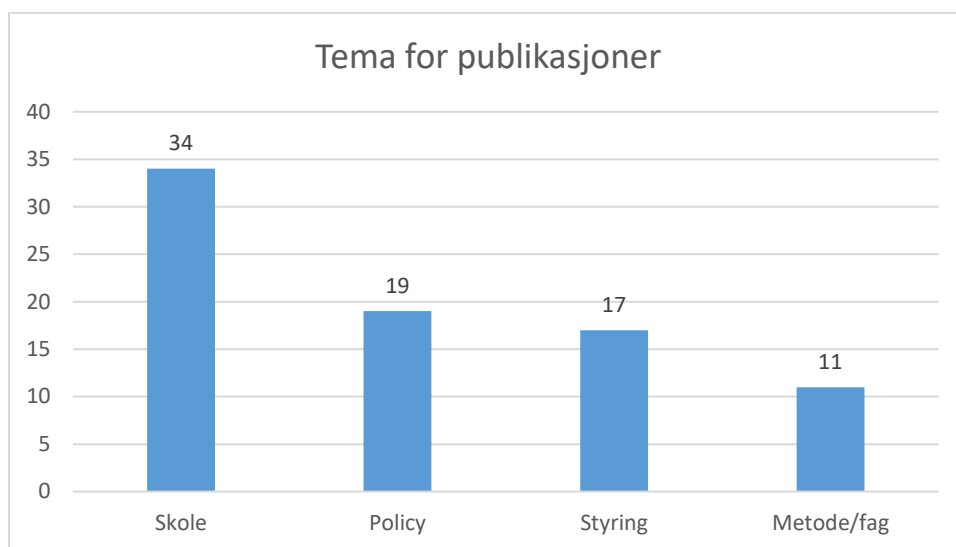
Det er ulike typer publikasjoner. Det er fagfelleverderte publikasjoner som artikler og kapitler. Videre er det mastergradsoppgaver og rapporter. Blant de publikasjonene vi fant er det 38 fagfelleverderte publikasjoner (47%), 29 mastergrader (36%) og 14 rapporter (17%).

I denne perioden har Utdanningsdirektoratet utviklet nasjonale prøver, kartleggingsprøver, læringsstøttende prøver og karakterstøttende prøver. Vi har funnet 50 publikasjoner som kun tar for seg nasjonale prøver, 8 publikasjoner tar for seg kartleggingsprøver, 2 publikasjoner omhandler læringsstøttende prøver, 12 flere prøvetyper (eller prøver generelt) og 9 under kategorien «annet» som blant annet dekker artikler som tar for seg prøvefeltet mer indirekte.

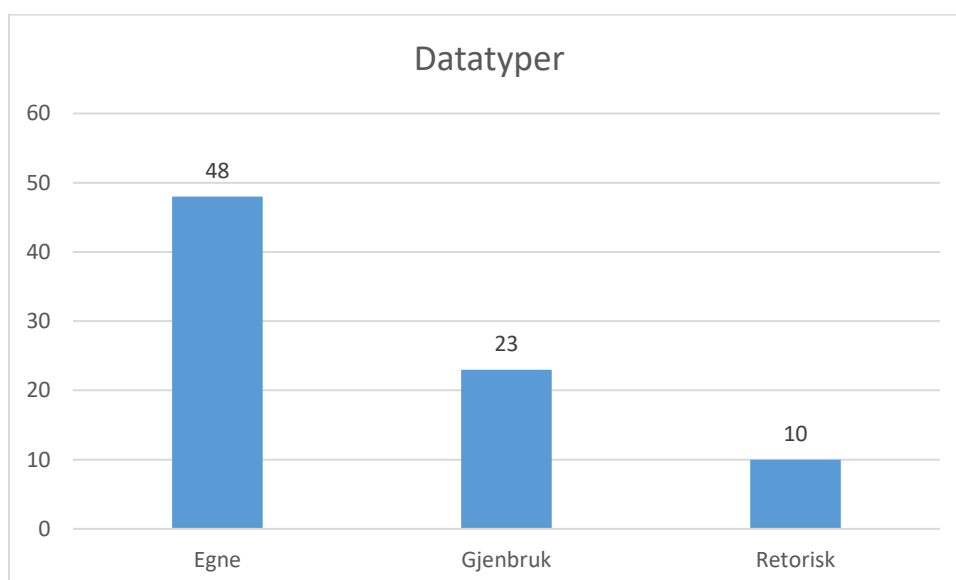
Når det gjelder fagfelt, så er det 36 publikasjoner som enten dekker flere fag eller ingen spesielle fag. Det er 24 publikasjoner som omhandler norskfaget (i hovedsak lesing), 12 publikasjoner omfatter matematikk (i hovedsak regning), 2 dekker digitale ferdigheter og 7 er under annet.

Det er 28 publikasjoner med analyser av intervju, 22 publikasjoner med analyse av prøver, 15 analyser av retorisk type, det er 7 publikasjoner med survey, det er 5 publikasjoner med mixed metode og 4 publikasjoner med annet.

Tema for skolen har vært inkluderingskriterier for publikasjonene. Det er 34 publikasjoner innen tema forbedre praksis ved skolene, 19 publikasjoner om tema policy, 17 publikasjoner om tema styring, og 11 publikasjoner innen metode/fag.



I 48 av 81 publikasjoner er det samlet inn egne data, mens det er 23 publikasjoner som har gjenbruk av data eller prøver (inkl. Oppgavesett fra en aktuell prøve). Dessuten er det også 10 publikasjoner som er av mer retorisk art, og som ikke har kvalitative eller kvantitative data som analyseres. De retoriske arbeidene er i hovedsak knyttet til policyområdet, mens det er mer jevn fordeling av datagrunnlag for de andre områdene.



Kort oppsummert ser vi en nokså jevn forskningsinteresse på prøvemarket over de siste 15 årene. Leseprøvene har fått mest oppmerksomhet. Det er en overvekt av arbeider som har skolen og praksiser der som utgangspunkt. Selv om det er en del gjenbruk av prøvedata, har de fleste

arbeidene gjort egne undersøkelser. Det er svært få studier som har benyttet mixed methods, selv om prøvefeltet egentlig er godt egnet for denne tilnærmingen.

Publikasjoner relatert til tema å forbedre praksis i skolene

Ett av områdene vi har sett på er hvordan prøvene er benyttet for å utvikle skolenes praksis. Prøvene fungerer på flere måter i denne sammenhengen, både som utgangspunkt for ledere, for læreres undervisnings- og vurderingspraksis og for skoleutvikling mer generelt. Noen av arbeidene vi fant rettet seg mer mot mestringsperspektiver. Det er 34 publikasjoner som er relatert til det å bruke prøvene for å forbedre praksis i skolene. Det er 12 fagfelleverderte publikasjoner, 17 mastergradsoppgaver og 5 rapporter.

Vi har organisert disse artiklene under fire **overskrifter**:

- **Administrasjon, tilrettelegging og forbedring av praksis**
- **Undersøkelser om hvordan prøvene fungerer i praksis**
- **Endringer i vurderingspraksis**
- **Mestring**

Administrasjon, tilrettelegging og forbedring av praksis

En viktig side av prøveområdet er hvordan skolene organiserer arbeidet rundt prøvene. Det varierer hvordan skolene administrerer de ulike prøvene, hvem som har ansvaret for gjennomføringen, hvilke forberedelser som gjøres og i hvilken grad prøvene brukes til å forbedre praksis.

Berger (2008) gjennomførte en mastergrad med tittel «Nasjonale prøver 2007: arbeidet skolene etter intensjonen?: hvordan takler skolene presset?». Tema er om de nasjonale prøvene i realiteten er vurdering *for* læring slik intensjonen er formulert fra Utdanningsdirektoratet, eller er de vurdering *av* læring? Oppgaven baserer seg på en survey til lærere fra tre skoler. I tillegg er det intervjuer med fem lærere og en rektor fra to skoler. Berger tar opp hvilke dilemmaer som oppstår i tilknytning til de nasjonale prøvene og hva skolen gjør i forhold til eventuelle dilemma og hvordan de oppfatter eventuelt press knyttet til resultatene fra de nasjonale prøvene. Skolene i prosjektet satte seg inn i relevant materiell. De forberedte seg grundig på gjennomføring av prøvene. Skolene har også systemer for oppfølging i etterkant av prøvene. Skolene er likevel usikre på rollen til de nasjonale prøvene, og de oppfatter at de nasjonale prøvene har en kontrollfunksjon.

Eriksen (2011) har en masteroppgave om hvordan nasjonale prøver har påvirket til endring av praksis på to norske skoler. Hensikten er å belyse informantenes syn på «nasjonale prøver, hvordan man jobber med forbedringsarbeid på skolen sammenlignet med tidligere, om og hvordan

skoleledelse har endret seg, om og hvordan informantene opplever at nasjonale prøver bidrar til konkurranse og markedsliknende mekanismer i skolen». Det er intervju av to skoleledere og ni lærere fra to skoler. Det ser ut som begge skolene er suksessrike fordi de over tid har levert gode resultater på nasjonale prøver. På begge skolene opplevde informantene at det var blitt endringer i praksis, og at dette er endringer som kan knyttes til nasjonale prøver. Skolene var gjennomgående positive til nasjonale prøver, og det ble jobbet bevisst med oppfølging av resultater fra prøvene. Men samtidig fører prøvene med seg større fokus på konkurranse og sammenlikning med andre skoler.

Gunnulfson (2017) har publisert en artikkel med analyse av hvordan lærere og skoleledere forsøker å forholde seg til myndigheters forventninger om at skolen skal bruke resultat på nasjonale prøver for å utvikle kvalitet. Det er samlet inn og analysert opptak fra møter med skoleledelse og lærerteam. Gunnulfson finner at lærere og skoleledelse tar ulike roller i dette arbeidet. Skoleledelsen er mer entusiastiske fortellere, mens lærere tar ansvar for å håndheve og iverksette. Gunnulfson konkluderer med at forventning om å bruke nasjonale prøveresultater i kvalitetsutvikling forsvinner i oversettelse av prøvene.

Nes (2013) master med tittelen «Nasjonale prøver på mellomtrinnet» som undersøker hvordan de nasjonale prøvene brukes i skolen. Med bruk i skolen menes det «hvordan lærerne arbeider under forberedelsene av sine elever i forkant av prøvene». Lærernes arbeid med de nasjonale prøvene brukes som en case for å undersøke hvordan ansvarlighet («accountability») påvirker læreres arbeid. Metode inkl. survey med 211 lærere og intervju med elever. Hovedfunnene «er tid, elever som gruer seg og å øve på prøve». Tid innebærer tid til å øve og tid til å gjennomføre. Elever som gruer seg, og at man derfor kan øve seg til prøven. Det ser også ut til at lærerne opplever stor grad av autonomi under arbeidet med forberedelsene mot de nasjonale prøvene, men at lærerne må sikre at elevene kjenner seg trygge. Det er under 50% som opplever at nasjonale prøver er gode redskaper.

Nortvedt (2018) kombinerer analyser av prøveresultater fra kartleggingsprøven i regning på 1.-3. trinn med dybdeintervjuer av lærere i et mixed methods design. Målet er å undersøke hvilke erfaringer lærerne har med kartleggingsprøvene og hvilke trender vi kan se på elevsiden. Over fire års gjennomføring med samme prøve (2014-2017) finner Nortvedt at andelen elever under bekymringsgrensen er stabil. Nortvedt peker på at det normalt ville være forventet at andelen ville synke fordi prøven er kjent og at lærerne da bedre kan forberede elevene og fordi tiltak burde fått tid til å virke. Videre peker hun på at elevene virker å ha grunnleggende tallforståelser med økende forståelse med alder, men at det grunnleggende tallbegrep ikke er godt nok utviklet. Hun peker på at 3. klassingene under bekymringsgrensen ikke klarer addisjonsoppgaver med tieroverganger og at dette indikerer at skolene kan bli flinkere til å følge opp elever som strevet med dette. Lærerne forteller fra intervjuene at det ikke alltid er så enkelt å bruke prøvene for å utvikle egen praksis, altså å finne produktive og gode praksiser som kan hjelpe elevene. Ikke alle lærerne hadde god nok forståelse av prøveformater og hensikter med prøvene. Lærerne snakker også om «svake elever» og indikerer dermed en svekket tro på hva disse elevene kan få til. Nortvedt peker på at bruken av prøvene må bli bedre for at de skal kunne bidra til utvikling av elevenes ferdigheter.

Tavas (2006) har i sin mastergrad tatt for seg læreres oppfatninger av leseprøven i nasjonale prøver. Vi snakker her om de første nasjonale prøvene som ble gitt. Tavas finner at lærerne er positive til prøven som de mener har flere gode egenskaper. Lærerne opplever likevel at prøvene ikke tar utgangspunkt i læreplanen og læreverkene. De peker på også på at prøvene mangler stabilitet og at dette blant annet hindrer måler over tid. Likevel, Tavas finner mindre endringer i lærernes praksiser i klasserommet. Det er problemer knyttet til gjennomføringen av prøven, ikke minst når det gjelder informasjon. Tavas finner også at lærerne mangler kompetanse, blant annet i den videre leseopplæringen.

Vestheim (2018) har intervjuet rektorer og lærere for å undersøke hva som kjennetegner skoler som har oppnådd gode resultater på nasjonale prøver. Det er praksiser ved disse skolene som er av spesiell interesse. Profesjonelle og kollektive praksiser kjennetegner skolene som gjøre det godt på nasjonale prøver. I slike kollektiver ligger det rom for utvikling.

Kort analyse

Et sentralt tema knyttet til hvordan ulike prøver brukes til forbedring av praksis er knyttet til rollene aktørene i skolen har. Ideelt sett burde rollene være tydelig avklart slik at ansvaret for de ulike prosessene skolen må gjennom i forbindelse med prøvene er avklart. Det virker å være usikkerhet rundt roller og ansvar. Det kan være behov for å tydeliggjøre de ulike aktørenes rolle i arbeidet med prøvene. Videre er fremdeles problemstillinger knyttet til konkurranse og sammenligning eller rangering til stede. Det er nokså unisont fra praksisfeltet at dette ikke så positivt og at den pedagogiske bruken kommer i skyggen av det prestasjonsrettete perspektivet. Mange lærere opplever ikke nasjonale prøver (i 2013) som så nyttige og det er også pekt på at kartleggingsprøvene må bli bedre. En av utfordringene som ses i mange av rapportene er det å «oversette» prøveresultater til aktive handlinger som kan bedre praksis. Det er interessant å sette dette i perspektivet «profesjonelle og kollektive praksiser». De store ulikhetene mellom skolene når det gjelder arbeidet med de ulike prøvene oppleves av noen som uheldig.

Undersøkelser om hvordan prøvene fungerer i praksis

Mange av arbeidene i denne seksjonen tangerer temaer om metode/fag og styring. Vi har valgt å samle disse under praksis-fanen fordi de gir informasjon om hvordan elever, lærere og ledere forholder seg til de ulike prøvene, hvordan de forstår begreper og hvordan de løser oppgaver. Det er også temaer som er knyttet til ressurser og forhold rundt styring, men arbeidene her er kontekstualisert i praksis, dermed er de tatt med her.

Hatlevik & Radtke (2017). Standard Setting in a Formative Assessment of Digital Responsibility Among Norwegian Eighth Graders. Dette er et fagfelleurdert kapittel om læringsstøttende prøver. Det dreier seg om å beskrive en prosess med hvordan fastsette mestringsnivå. Det viser seg at en

gruppe med lærere, skoleledere og forskere har litt ulike synspunkter på hvordan sette grenseverdier. Dette viser utfordringer når det gjelder å komme med en omforent forståelse av mestringsnivå.

Holthe (2007). Ordforråd og syntaktisk kompleksitet i argumenterende elevtekster: en studie av et utvalg tekster skrevet av tiendeklassinger til den nasjonale prøven i skrijving våren 2005. Målet med denne masteroppgaven er å studere på ordforråd og syntaktisk kompleksitet i et utvalg argumenterende elevtekster skrevet av tiendeklassinger. Det er plukket ut 20 av besvarelsene til den første oppgaven på den nasjonale prøven våren 2005. Ut fra analysekriteriene, har de sterke tekstene gjennomgående et langt mer avansert ordforråd, og en mye høyere syntaktisk kompleksitet enn de svake tekstene.

Landmark (2018) tar opp nasjonale prøver som kvalitativt styringsverktøy for elever. I denne mastergrad er det intervju med fire lærere og to rektorer. Informantene har ulikt syn og praksis på hvordan man oppfatter formålet til nasjonale prøver, hvordan man praktiserer muligheten for fritak og øving, og hva man benytter prøvene til i etterkant. Selv om dette kun er seks stemmer viser det at denne forskjelligheten kan få konsekvenser for enkelte elever og skoler.

I en undersøkelse basert på mixed methods, undersøkte Nortvedt (2011) hvilke strategier elever med normalt gode matematikkferdigheter, men med svakere leseferdigheter, brukte når de løste teksttunge oppgaver i matematikk («word problems»). Forskeren finner at elever i den aktuelle gruppen benytter nøkkelord og stereotypisk forståelse av «word problems» i løsningen. Dette hjelper dem og kompenserer for de svakere leseferdighetene.

Nortvedt og Rohatgi (2018) har sett på bruken av kartleggingsprøver i matematikk i Oslo kommune. Innledningsvis fremmer de teorier om «instructional leadership» som en effektiv form for ledelse. Sentralt i denne retningen er at ledelsen aktiv veileder for, stiller krav og legger til rette for gode læringsmiljøer. Lærere og ledere arbeider sammen for å heve undervisning og elevenes læring. Kartleggingsprøvene kan her spille en rolle, ikke minst med tanke på oppfølgingsarbeidet i etterkant av prøvene. Rapporten har mange funn både fra det forberedende arbeidet, fra selve gjennomføringen og fra arbeidet i etterkant. Ett interessant funn er at mens skolens ledelse i stor grad peker på læreren som sentral i oppfølgingsarbeidet, rapportere omkring halvparten av rektorene at de mangler ressurser for å oppnå god oppfølging i etterkant av prøvene. Dette går på antall lærere, lærernes fagdidaktiske kompetanse og tilgang på utstyr. Det er lagt til rette for at så mange elever som mulig skal kunne gjennomføre prøvene og det er på mange skoler systemer for å fange opp elever som for eksempel er fraværende under prøven. Det er også stor grad av samarbeid mellom lærerne i oppfølgingsarbeidet. Skolens ledelse er sentralt og har et ansvar for gjennomføring og tilrettelegging i etterkant. Det er med andre ord mange forutsetninger for god oppfølging av prøvene, selv om det virker å være ressurs-/kapasitetsutfordringer.

Sandset (2015) har i sin mastergradsoppgave sett på hvilke utfordringer elever møter i den nasjonale leseprøven, hvilke strategier de bruker og hvordan elevenes leseprosesser stemmer overens med det som ligger til grunn for prøven. Hun undersøker seks elever gjennom bruk av en høyttenkningsmetode og intervju. Hun finner at elevene hadde problemer med å løse oppgavene på nivå 2-5, mens det laveste gikk greit. Elevene strever med å kontrollere egen leseforståelse. Sandset finner at elevene har utilstrekkelige kognitive strategier for å løse oppgavene de skulle løse. Elevene benyttet det de husket fra teksten de hadde lest og benyttet ikke andre strategier for å løse oppgavene selv om disse krevde dypere forståelse av teksten for å løse. Videre var det enkelte ord i tekstene som var vanskelige for elevene å forstå. Det er opplagt at elevene trenger ulike strategier og kompetanser for å løse oppgavene.

Strøm (2006) fant i sin hovedoppgave (tilbake helt i starten av nasjonale prøver) at lærerne var positive til prøven, men at de ikke ønsker offentliggjøring av resultater eller at resultatene skal brukes til lønnsfastsetting. Dette er gjenkjennbare utsagn fra denne fasen av nasjonale prøver. En god del av funnene i oppgaven er mindre interessante i denne konteksten fordi prøvefeltet i dag har utviklet seg i stor grad, men oppgaven er interessant som et historisk referansepunkt.

Kort analyse

Det er nyttig å involvere lærere i prøvearbeidet. For eksempel kan prøveutviklere/eksperter ha et annet syn på mestringsnivåer og bekymringsgrenser enn lærere. Videre er det flere undersøkelser som viser at lærere og ledere ønsker å benytte prøvene til utvikling, men at dette ofte er vanskelig. Et annet moment er at elevene trenger strategier for å løse oppgavene og at dette er noe som skolene må ha i mente. Kjønnforskjeller er et sentralt tema på prøvefeltet, både som fenomen i seg selv og i forbindelse med validitet og invariansvurderinger (at prøvene behandler ulike grupper rettfærdig).

Vurderingspraksis

Et svært sentralt tema i denne rapporten er hvordan ulike prøver brukes i vurderingsøymed. Vi har her samlet en rekke arbeider som undersøker hvordan lærere, men også andre aktører i skolen, bruker prøvene som deler av sitt vurderingsarbeid. Sentralt i mange av arbeidende er status for hva skolene gjøre med prøvene og resultatene, og spesielt er dette knyttet til vurdering for læring og koblingen mellom prøveresultater og endringer i undervisningspraksis.

Almaas (2017) har en masteroppgave med tittel «Hvordan brukes elevresultater til læring i skolen». Almaas ser på nasjonale prøver fra 5. og 8. trinn i flere fag. Almaas tar opp hvordan elevresultater brukes til læring i skolen. Utgangspunktet er nasjonale prøver som gjennomføres på 5. og 8. trinn i engelsk, lesing og regning. Analyse ut fra intervjuer med tre lærere og en rektor. Skolen har ikke noe tydelig mål for nasjonale prøver. Lærerne kjenner heller ikke til dette målet. Det viser seg også at resultatene fra prøvene i større eller mindre grad følges opp.

Evensen mfl (2016) har en artikkel som bygger på en antakelse om at vurderingskriteriene fra de nasjonale prøvene kan ha pedagogisk betydning. Sentralt i NORM-studiet er standarder og normer. Metodisk har de en stegvis tilnærming. Steg 1 innebærer at de ut fra eksisterende læreplaner og relevant litteratur utvikler et sett av kriterier aktuelt for nasjonale prøver i lesing (2005) for 4. trinn, 7. trinn, 10. trinn og Vg1. I steg 2 blir kriteriene gjennomgått og vurdert av erfarne lærere. Steg 3 innebærer at det gjennomføres analyser av utskrifter fra intervjuer. De kriteriene som ble identifisert ble anvendt og prøvd ut i daglige klasseromskontekster. Forfatterne finner at lærerne kan kombinere formativ og summativ vurdering fra samme prøve. Dette peker de på at er krevende, men likevel mulig å få til.

Høihilder & Engh (2006) står bak en rapport om «Elevvurdering på grunnskolens sjuende trinn i Vestfold». Dette er en rapport om læreres erfaring med vurdering. De har også med spørsmål til lærere om nasjonale prøver. Formålet med undersøkelsen var å innhente informasjon om lærernes praksis. De finner en ganske lik fordeling av positive og negative svar på spørsmålene om de nasjonale prøvene har ført til endringer i vurdering og undervisning.

Johansen (2017). Nasjonale prøver i lesing. Hvordan tar lærerne resultatene i bruk? Masterstudien fokuserer på Fjellbekken skole og hvordan noen av lærerne på denne skolen arbeider med resultatene fra nasjonale prøver. Johansen gjennomførte et kvalitativt forskningsintervju med tre informanter (erfaring fra 5. trinn). Resultatene fra studien tyder på at «lærere tar i bruk resultatene fra de nasjonale prøvene i lesing med det formålet at det skal føre til vurdering for læring». Dette skjer i etterkant av prøven.

Ljønes (2012) har en mastergrad om nasjonale prøver. Det er en undersøkelse av 5. trinnelevers lesekompetanse i lys av nasjonal leseprøve. Det dreier seg om prøvebesvarelsene som et «redskap for vurdering av og videre veiledning knyttet til elevenes leseforståelse og leseopplæring». Prosjektets enkeltcase er en liten kommune med fire deltagereskoler. Det er til sammen 98 elever på 5. trinnet som utgjør analyseenhetene. Det er hentet inn tilleggsinformasjon gjennom en spørreundersøkelse rettet mot 5. trinns lærerne. En konklusjon er at lærergruppen stort sett har kjennskap til klassen og den enkelte elevs resultat på leseprøven. Samtidig gir lærerne uttrykk for at resultatene har liten innvirkning på leseopplæringen.

Myhre (2014) har en master i matematikdidaktikk. Det er en kvalitativ studie av hvordan nasjonale prøver i regning blir brukt i skolen. Han er opptatt av hvordan nasjonale prøver i regning blir brukt i skolen ut ifra sitt formål. Utvalget består av en ren barneskole, en ren ungdomsskole og en 1-10-skole. Til sammen ble det fire klasser (fra tre skoler). Det er brukt observasjon, intervju og dokumenter for å innhente data fra fire klasser. Han ser på veiledning til prøvene og beskrivelse av analysene fra prøvene. Prøvene kan bli brukt i en formativ vurderingspraksis i etterkant, men han

finner at det er manglende bruk av prøvene i klasserommene i etterkant. De nasjonale prøvene bidrar til sammenligning og konkurranse mellom skolene. Det kan føre til at fokus blir resultatet, og ikke læringsprosessen til elevene. Informantene viser til hvordan skolene kan ha nytte av prøvene, men det er eksplisitt om hvordan lærere/skole kan bidra til at elevene får direkte nytte av prøvene.

Nilsen (2014) skrev sin masteroppgave om hvordan kartleggingsprøven i lesing påvirker skolen. Hun intervjuet tre lærere og benyttet fenomenologisk tilnærming. Nilsen finner at grunnleggende ferdigheter ikke blir vektlagt i stor grad og at lærere i andre fag i liten grad opplever at de også er lærere i lesing. Videre finner hun at lærerne ikke vektlegger lesestrategier. Når det kommer til vurderingsarbeidet er heller ikke prøvene her tatt i bruk på optimalt vis, det er mye å hente. Her peker hun på gode tilbakemeldinger på elevenes leseprestasjoner koblet med tydelige fremovermeldinger. Nilsen peker på andre forskere når hun argumenterer for at tilbakemeldinger må være forståelige, dermed blir lange og kompliserte tilbakemeldinger mindre effektive. Hun gjør et litt overraskende funn blant sine tre informanter når hun skriver at lærerne oppfatter læringsstrategier som måter lærere planlegger og tilrettelegger for i undervisningen, ikke koblet til elevenes egne strategier som kanskje er den vanligste forståelsen. Det virker som om et såpass sentralt begrep som læringsstrategi oppfattes grunnleggende ulikt. På samme måte er også begrepet tilpasset opplæring snevert forstått av lærerne. Det er ikke gitt at lærere deler prøveresultater med andre lærere selv om det i utgangspunktet er et felles ansvar å utvikle elevenes grunnleggende ferdigheter. Det ender med at det er norsklærerne som følger opp. Vurdering som begrep virker også nokså unyansert forklart av lærerne, selv om dette kanskje kan ha endret seg til det bedre siden 2014. Hvis man forventer at kartleggingsprøvene skal utløse relevante og gode tiltak for elevene som trenger dette, bør momentene Nilsen finner klart adresseres. Nilsen skriver avslutningsvis om ansvaret for oppfølging av kartleggingsprøven og peker på svakheten ved at det er enkeltlærere som har dette. Ledelsen bør involveres.

I en artikkel som sammenligner Norge og Portugal på temaet vurdering, peker Nortvedt, Santos og Rinto (2016) på at lærere generelt strever med formativ vurdering, eller «assessment for learning» som de bruker som begrep. De har sett spesielt på matematikk og skriver at læreres «beliefs» om matematikk er knyttet til vurderingspraksisen deres. Artikkelen poengterer at lærere ofte ikke er flinke til å forklare hva de forventer av elevene. Videre at lærere i for stor grad venter med å sette inn tiltak i håp om at vansker skal gå over av seg selv. Lærere mangler kompetanse og at læreres autonomi bidrar til store forskjeller i vurderingspraksisen. Forskerne mener også at det mangler mer forskning på læreres vurderingspraksis innen matematikkfaget.

Rangnes og Fodstad Gourvenec (2018) har undersøkt hvordan lærere reflekterer over flespråklige elevers prøveresultater i den læringsstøttende prøven som tar for seg elevens vokabular. Gjennom bruk av fokusgruppeintervjuer med åtte lærere finner forskerne blant annet at lærerne oppfattet at prøven var vanskelig for mange elever. Flere gav opp. Forhold som alder, om eleven var stresset, jobbet for raskt, var vimsete eller unøyaktig eller ikke husket undervisningen hadde betydning. Elever som hadde relativt store vansker, uavhengig av morsmål, endte opp med mye gjetting. Hvorvidt de

flerspråklige elevene presterte som forventet varierte, flere oppgaver at disse elevene presterte bedre enn forventet. Videre, lærerne oppgav at prøven hadde verdi fordi de da fikk oversikt over mangler ved elevenes vokabular. Det er åpenbart at prøvene er informative for lærerne, ikke minst fordi de ser hvilke ord som det må øves på. Det er interessant hvordan lærerne i undersøkelsen er kritiske til egen praksis og peker på svakheter ved denne.

Ravlo (2009) ser på hvordan resultater fra nasjonale prøver i regning kan brukes av lærere. Forskeren vektlegger at vi her snakker om en grunnleggende ferdighet og at elevenes prestasjoner dermed er relevante også for andre fag. Videre vektlegger hun at prøvene har en verdi i vurdering for læring, samtidig som prøvene har en funksjon i styringsøyemed. Ravlo mener at det er nyttig å hente ut detaljert informasjon om elevene på itemnivå, selv om det kan ta litt tid.

Stornes, Tvedt og Bru (2013) vurderer normative tester i skolen i lys av motivasjonsteori. Forskerne peker på at skolemotivasjon synker tydelig fra 4. – 10. trinn og argumenterer for at mange elever opplever økt press og negative forventninger i en konkurransesituasjon drevet av normative prøver. Hvis læreren fokuserer på rangering av elevene, vil elever som presterer svakere på prøvene oppleve redusert tro på egne evner. Videre kommer artikkelen inn på viktigheten av at prøveresultatene kan gi individuelle tilbakemeldinger. Det refereres til Harlen & Crick (2003) når det argumenteres for at overdrevent fokus på sentralt gitte prøver, som er innrettet for å måle relativt kunnskapsnivå, kan medføre at læreren i mindre grad benytter seg av egen kunnskap og slik påvirker den individuelle vurderingen negativt. Artikkelen konkluderer med at nasjonalt gitte, normative prøver har en funksjon for kommune og stat, men at verdien i det læringsstøttende arbeidet er liten. I verste fall kan slike prøver være negative for elevene, spesielt de som presterer lavere.

Berge (2005) har en artikkel om skriving som grunnleggende ferdighet og som nasjonal prøveideologi og strategier. Berge trekker fram og diskuterer hvordan prøveresultatene fra nasjonale prøver er ment brukt. Videre kommer artikkelen på fundamentale forskjeller på lesing og skriving som ferdigheter som skal måles. Berge er også opptatt av en mulig «pedagogisk konflikt mellom skrivepedagoger som kan karakteriseres som kunnskapsorienterte formalister på den ene siden og skrivepedagoger som er elevorienterte situasjonister på den andre».

Falk (2006) har en masteroppgave om lærerperspektiver på nasjonal prøve i 1MX våren 2005. Han har analysert 60 besvarelser og snakket med lærer. Han analyserer hvilken didaktisk informasjon prøven gir til læreren. Falk finner blant annet et fokus på representasjonskompetanse og kompetanse i symbolbruk. Oppgaven som bygger på en tidlig versjon av nasjonale prøver kritiserer klassifiseringssystemet (kompetanse i matematikk er delt opp i åtte delkompetanser) i prøven, det virker å ha vært for ambisiøst. Falk kommer inn på flere momenter knyttet til lærernes bruk av prøveresultatene, blant annet rangering, tidspunkt på året for gjennomføring og opplevd nytte.

Kort analyse

Det er noen funn som tyder på at de ulike prøvene har begrenset impact på lærernes vurderingspraksis, selv om det også er funn som er litt mer optimistiske her. Ofte kjenner læreren klassen allerede og dermed har ikke nødvendigvis prøvene så stor effekt. På den andre siden er det også rapporter som peker på at lærernes oppfatninger om klassen kan bli overstyrt av prøveresultater. Det kan være nyttig å bruke enkeltitems in det videre vurderingsarbeidet, selv om dette kan være litt tidkrevende. En utfordring ligger i at lærerne ikke har omforent forståelse av sentrale begreper som læringsstrategier og tilpasset opplæring, en annen utfordring ligger i at lærerne venter for lenge med å sette inn tiltak. Noen elever gir opp for tidlig og dette kan gi misvisende informasjon.

Mestring

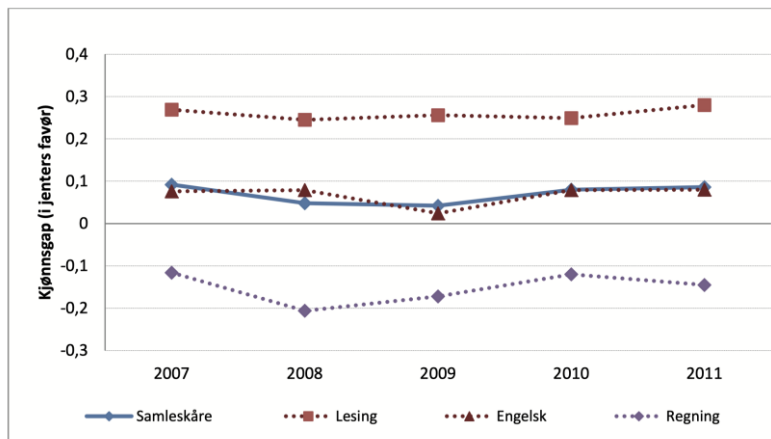
En rekke arbeider er opptatt av elevenes mestring. Mange av disse er tekniske rapporter og ulike faglige arbeider. Arbeidene som er tatt med her forsøker å gi et bilde av hva elevene mestrer eller hvordan ulike grupper (for eksempel gutter/jenter) presterer på prøvene.

Almestrand (2012) «Hva gjør de gode leserne?». I denne masteroppgaven bygger Almestrand på oppgaver brukt i de nasjonale prøvene. Almestrand har tatt tre oppgaver (av sju) fra leseprøvene (nasjonale prøver). Disse oppgavene blir løst av fire elever som er predefinert som gode lesere. I sine analyser av resultatene fra prøvene legger Almestrand vekt på hva som kjennetegner de strategiene som brukes av de god leserne. En mulig konklusjon kan være å forbedre praksis ved å hjelpe lærere til å bli klar over elevers lesestrategier.

Alseth mfl (2009) har en rapport som omhandler to kartleggingsprøver i tallforståelse og regneferdighet. Det er to deler. Del 1 omhandler data fra 2. årstrinn. Det er analyser av kjønn, hjemmeforhold og minoritetsstatus opp mot resultater fra kartleggingsprøven på 2. årstrinn. Del 2 omhandler data fra videregående skole, Vg1. På data fra videregående skole finner de at det er forskjeller mellom yrkesfag og studieforberedende. For yrkesfagene er det en sammenheng (korrelasjon) mellom skår på prøven og utdanningsambisjoner. Denne sammenhengen finnes ikke for studenter ved studieforberedende. Dette forklares med at mens det i større grad er de høytpresterene yrkesfagsstudentene som har ambisjoner om mer utdanning, så er de fleste studentene ved studiespesialiserende innstilt på lengre utdanninger.

Bakken & Elstad (2012) er en omfattende rapport som har som mål å vurdere om Kunnskapsløftet har bidratt til utjevne forskjeller mellom ulike grupper (kjønn, sosioøkonomi, etc.). Her ser de på endring i karakterer fra før og etter KL06. De har også med resultater på nasjonale prøver og dette kobles til kjønn.

Figur 5-4 Kjønnforskjeller på nasjonale prøver, 8.klassetrinn, under Kunnskapsløftet i årene 2007 – 2011, målt i standardavvik



Jentene skårer bedre enn guttene i lesing, noe bedre i engelsk, mens guttene har et lite forsprang i regning. Alle prøvene under ett skårer jentene noe bedre. Forfatterne finner at sosioøkonomiske forskjeller tenderer å øke for de fleste fag, grupper og mestringsnivåer. Rapporten kobler ikke nasjonale prøver til KL06 eller til karakterer.

Bråthen (2011). Dette er en masteroppgave om hva som kjennetegner 2. klassingers leseferdigheter slik dette blir kartlagt i nasjonale kartleggingsprøver. Bråthen undersøker hva teori fremhever som viktig ved lesing og hva som måles i en kartleggingsprøve i lesing. Det er en undersøkelse med 81 elever fra to skoler. Det er gjort analyser av frekvenser. Bråthen finner forskjell mellom elever med norsk som førstespråk og elever fra språklig minoritet. Forskjellen er størst når det gjelder ordforråd. Bråthen mener at tidsbegrensningene i leseprøven kan ha betydning for en del elevers prestasjoner, og da kan det være vanskeligere å fastslå hvilke pedagogiske tiltak elevene trenger.

Hultin (2017) har en masteroppgave hvor hun undersøker hvordan den digitale kompetansen endrer seg hos elever på 4. trinn fra 2013 til 2016. Det er noen systematiske forandringer i besvarelsene. Det ser ut som oppgaver som inneholder teknisk innhold og oppgaver om e-post har en reduksjon i andel riktige svar. Mens resultatene typer på mer stabilitet i oppgaver om kildekritikk og digital dømmekraft. Ut fra en generell forventning om at elevkullene normalt skårer noe bedre på samme prøve over tid, er funnene her litt oppsiktsvekkende.

Jetne (2017) har en masteroppgave om å lese i samfunnsfag. Han ser på samfunnsfaglig lesing i nasjonale leseprøver for 8. trinn (2007–2016). Det er hentet data fra samfunnsfaglige tekster i NP 2007–2016. Funn tyder på at kjønnsdifferansen mellom gutter og jenter minker noe når prøvetekstene i de nasjonale leseprøvene kan knyttes til samfunnsfag, men at det spesifikke temaet for tekstene fortsatt virker ulikt inn på jentenes og guttenes engasjement og mestring.

Eriksen & Roe (2013). Den nasjonale prøven i lesing på 8. og 9. trinn, 2012. Rapport basert på populasjonsdata. Datagrunnlaget for undersøkelsen er 58 171 elever på 8. trinn og 59 019 elever på 9. trinn i 2011. Analyse av nasjonale prøver i lesing. I rammeverket for leseprøven på 8. trinn er det definert fem kompetansenivåer med konkrete beskrivelser av hva lesere innenfor et visst nivå er forventet å mestre innen de tre lesemåtene. Den gjennomsnittlige p-verdien på hele leseprøven er 62,1 på 8. trinn og 67,9 på 9. trinn. Det vil altså si at en gjennomsnittselev på 8. trinn klarer 62,1 prosent av oppgavene, mens en gjennomsnittselev på 9. trinn klarer 67,9 prosent av oppgavene. I 2010-prøven var p-verdiene 53,2 på 8. trinn og 60,0 på 9. trinn. 2011- prøven er med andre ord lettere for elevene enn prøven året før.

Kort analyse

Lesestrategier er sentrale i leseprøvene og lærerne må kan innsikt i hvordan elevene benytter disse. Forutsetningen for å prestere på leseprøvene er ulike for elever med norsk som førstespråk versus elever som har norsk som andrespråk. Prøver med tidsbegrensning kan ramme sistnevnte gruppe spesielt. Det kan være vanskelig å velge gode pedagogiske tiltak. Valg av tema i lesetekster betyr en del for elevenes prestasjoner og det er kjønnsforskjeller her avhengig av tema. Elevenes digitale ferdigheter virker å være relativt stabile, selv om ferdigheter knyttet til kommunikasjon og tekniske forhold faktisk har hatt en liten nedgang.

Publikasjoner under tema styring

Det er 17 publikasjoner som kan passe inn under tema styring. Det er åtte mastergrader, sju fagfelleverderte publikasjoner og to rapporter.

Når det gjelder styring har vi delt disse inn i kompetanseutvikling (1 publikasjon), skolen som organisasjon (2 publikasjoner), kommune/skole (3 publikasjoner) og ansvar for styring (11 publikasjoner). Vi vil oppsummere temaet styring samlet på slutten av kapitlet.

Kompetanseutvikling

Kompetanse og kompetanseutvikling er sentralt for å få utbytte av prøvearbeidet ved skolene. Lærerne trenger kompetanse både i å tolke prøveresultater, men ikke minst i forbindelse med etterarbeid og tiltak. Temaet er delvis til stedet i flere arbeider, men her er det tatt med et arbeid som har dette litt tydeligere med.

Det er en publikasjon under tema kompetanseutvikling. Brandlistuen Schjølberg, Tambs & Aase (2016) har en rapport om «Læringsmiljø, skoleferdigheter og utviklingsvansker» som del av den norske mor-barn undersøkelsen. Et mål i denne rapporten er å beskrive skoleferdigheter og utviklingsvansker hos elevene som deltar i mor-barn undersøkelsen (1992 8-åringer). Rapporten har med spørreskjemadata fra lærere til 1992 8-åringer som deltar i skolestudien. Brandlistuen mfl (2016) finner at over 90 prosent av lærerne i denne studien svarer at de bruker resultatene fra kartleggingsprøvene i praksis. I den forbindelse påpeker de at dersom lærerne ikke har tilstrekkelig

kompetanse i å velge pålitelige kartleggingsprøver, gjennomføre dem på en valid måte samt å tolke resultatene, kan det få store konsekvenser for enkeltbarn.

Skolen som organisasjon

Organisering og organisasjonsutvikling er noe mange skoler er opptatte av. Ofte er dette på en eller annen måte koblet til kvalitetsarbeid ved skolene. Det er to arbeider under tema skolen som organisasjon.

Solli og Isaksen (2014) skrev sin masteroppgave i skoleledelse om nasjonale prøver i et organisasjonsperspektiv. I et nokså sterkt design intervjuer de skoleeier (politikere og administrasjon), lederteam og lærere. Sentralt i oppgaven er nasjonale prøver som instrument for å utvikle kvalitet på den enkelte skole. De finner at alle ledd i skolen er godt kjent med rammene for nasjonale prøver og hva som forventes. Forfatterne peker på at skolene virker å gjennomføre nasjonale prøver instrumentelt og uten å fullt ut skape en felles forståelse for formålet med prøven og derigjennom arbeider for at dette nås. Det er for eksempel enighet om at elevene som skårer på nivå 1 skal få oppfølging, men svake systemer for å sikre at dette skjer. Selv om det er finnes materiell fra Udirs side for å følge opp resultatene etter nasjonale prøver, er det manglende kunnskap og evne til å nyttiggjøre seg disse på en god måte. Noen lærere etterspør kurs. Det virker som om rammene for å gjennomføre nasjonale prøver ligger greit forankret, men at rutinger og systemer for å løfte prøvene over i en mer konstruktiv posisjon mangler. Det virker å være en utfordring med kontinuitet. På eiersiden byttes politikere jevnlig ut og på lærersiden går det ofte for lang tid mellom hver gang læreren gjennomfører nasjonale prøver (typisk tre år). Flere mener at nasjonale prøver blir en «happening». Isaksen og Solli finner at skolekulturen er i spill i arbeidet med nasjonale prøver og at tidligere praksis står sentral («sånn gjør vi det hos oss»). Når det gjennomføring og spesielt oppfølging blir overlatt til den enkelte lærer blir det vanskelig å sikre god bruk av prøveresultatene. Lærerne opplever også mye testing i perioder.

Johnsen (2010). Hvordan opplever rektor rollen som aktør, i skjæringspunktet mellom ytre krav og indre liv i organisasjonen?: nasjonale prøver i rene ungdomsskoler. Dette er en masteroppgave hvor Johnsen har fokus på hvordan rektor opplever rollen som aktør, i skjæringspunktet mellom ytre krav og indre liv i organisasjonen. Det er intervjuer med tre rektorer i rene ungdomsskoler. Ett av funnene er at rektorene kjenner til mye av situasjonen ved egen skole allerede og at det de trenger er hjelp til å komme i gang med tiltak. Skolelederne i denne undersøkelsen beskriver at fokus på fenomenet nasjonale prøver i sentrale og rurale strøk er ulikt. Skoler som hører til sentrale strøk er ofte mer ansvarliggjort fra det kommunale nivået, mens det ved distriktskoler kan skolelederene oppleve dette i mindre grad.

Kommune/skole

Arbeidet med prøvene utfordrer samarbeidet mellom skoleeier/kommune og den enkelte skole. Selv om flere arbeider berører denne problematikken, har vi her inkludert tre arbeider under kommune/skole.

Aasen & Sandberg (2010). «Hvem vet best? Om styring av grunnsopplæringen under Kunnskapsløftet. Skole, skoleeier og policy». De undersøker virkninger av Kunnskapsløftet som styrereform. De har data fra alle styringsnivå (kvalitativt og kvantitativt). Kunnskapsløftet ble av nasjonale utdanningsmyndigheter introdusert som en reform der forutsetningen for å lykkes innebar et systemskifte i styringen av norsk grunnsopplæring. Deres funn tyder på at «Kunnskapsløftets signaliserte systemskifte ikke virker etter hensikten». Skoleeier opplever ikke lokal handlingsfrihet fordi Kunnskapsløftets fristilling begrenses av sentralt styring gjennom rapportering og tilsyn.

Grytten (2015) har en mastergrad som tar opp hvordan prøveresultatene blir analysert og brukt av skolene. Det er samlet inn data fra rektorer fra skoler i to kommuner. En konklusjon er at rektorene har en klar forståelse av hva nasjonale prøver er og at de er fornøyd med å ha disse prøvene som et verktøy i sin skole. Rektorene er opptatt av kvalitetsutvikling som prosess, og det er negative til at prøveresultatene brukes til sammenligning.

Vassli (2014) skriver sin mastergradsoppgave om hvordan noen kommuner oppnår fremgang på nasjonale prøver. Oppgaven bygger på omfattende empiri av 32 skolesjefer og rektorer i 16 kommuner som har oppnådd fremgang på nasjonale prøver. Vassli trekker frem seks hypoteser som kan forklare fremgang på nasjonale prøver. Hypotesene undersøkes gjennom intervjuer. Informasjon (fra prøvene og andre undersøkelser) og undervisningspraksis er de sentrale faktorene, mens prioriteringer og ressurser spiller mindre rolle. Juks (for eksempel ved å ta ut flere elever enn nødvendig) forklarer ikke endringer. Ukjente faktorer er en egen kategori som det er vanskelig å kommentere. Skoleeiere mener at lederavtaler med klare mål er viktige, mens rektor ikke vektlegger dette. Vassli finner stor forskjell på skolene når det gjelder å forberede elevene. Ved en skole jobbet man med gamle nasjonale prøver intensivt i to perioder før selve prøven, men ved andre skoler var det ingen spesiell forberedelse. Vassli finner også at skolenivået er viktigere enn kommunenivået for å forklare fremganger.

Ansvar for styring

Klare roller og ansvarsfordelinger bør ligge til grunn for arbeidet med ulike prøver. At dette ikke alltid er på plass kan den store interessen for dette temaet tyde på. Det er 11 arbeider som er plassert inn under tema ansvar for styring.

Bonesrønning og Iversen (2008) om suksessfaktorer i grunnskolen med en analyse av nasjonale prøver 2007. Formålet med analysene som presenteres i denne rapporten er å kartlegge suksessfaktorer bak prestasjonene på disse prøvene, på individ-, skole- og kommunenivå. Analysene er primært rettet mot 5. trinn. De bruker primært samleskår på elevers prestasjoner (dvs engelsk, lesning og regning). De finner at kjønn og sosial bakgrunn kan forklare forskjeller. «Det viser seg at enkeltelevers prestasjoner er signifikant høyere på skoler med høyt gjennomsnittlig utdanningsnivå

blant foreldre, og signifikant lavere på skoler med stor gutteandel». Videre undersøker de skolestørrelse, skoletype (1-7 eller 1-10) og kommunestørrelse, men samtidig ser de at det er forskjeller i sosial bakgrunn i forhold til skolestørrelse og kommunestørrelse. Dette kan ha betydning for styring av skole.

Chavannes, Engesveen og Strand (2011) tar for seg nasjonale prøver som grunnlag for skoleutvikling og kontroll. De er opptatt av hvordan nasjonale prøver bidrar til å gi retning i skolens utviklingsarbeid. De bruker data fra et eksisterende forskningsprosjekt (FIRE-prosjektet) (Møller og Ottesen, 2010). De bruker intervjuer med skoleledere fra seks grunnskoler. De nasjonale prøvene har ikke vært et hovedtema i prosjektet, men inngår som en del av undersøkelsen av Kunnskapsløftet som styringsform. De mener at funn fra undersøkelsen «kan tyde på at rektorene vi har intervjuet i liten grad bruker prøvene som grunnlag for skoleutvikling, men at man har mer fokus på rapportering av resultat og igangsetting av tiltak som er lett å måle i forhold til kontrolldimensjonen».

Døsrønningen (2007) er opptatt av hvordan ledere i skolen gjøres ansvarlig for styring av skolen. I denne masteroppgaven er det en sammenligning av to kommuner og deres bruk av vurderingssystemer og strategier for styring. Døsrønningen har intervjuer med skoleeiere fra to kommuner. Resultater fra prøver utgjør sammen med annen statistikk og kvantitative mål en vesentlig del av styringsgrunnlaget i kommunene. Det er ulikt mellom kommunene hva som gjøres med tallene og dermed hvilken verdi de får. Han mener at vurderingssystemene bør så godt som overhodet mulig gi et bredt bilde av elevens faglige og sosiale kompetanse.

Engeland & Langfeldt (2009) tar opp forholdet mellom stat og kommune i styring av norsk utdanningspolitikk 1970-2008. De har en retorisk innrettet fagfellevurdert publikasjon om styring og endring av styringsform over tid. Hva kjennetegner forholdet mellom stat og kommune i styring av norsk utdanningspolitikk 1997 og 2008? De har klart å følge noen kommuners skoleforvaltning med 10 års mellomrom, i 1997 og 2007. De mener at det nasjonale nivået er i ferd med å etablere et styringsgrep på skolesiden som er så omfattende at det mangler sidestykke i moderne tid. Det nye er at dette skjer innen et system hvor inntektssystemet gir kommunene ansvar for undervisningen, og hvor statens mulighet til å bruke økonomiske styringsmidler er redusert.

Gjelsvik (2012) har en masteroppgave med tittel «Kartlegging-og så kva?: Kva blir gjort for elever som kjem under kritisk grense i leseferdigheit i 2. klasse?». Fokus er på bruk av kartleggingsprøver i lesing for elever på 2. trinn. Det er intervju med lærere fra fire skoler. Gjelsvik har valgt ut lærere som hadde erfaring fra prøvene. Gjelsvik ville finne ut meir om kva som skjer etter ei slik kartlegging, og kvar ansvaret for det som evt. skjer ligg. Det er hentet inn lærerar sine erfaringar og tankar kring bruken. Alle lærerne mener at de reflekterer rundt kva tiltak dei skal sette i verk. Gjelsvik mener at upresise planer (f.eks. manglende meldeskjema) for arbeid i etterkant av kartleggingsprøvene kan bidra til tilfeldig oppfølging av elever. Rektorene mener at det er lærernes ansvar å sette i gang tiltak

(med rektor som delansvarlig). Men at rektorene har ansvar for systematisering og forvaltning av ressursene som lærerne trenger.

Grandemo (2017) er opptatt av hva som kjennetegner ledelsens tenkning ved skoler som over tid scorer høyt på nasjonale prøver. Det er intervju med sju skoleledere fra skoler som har jevnt over gode resultater på nasjonale prøver. Analysene tyder på at resultatene på de nasjonale prøvene kommer som en følge av systematisk arbeid over tid og at skolene har et heller avslappet forhold til prøvene. Skolene utvikler seg gjennom kollektive prosesser. Refleksjon og erfaringsdeling er viktig. Skolene, som skårer høyt, er ikke opptatte av å drille i forkant.

Gunnulfsen & Roe (2018) undersøker hvordan lærere oppfatter skoleledelse ved bruk av resultater fra nasjonale prøver på egen skole. Survey med 176 lærere fra ungdomsskolen (flere skoler i en kommune). Resultatene viste at det mellom skolene var ulike oppfatninger av skoleledelsen. Det virker som lærerne har et uavklart forhold til intensjonene som ligger i politikken. Gunnulfsen & Roe konkluderer med at en nær sammenheng mellom rektorenes tilrettelegging for nasjonale prøver og lærernes praksis for å utnytte resultatene.

Holmen & Lyngnes (2016) har forfattet en artikkel om praksis på skoler med gode resultater på nasjonale prøver (Continuous Collective Development as a Road to Success in Primary School). De tar opp hva som kjennetegner praksis på skoler som over tid skårer høyt på prøvene. De har gjennomført intervju med lærere og ledere ved fire skoler for å få frem skolenes egne forklaringer og refleksjoner rundt prestasjoner på prøver. De trekker fram viktighet av kollektiv forståelse, lederskap ved skolene og fokus på lesing.

Hovet & Sætre (2014). Masteroppgave med tittelen «Eg blir både ein overvakar og hjelpar». Tema er knyttet til kartleggingsprøver på vidaregåande skole. Korleis bruker rektor kartleggingsprøver i utøving av styring og leiing, og korleis oppfattar lærarane at prøvene blir brukt? Ein skule, ein rektor og tre lærere er valgt. Hovudfunnet i denne studien er at rektor vert både ein overvakar og hjelpar. Det kan være ei komplisert og utfordrende rolle.

Jenssen (2011). Et skolelederperspektiv på tilpasset opplæring. Jenssen intervjuer utvalgte rektorer om hvordan de tolker sitt oppdrag som ansvarlige for at tilpasset opplæring blir initiert og fulgt opp på skolen. I artikkelen drøftes rektorenes forståelse av tilpasset opplæring, opplevde utfordringer og hvilke handlingsvalg de gjør når de leder tilpasset opplæring ved egen skole. En konklusjon er at rektors handlingsvalg har indirekte innflytelse på praksisen i klassene.

Møller, Ottesen og Hertzberg (2010) analyserer to aspekter (grunnleggende ferdigheter og kvalitetsvurdering) ved Kunnskapsløftet ut fra skolenivå. De har empiri fra seks grunnskoler og fire videregående skoler (maksimal variasjon var strategi for seleksjon av skoler). Deres analyser tyder på at Kunnskapsløftet (som reform) ikke har medført «store endringer i skolenes profesjonsforståelse, slik det er eksemplifisert gjennom beretninger om kvalitetsarbeidet og de grunnleggende ferdighetene» (Møller mfl, 2010). De mener at lærerne mangler tilstrekkelig informasjon til å «oppfatte hvor dypt læreplanmålet om grunnleggende ferdigheter griper inn i de enkelte fagene, eller hvordan det nasjonale systemet for kvalitetsutvikling kan bidra til å styrke lærerarbeidet». Det ser ikke ut som de politiske ambisjonene har nådd ut til lærere og rektorer. Deres funn tyder på at det er ulike spenninger, f.eks. spenning mellom profesjon og politikk, samt spenning mellom sentralisering og desentralisering. Det ser ikke ut til at reformen har bidratt til profesjonelt selvstyre. Dette kan forbedres gjennom å styrke relasjon og sikre dialog mellom de ulike styringsnivåene og den lokale skole.

Kort analyse

Det krever kompetanse og innsikt for å bruke kartleggingsprøvene på en god måte, faren er at enkeltelever ikke får hensiktsmessig hjelp. Det pekes på at uheldig bruk av kartleggingsprøver kan få store konsekvenser for enkeltelever. Skolen har systemer for prøvene, men det er ikke alltid slik at disse følges opp. Dermed kan man ha kunnskap om enkeltelever som ikke mestrer, men uten at det settes inn tiltak. Rektorer opplever å både ha roller knyttet til både kontroll og støtte, noe som kan være komplisert og utfordrende. Det varierer i hvor stor grad rektorene bruker prøvene til utvikling. Rektors rolle er sentral for hvordan elevene blir fulgt opp, men det er interessant å se at skoler som skårer godt på nasjonale prøver likevel kan ha et avslappet forhold til prøvene. Skoler som gjør det godt virker å ha godt lederskap, sterk kollektiv forståelse og klare faglige fokus.

Publikasjoner med tema policy

Det er 19 publikasjoner med tema policy med 12 fagfelleverderte artikler (inkl en artikkel som del av doktorgradsavhandling), to mastergradsoppgaver, fire rapporter og en artikkel fra Bedre skole.

Arbeidene som inngår i denne delen har nokså ulike problemstillinger og vinklinger. Det er ikke uten videre enkelt å kategorisere disse, men vi ser at en del av tekstene er rettet mot det nasjonale kvalitetsvurderingssystemet, mens andre mer direkte retter seg mot nasjonale prøver. Noen tekster passer ikke direkte under noen av disse to overskriftene og er samlet under overskriften «andre temaer».

Nasjonalt kvalitetsvurderingssystem (NKVS)

Allerup mfl (2009) har en rapport om evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnopplæringen. Rapporten vurderer NKVS i et internasjonalt perspektiv, ser på NKVS i nasjonal sammenheng og dessuten på hvordan grunnskolen og vgo forholder seg til NKVS. I denne rapporten tar de for seg nasjonale prøver og kartleggingsprøver. Her ser forfatterne blant annet på effekter av

prøver (f.eks. læring og wash-back). De gjennomgår også forskning på hva internasjonal forskning viser om pedagogisk bruk av tester. Blant mange interessante funn og konklusjoner (rapporten er på 300+ sider), er blant annet behovet for et enklere system verdt å merke seg.

Augedal, PK (2006) Målstyring og desentralisering—en øvelse i motsigelser?: vil det nasjonale kvalitetsvurderingssystemet sikre kvalitet i skolen innenfor dagens styringssystem? Målet med hovedoppgaven er å peke på motsigelser i forbindelse med innføringen av et nasjonalt kvalitetsvurderingssystem samt det samtidige fokuset på desentralisering. Dette er en teoretisk og retorisk masteroppgave. «Ved å drøfte de nevnte begrepene med referanse til ulike forfattere, er målet å bedre kunne forstå hvordan begrepene kan forstås og benyttes fra sentralt hold og hvordan de kan oppfattes av lokale aktører». Augedal mener å finne at det er grunnleggende uforenelighet mellom sentral målstyring og desentralisering. Augedal diskuterer også økonomi vs kvalitet som bakgrunn for styring av skolen. Dessuten at evaluering fungerer som en type styring.

Blömeke og Olsen (2018) har en artikkel som heter «På vei mot et sammenhengende nasjonalt kvalitetsvurderingssystem». De trekker fram seks sentrale betingelser for at målesystemet skal gi god støtte til langsiktig og evidensbasert skole- og politikktutforming: 1) Monitorering på systemnivå, 2) Måling av progresjon, 3) Helhetlig måling av elevers kompetanse, 4) Måling av hva som skjer i klasserommet: Undervisningskvalitet, 5) Kvalitetssikring og validering, 6) Rapportering og kommunikasjon av usikkerhet og endring. Artikkelen presenterer de mest sentrale verktøyene for kvalitetsmonitorering som har blitt innført som en del av det «Nasjonale kvalitetsvurderingssystemet for grunnsopplæringen» (de nasjonale prøvene, kartleggingsprøvene og elevundersøkelsen), i all hovedsak ut fra et perspektiv om at dette er målinger som skal bidra med informasjon til skoler, skoleeiere og nasjonale beslutningstakere.

Langfeldt & Lauvdal (2006) har en rapport om evaluering av det nasjonale kvalitetsvurderingssystemet (NKVS). Rapporten forsøker å presentere hvordan «kvalitetsvurdering beskrives i kommunikasjonen mellom departement og Storting», samt hvordan kvalitetsvurdering er drøftet. De har fulgt og analysert utviklingen fra 1988 og frem til 2004, fordi det er i «denne perioden at den politiske diskurs om et nasjonalt kvalitetsvurderingssystem er tenkt, omsnakket, besluttet og realisert». De viser hvordan det har vært ulike synspunkter på behov for et nasjonalt system for måling, systematisering og rapportering av kvalitet. De viser også hvordan nasjonale prøver inngår som en del av NKVS, og hvordan nasjonale prøver har krevd et omfattende arbeid når det gjelder utviklingsarbeid i ulike fagmiljøer. I Skoleporten kan myndighetene gi informasjon om kvalitet på struktur, prosess og læringsutbytte (inkl nasjonale prøver og kartleggingsprøver). De tror at evalueringen av NKVS kan gi ny kunnskap om gyldigheten av prestasjonsmåling og at den kan gi transparens.

Skov (2010) viser til egne erfaringer som forsker knyttet til evalueringen av NKVS i siste halvdel av 2000-tallet. Han peker på at det i denne perioden er generell enighet mellom kommuner, rektorer og

lærere om målene for nasjonale prøver, men at det er uenighet om hvordan prøvene faktisk brukes. Generelt er skoleeiere (kommuner) og rektorer mer positive til at prøvene har en stor betydning for positiv utvikling på skolen sammenlignet med lærerne. Prøvene brukes ofte for å avdekke elever med svake ferdigheter slik at disse kan følges opp videre, men ut over dette er ikke bruken rettet mot den individuelle eleven i stor grad. Det etterspørres mer informasjon om og fra prøvene for å styrke lærernes muligheter til å nytte prøveresultater i pedagogisk-diagnostisk sammenheng. Videre peker artikkelen på at skolekulturen, her definert som enten individualpraktiserende eller samarbeidene, påvirker hvordan prøvene oppleves. Ved de samarbeidende skolene er lærerne mer positiv enn ved skoler som har mer individualorientert kultur. Skoleeiers involvering er også viktig, det er sammenheng mellom opplevd støtte fra kommunen og opplevelsen av nytte i prøvene.

Nasjonale prøver

Henriksen, Eira, Keskitalo & Øzker (2018) har en artikkel om nasjonale prøver i lesing på samisk, og på hvilke vilkår disse prøvene i samisk for 5., 8. og 9. klassetrinn er utviklet. Forfatterne utdyper hvordan de nasjonale prøvene i lesing på samisk kan være med «å sette samisk språk på dagsorden, både i skolen, blant foreldre og hos skolemyndigheter på ulike nivå». De mener at dette forhåpentligvis kan bidra til å «høyne statusen for samisk som språk og befeste samiske språks stilling som fag og som undervisningsspråk». Det er også forfatterens «oppfatning at de nasjonale prøvene medvirker til å gi samiske lærere en arena hvor de kan drøfte ulike sider ved leseopplæringen».

Huse (2006) har en mastergrad om «Evaluering som resultatevaluering: en diskursanalyse av beslutninger knyttet til innføringen av nasjonale prøver i norsk grunnsopplæring». Huse er opptatt av hvordan beslutninger om nasjonale prøver i det norske utdanningssystemet preges av en hegemonisk diskurs om resultatstyring, og hvordan dekonstrueres en slik betydning gjennom konkurrerende betydninger av evalueringens hensikt og bruk? De sentrale kildene er tekster og intervjuer med personer som er tilknyttet beslutningsprosessen om nasjonale prøver. En konklusjon fra analysen er at forståelsen av nasjonale prøver som sanksjoner ikke framstår som dominerende i beslutningene om nasjonale prøver.

Kavli (2008) har en brukerevaluering av gjennomføring av nasjonale prøver. De inviterte skoleledere, lærere, elever og foreldre (5. og 8. trinn) ved 500 grunnskoler til å delta. Kavli finner at det stort sett er læreren som vurderer prøvebesvarelsene, men at 75% av skolelederne legger opp til samarbeid, primært gjennom personalmøter. Blant lærerne er det mest brukte opplæringstiltaket å ha samtale med eleven om besvarelsen. Dessuten gir ca 25% av lærerne gir skriftlig tilbakemelding. Skolelederne mener at nasjonale prøver kan styrke opplæringen til elevene, og det er ca 50% av elevene som opplever at lærerne gir råd om arbeid med lesing og regning (25% av elevene har ikke fått slike råd). Både skoleledere og lærere ønsker kompetanseheving som følge av de nasjonale prøvene. Skolelederne trekker fram vurdering og pedagogisk oppfølging, mens lærerne trekker fram tilpasset opplæring og fagdidaktikk/metode.

Mausethagen (2013) har en artikkel som inngår i en doktorgradsavhandling "Reshaping teacher professionalism. An analysis of how teachers construct and negotiate professionalism under increasing accountability". Når det gjelder hvordan lærerprofesjonen i Norge fyller 'lærerprofesjonalitet' med innhold, så tar studien utgangspunkt i det økte politiske fokuset på elevresultater og det å undersøke ansvarliggjøring av læreren (accountability). Det er intervjuer med lærere. Mausethagen undersøker om og i hvilken grad slike endringer påvirker læreres oppfatning av profesjonalitet. En konklusjon fra studien kan være at nasjonale prøver bidrar til en økt bevissthet rundt lærerrollen.

Seland, Hovdhaugen og Vibe (2015) har samlet inn data gjennom en spørreundersøkelse og til alle nivå (kommune, skoleleder, lærer) i forvaltningen av nasjonale prøver. Det er også gjennomført intervjuer. Formålet er å se på hvordan prøvene brukes i kvalitetsutviklingsarbeid. Forskerne tar utgangspunkt i OECD sine anbefalinger om hvordan prøvene skulle benyttes som verktøy for kvalitetsutvikling i skolene. Basert på funn, peker forskerne på at skoleleiernivået i varierende grad er engasjerte i arbeidet med nasjonale prøver og der de største kommunene viser størst engasjement. Rektorene mener at det kan være vanskelig å omsette prøveresultater til konkrete endringshandlinger, samtidig som det er akkurat dette som lærerne peker på som nødvendig for å oppleve nytte av prøvene. Lærerne har ønske om å benytte prøvene til pedagogiske formål og forskerne kommer inn på begrensninger i prøveformatet når det gjelder å gi konkrete tilbakemeldinger til den enkelte elev. Utdanningsmyndighetene får litt ros for arbeidet med å utvide prøvenes nytte gjennom ulike tiltak.

Seland, Vibe og Hovdhaugen (2013) har gjort en omfattende datainnsamling med nasjonale prøver som tema. Rapporten er omfattende og bør leses separat. Rapporten viser en rekke interessante funn om hvordan nasjonale prøver fungerer i skolen. Mange av funnene er i tråd med annen forskning, slik som at skolene i varierende grad øver og at skoleledere er mer positive til nytten enn lærere. Det er mange mer forvaltningsmessige perspektiver som tas opp i rapporten som ikke er informativ direkte om bruken, men slike perspektiver kan nok være med å forklare noe om betingelsene prøvene virker under.

Skedsmo og Mausethagen (2015) peker at nasjonale prøver er en del av resultatstyring av skolen og at det er snakk om en dobbel funksjon. På den ene siden skal resultatstyringen gi informasjon om elevenes prestasjoner, mens det på andre også skal gi informasjon om hvordan prestasjonene kan forbedres. Artikkelen gjennomgår norske, empiriske studier på prøvefeltet. De finner blant annet at det mangler kunnskap om hvordan lærere og skoleledere oppfatter de nye styringsformene og at det er en tendens til «for» eller «mot». Det er dessuten ikke mange arbeider som anerkjenner at vurderingsverktøy også er styringsverktøy. I artikkelen kommer forskerne inn på at nasjonale prøver anses som det viktigste elementet i kommunenes og skolenes kvalitetssystemer og at resultatene både brukes til å følge elevers prestasjoner over tid og i forbindelse med mer langsiktige satsninger. Det pekes på at rektorer prioriterer nasjonale prøver på grunn av prestisje ved å gjøre det bra her. På den ene siden brukes prøvene i skolenes forbedringsarbeid, mens det på den andre siden virker å

være mindre integrert formativ praksis enn hva politikere og byråkrater forventer. Det pekes på et behov for økt kompetanse blant lærerne i hvordan prøvene kan bidra til forbedret praksis. Forskerne peker på et økt resultatkrav med detaljerte indikatorer på den ene siden og lærerne som profesjonelle og ansvarlige aktører på den andre.

Smith (2009) skriver om forhold rundt evaluering og vurdering. I artikkelen vektlegges det at prøveresultater er en av flere informasjoner om eleven og at resultatene må brukes i fremtidig planlegging av undervisning. Det er ikke hyppighet av tester eller karakterer som er avgjørende. Forfatteren kommer inn på prøveangst og refererer til andre forskere som har funnet at 20 % av elevene opplever prøveangst i såpass stor grad at det vil påvirke prestasjonen deres. Å utsettes for unødvendig mye prøver vil ikke gavne denne gruppen.

Volckmar (2014) har i kjølvannet av Stortingsvalget i 2013 sett på politisk retorikk slik de fremkommer i partiprogrammene og Sundvollenerklæringen. Det er flere temaer som berøres, blant annet private vs offentlige skoler, forsøk og forskning, like muligheter og Kristendommens plass, men Volckmar kommer også inn på partienes oppfatninger av styringsformer og -instrumenter. Det virker å være nokså tverrpolitisk enighet om innretningen av læreplanen og statusen til nasjonale prøver. Det vil si, Ap, H og Frp ønsker å fortsette med nasjonale prøver som i dag, mens SV, Sp, KrF og R ønsker å gå over til utvalgsprøver. Og mens H og Frp ønsker mer offentliggjøring av prøveresultater, er Ap, SV, Sp, V, KrF og R imot dette og ønsker i stedet at prøvene skal fylle en pedagogisk funksjon. Volmarck peker på at H har et tydelig avtrykk i regjeringserklæringen og at dette burde bety mer styring basert på kartlegging og resultatkontroll.

Andre temaer

Bjørnson (2018) tar opp lenkefeil og ekvivaleringsmetoder på nasjonale prøver. Temaet er interessant i konteksten av utvikling av ferdigheter over tid. For å kunne avgjøre om det er signifikante endringer over tid, så er det «nødvendig å kvantifisere den usikkerheten som er knyttet til prosedyren for lenking fra år til år. Denne usikkerheten betegnes som lenkefeilen». Artikkelen redegjør for ulike måter å forholde seg til lenkefeil i regning og engelsk for 5. og 8. trinn. Konklusjonen er at det er akseptabelt med lenkefeil, men at det er viktig å minimere lenkefeilen gjennom prøvedesign og riktige metoder for analyse.

Gunnulfsen & Möller (2017) undersøker hvordan skoleledere forstår reformer og hvordan de overfører disse reformene til egen skole. De har intervjuer med 12 skoleledere. De trekker fram distinksjonen mellom indre og ytre ansvarlighet, samt at det er ulike mekanismer for hvordan forholde seg til endringer (avvising, avkobling, parallelle strukturer, tilpasning og innkvartering). Deres funn antyder at skolelederne gjennom sine valg og posisjoneringer, ikke ble direkte styrt av den nye politiske konteksten.

Nortvedt og Bucholtz (2018) adresserer flere sentrale temaer i en tematisk bredt anlagt artikkel om vurdering i matematikk. Artikkelen tar ikke for seg Utdanningsdirektoratets prøver direkte, og dermed er litt på siden av temet for denne rapporten. Likevel, artikkelen er interessant fordi den gjennom en omfattende review av eksisterende forskning berører flere temaer som også er relevante for prøvefeltet. Forskerne peker på at studenter, lærere, policy-makers og også forskere kan ha en intuitiv oppfatning om at vurderingsformene vi har er objektive og valid og at en enkelt test eller observasjon kan gi et riktig eller sant bilde av studenters, læreres eller om så hele utdanningssystemets prestasjoner. Det er et sentralt poeng at ikke all bruk av prøvene følger av formålet de ble utviklet for. Forfatterne tar opp at forskning på undervisning i matematikk er et relativt sett nytt forskningsområde og setter dette i konteksten av vurdering og problemstillinger her. For eksempel peker de på at det er mye vanskeligere å måle problemløsning som ferdighet kontra ferdigheter i regning. De finner også at prøver i matematikk ofte ikke måler mer sammensatte kompetanser. Det er en rekke problemstillinger knyttet til vurdering og testing i matematikk. Ofte skal prøvene gi informasjon til flere nivåer samtidig. Dette er problematisk. På samme måte som lærere har også politikere og andre policy-makere ofte problemer med å tolke prøveresultater. Knyttet til policy ligger også spenningen mellom kontroll og støtte, forskerne finner at lærerne kan føle seg kontrollert. Hvis skoleledere og lærere sammen tolker og bruker prøveresultatene, kan dette ha positiv effekt for elevenes prestasjoner. Policy-perspektiver knyttet til ulike prøver har innvirkning på undervisningen, for eksempel gjennom effekter som «teaching for tests» og ved at innholdet i undervisningen styres av prøvene. Artikkelen berører også temaer som like muligheter (equity) og ulike perspektiver knyttet til validitet. Forskerne peker på fire sentrale områder for å utvikle vurderingen i matematikk: Utvikle det metodiske og høyne kvaliteten i prøvene («assessment»), bedre sammenhengen mellom forskning og praksis, bedre likeverdighet («equity») og benytte teknologi for å utvikle bedre mål på matematisk kompetanse.

Olsen, Hopfenbeck og Lillejord (2013) gjennomførte en syntese av tidligere artikler og rapporter om elevenes lærings situasjon etter innføringen av Kunnskapsløftet. Utover å kommentere utviklingen for en de mest sentrale internasjonale prøvene Norge deltar i, kommer forskerne inn på flere interessante temaer. I forbindelse med de nasjonale satsingene på vurdering for læring viser en av undersøkelsen av det ikke var signifikante forskjeller på resultater fra nasjonale prøver mellom skoler som deltok i satsingen Vurdering for læring og de som ikke gjorde dette. Samtidig viser undersøkelser at vurdering har blitt viktig satsing lokalt. Det vises til forskning som antyder at lærere opplever en forbedret vurderingspraksis, selv om det også er en bekymring for at testing kan føre til vurdering for vurderingens skyld. Rapporten kommer inn på flere områder som rører ved elevenes lærings situasjon, slik som læringsmiljø, men disse er ikke klart knyttet til prøvefeltet. Likevel, det er interessant at forskerne peker på at vi vet mer om skolens innhold nå enn bare for noen år siden og at vurderinger av effekten K06 og reformer krever tid.

Tveit (2018) peker på internasjonal forskning (e.g. som definerer tre formål med «educational assessment»: *Selection and certification*, *Determining and raising standards*, og *Formative assessment – Assessment for learning*). Artikkelen sammenligner policy-dokumenter i Norge og Sverige over de siste to tiårene. Prøvesystemene i de to landene har begge som mål å gi nasjonal

styringsinformasjon og samtidig fylle en formativ vurderingsrolle i skolen. I Sverige har prøvene også en sertifiserende rolle som man i Norge ikke har gått inn på. Generelt har artikkelen en oversiktlig gjennomgang av en del relevante rapporter og «white-papers».

Kort analyse

Nasjonale prøver utfordrer forholdet mellom sentralisert og desentralisert ansvar og styring. Prøvene er på den ene siden vurderingsverktøy, men på den andre også styringsverktøy. Det er generelt politisk enighet om NKVS, samtidig som det er litt ulike syn for eksempel på om nasjonale prøver skal være obligatorisk for alle eller endres til en utvalgsprøve. Skoleeiere og -ledere er mer positive til nasjonale prøver enn lærerne. Hvis skoleeier er engasjert og støtter skolene øker opplevelsen av nytte av prøvene. Lærere ønsker å dra nytte av prøvene i pedagogisk sammenheng, men trenger hjelp til dette. I denne sammenheng trekkes det frem at Utdanningsdirektoratets økte satsing på dette er bra. For samiske skoler har nasjonale prøver gitt legitimitet til samisk, også som undervisningsspråk, men at de også skaper et grunnlag for faglige diskusjoner.

Publikasjoner med andre metode/fag

Forskere bruker prøvene som utgangspunkt for å vurdere ulike sider ved eget fagområde, for eksempel for å se på hvilke strategier elever benytter for å løse oppgaver eller for å vurdere hvilke ferdigheter elevene trenger. Ved å koble teoretiske og metodiske perspektiver til prøvene, får forskerne ofte frem gode faglige og metodiske momenter. Det er 11 publikasjoner innen tema metode/fag. Det er 7 fagfelleverderte publikasjoner, 2 masteroppgaver og 2 rapporter.

Ottem, Frost & Veia (2014) har undersøkt sammenhenger mellom prøver som kartlegger ferdigheter i språk og lesing med krav til relasjonell tenkning. Forskerne benytter Raven's Standard Progressive Matrices (SPM) som et mål på intelligens og korrelerer så denne mot ulike egenskaper og prøver. Logisk bør SPM korrelere høyt med prøver som krever høy relasjonell tenkning. Forskerne finner at ulike prøver lar seg rangere basert på krav til relasjonell prosessering. De finner at nasjonale prøver i lesing krever høy grad av relasjonell tenkning sammenlignet med andre leseprøver. Dette betyr at elever kan skåre bra på lesetester og samtidig dårlig på nasjonale prøver. Relasjonell tenkning er også korrelert med akademisk selvoppfatning, vokabular og setningsminne.

Engen, Begnum, Heyen & Solheim (2007) har en rapport om leseferdighet på 2. årstrinn. De har data fra Lesesenterets oppfølgingsundersøkelsen av lesing på 2. årstrinn skoleåret 2006/2007 (1551 elever fra 2007). Disse er sammenholdt med tilsvarende data fra fjorårets undersøkelse (Engen med flere 2006b). Resultater fra standardiseringen våren 2000 (jf Læringscenteret 2001) er benyttet som referanse. Våren 2007 er prosentandelen som skårer på/under bekymringsgrensene betydelig lavere enn ved standardiseringen i 2000 og ved forrige års undersøkelse. Samtidig viser de en økning i prosentandelen elever som klarer alt.

Høie (2011) tar for seg skolevandring i en kommune. Høie stiller spørsmål om praktisering av Skolevandring fører «til endring i læringstrykk og synliggjøres dette i resultatene på nasjonale prøver?». Dette er en masteroppgave med fokus på 5. trinn (barneskole). Det viser til nasjonale prøver i 2007-2010 i flere fag. Det er intervju med lærere/rektor, samt analyse av nasjonale prøver. Funnene viser at en overvekt av rektorene i kommunen og alle de intervjuede lærerne oppfatter at læringstrykket øker gjennom Skolevandring. For de aktuelle skolene viser læringsresultatene på nasjonale prøver at det er en liten eller ingen endring i forhold til landsgjennomsnittet.

Lie mfl. (2005). Dette er et analyse av de første rundene av de nasjonale prøvene. Det er rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005. De mener at det er en del svakheter ved prøvene og de anbefaler at prøvene settes på vent slik at forarbeid og gjennomføring av prøvene kan kvalitetssikres bedre.

Midtbø, Rossow & Sagbakken(2018) har en artikkel om måling av sensorreliabilitet ved vurdering av norskprøve i skriftlig framstilling. De undersøker informasjon om voksne innvandrere som tar norskprøven. Data er hentet fra 77 sensorer sin vurdering av ca 7600 besvarelser. Many-Facets Rasch Measurement (MFRM) er brukt for å analysere sensorkorpsets reliabilitet ved sensurering av Norskprøvens delprøve i skriftlig framstilling. En konklusjon er at det er klare forskjeller i strenghet mellom sensor.

I en argumenterende artikkel tar Nortvedt (2006) opp noen sentrale problemstillinger som gjorde seg gjeldende i forbindelse med oppstarten av nasjonale prøver i regning etter at prøvene hadde vært tatt av etter de to første gjennomføringene. Nortvedt poengterer at prøvenes innretning dreier fra å skulle måle matematisk kompetanse til å rette seg mer spesifikt mot grunnleggende ferdigheter i regning. Hun peker på at oppgavetyper vil måtte endres, men også at veiledningsmaterialet må inkludere vurderingsaspektet bedre

Roe og Vagle (2013) undersøkte kjønnsforskjeller i lesing basert på nasjonale prøver i perioden 2007-2011. Jentene gjør det generelt en del bedre enn guttene og i denne perioden har forskjellene vært stabile. Det er imidlertid forskjeller mellom ulike oppgavetyper. Åpne oppgaver og oppgaver som krever kritisk vurdering eller refleksjon viser større kjønnsforskjeller enn avkryssingsoppgaver. Oppgaver som krever generell forståelse, tolkning og informasjonsinnhenting viser igjen mindre forskjeller mellom kjønnene. Forskerne diskuterer også ulike tekstformer og ser hvordan gutter og jenter presterer på disse. For eksempel, skjønnlitterære tekster og kunnskapsformidlende tekster gir tydelige kjønnsforskjeller, det samme gjelder argumenterende tekster.

Skar og Aasen (2018) har en artikkel hvor de gjennomgår og diskuterer utfordringer med å utvikle læringsstøttende skriveprøver for skolen. De beskriver kjennetegn ved utvikling av prøvene, og redegjør for hva de har gjort for å dokumentere kvalitet på prøvene. Å måle skriveferdigheter er svært komplisert. Kulturelle og kontekstuelle forhold er involvert. Forfatterne peker på at undersøkelser på av reliabilitet på eksamen i skrijving har vist svakheter.

Stensen (2005) skrev i en tidlig hovedoppgave om den nasjonale leseprøven på 10. og 11. trinn. Hun finner tar utgangspunkt i et teoretisk perspektiv som vektlegger at lesing handler både om å hente ut og også konstruere mening i teksten. Denne forståelsen finner hun at sammenfaller med innretningen av den nasjonale prøven. Hun peker imidlertid på at det er mindre klart hvilke kognitive prosesser som kreves for å løse oppgavene og at dette er uheldig for det videre arbeidet.

Tonne og Pihl (2013) finner at skåren i nasjonale prøver i lesing på 5. trinn er avhengig av elevenes morfologiske bevissthet og grep om spesielle typer vokabularer. De finner dette ved å korrelere nasjonale prøver i lesing på 5. trinn med kartleggingsprøvene på 3. og 5. trinn. Forskerne peker på at nasjonale prøver er ment å måle leseforståelse, men at det da altså er andre faktorer involvert. De peker på at dette fenomenet er skaper validitetsproblemer når det er andrespråkselever som skal vurderes, disse har ofte lav språklig bevissthet. Artikkelen kommer med pedagogiske anbefalinger i arbeidet med å utvikle andrespråkselevers språklige bevissthet.

Artikkelen «Kartleggingsprøver i lesing – tid for nytenking?» (Walgermo, Uppstad, Lundetræ, Tønnessen, & Solheim, 2018) bruker analyser av kartleggingsprøvene i lesing for å peke på mulige forbedringer i disse. Innledningsvis peker forskerne på fem særlige utfordringer med kartleggingsprøvene: 1) prøvens omfang, 2) utilsiktede konsekvenser for undervisningspraksis, 3) feil bruk av prøveresultatene, 4) ulik progresjon i begynneropplæringen og 5) høy terskel for bruk av prøveresultatene. Artikkelen gjennomgår alle disse fem utfordringene grundig. Det empiriske materialet består av et utvalg elever som har gjennomført kartleggingsprøven på 1. og 3. trinn. Interessante funn viser at en enkel setningsleseprøve ved slutten av 1. klasse klart predikerer vansker med leseforståelse i slutten av 3. klasse. På bakgrunn av dette viser forskerne at et prøveformat med en kortfattet innledende del (screening-del) og en mer utfyllende del er problemene tydeligere dokumenteres, kan være en mulighet. Forskerne diskuterer også funnene og dagens prøve i lys av nyere forskning om lesing.

Kort analyse

Det er stor bredde i arbeidene som tar for seg metodiske eller faglige problemstillinger eller temaer. Det er vanskelig å oppsummere arbeidene siden de spriker såpass mye i tematikk og innretning. Et arbeide peker på sammenhengen mellom relasjonell tenkning og prestasjon på leseprøver, mens en annen ser på hvordan elevenes skår på leseprøven har endret seg over tid. Andre ser på prøveformer og -formater. Likevel, disse arbeidene er sentrale og gir dybdeinnsikter som kan være svært nyttige i arbeidet med prøvefeltet generelt og ikke minst i utviklingen av konstruerte og faktiske prøver.

Oppsummering og avslutning

Utgangspunkt for dette arbeidet har vært å identifisere og beskrive vitenskapelige arbeider på prøvefeltet innenfor utvalgte temaområder. Vi har arbeidet ut fra tre kriterier for inkludering av vitenskapelige arbeider. Det dreier seg om publikasjoner som omfatter 1) bruk av prøvene i skolen for å forbedre praksis, 2) bruk av prøvene som styringsverktøy, og 3) prøvenes funksjon i politikktutforming og til andre policyformål. Vi har utelatt publikasjoner som primært er utformet for å beskrive og dokumentere tekniske spesifikasjoner ved prøvene.

Gjennom manuelle søk etter aktuelle publikasjoner og gjennomlesning av abstrakt satt vi igjen med 120 publikasjoner hvor vi ønsket å lese gjennom fulltekst. Vi forsøkte å hente inn digitale versjon av disse 120 publikasjonene for å lese de i fulltekst. Det var noen arbeider vi ikke fikk tilgang til digitalt, og disse valgte vi å utelate. Under gjennomlesning av fulltekst viste det seg at noen publikasjoner ikke dreide seg tilstrekkelig om de tre temaene for inkludering. Disse ble publikasjonene ble derfor ekskludert. Etter gjennomlesning gjenstod det 81 aktuelle publikasjoner som passer med kriterier for inkludering og ekskludering.

Om publikasjonene

De fleste publikasjonene har fokus på nasjonale prøver. Det er kun noen få publikasjoner som tar opp kartleggingsprøver eller læringsstøttende prøver. De nasjonale prøvene er obligatoriske, men det har også noen av kartleggingsprøvene vært. En forklaring på at nasjonale prøver er dominerende kan være at de har vært ganske stabile siden 2006, mens kartleggingsprøver og læringsstøttende prøver har mer kommet og gått.

I et flertall av publikasjonene (47 av 81 publikasjoner) er det samlet inn og analysert egne data for å belyse problemstillinger knyttet til prøver. Det er interessant at de nasjonale prøvene har skapt grunnlag for ytterligere forskning og utforming av spesifikke forskningsdesign. Et mindretall av publikasjonene (23 av 80 publikasjoner) baserer seg kun på gjenbruk av prøvedata.

Når det gjelder de tre temaene, så er det ikke gjensidig utelukkende begreper og det kan være en viss overlapp. Gjennomgang viser at det er flere publikasjoner innen tema skolen som praksis (34 arbeider) sammenlignet med publikasjoner på tema som policy (19 arbeider) eller styring (17 arbeider). Det kan skyldes en interesse for å undersøke hvordan prøvene brukes av lærere i arbeid med skolefagene.

Blant de 81 publikasjonene er det 37 fagfelleverderte arbeider og 29 mastergradsoppgaver. Det er litt flere mastergrader enn fagfelleverderte publikasjoner under tema skole som praksis (hhv. 17 og 12) og styring (hhv. 8 og 7), mens vi har plassert få mastergrader under tema som policy og fag/metode.

Policy, styring og utvikling av undervisning

Det er flere publikasjoner (Allerup mfl., 2009, Blöemecke & Olsen, 2018; Skov, 2010) som knytter de nasjonale prøvene til NKVS (nasjonalt kvalitetsvurderingssystem). De nasjonale prøvene utgjør en liten del av NKVS. Publikasjonen viser hvordan prøvene kan ha ulike funksjoner for skolen. De

nasjonale prøver kan legitimere fag (Henriksen mfl., 2018) og bidra til profesjonsforståelse hos lærere (Mausethagen, 2013), men samtidig utfordrer de nasjonale prøvene det desentraliserte ansvar som er gitt til skoler og lærerne (Berger, 2008; Eriksen, 2011).

Berger (2008) er opptatt av at prøvene har en dobbeltfunksjon knyttet til vurdering for og av læring. På den ene siden har prøvene en formativ funksjon, fordi prøveresultatene gir informasjon til lærerne om enkeltelever. På den andre siden skolens resultater offentliggjort, og dermed kan resultatene fremstå som summative for omgivelsene. Gjennom offentliggjøring av informasjon om prøveresultatene blir det mulig å sammenligne resultater på nasjonale prøver mellom skoler og mellom kommuner (Eriksen, 2011; Gunnulfsen, 2017).

Det er noen publikasjoner som ser på prøver i enkeltfag (Henriksen mfl, 2018) eller bestemte ferdigheter som regning (Nortvedt, 2018), lesing (Tavas, 2006), skriving (Skar og Aasen, 2018) og digitale ferdigheter (Hultin, 2016), men det er også publikasjoner som analyserer og diskuterer hvordan lærere og skoleledere forholder seg til nasjonale prøver (Brandlistuen mfl., 2016; Mausethagen, 2013;), uten å konkretisere til fag eller ferdighet.

Det er ikke entydige funn knyttet til hvordan lærere og skoleledere oppfatter prøvene. Det kan synes som skoleledere/-eiere er mer positiv til prøvene enn lærerne (Gunnulfsen, 2017). Felles for både skoleledere og lærere er at de kan oppleve et spenningsforhold knyttet til kontroll på den ene siden og utvikling på den andre siden (Berger, 2008; Chavannes mfl., 2011; Nortvedt & Buchholz, 2018; Skedsmo & Mausethagen, 2015; Volckmar, 2014). For skolelederne kan det dreie seg om at de både skal utvikle skolens lærere samtidig som de har ansvar for at skolen følger de regler og krav fra myndigheter og skoleeier.

For lærerne kan det være et spenningsforhold mellom det at alle elever skal ta prøvene og det at lærerne har ansvar for å følge opp elevene og deres faglig utvikling. Lærerne har frihet til hvordan de vil gjøre forarbeid til prøvene (Nes, 2013) og etterarbeid med prøveresultatene (Eriksen, 2011). Det er stort sett lærerne som vurderer elevresultatene fra prøvene, mens skoleleder forsøker å sikre samarbeid mellom lærere blant annet gjennom personalmøter (Kavli, 2008).

Flere publikasjoner påpeker at lærerne i etterkant av prøvene ønsker å bruke prøveresultatene i sin oppfølging av elevene. Det kan være vanskelig for lærerne å bruke prøvene aktivt i utviklingsarbeidet (Eriksen, 2011). Eksempelvis etterlyser Brandlistuen mfl (2016) hvilke kompetanse lærerne har fått til å identifisere og iverksette gyldige oppfølging av elever basert på analyser av prøveresultatene. De understreker betydning av kompetanseutvikling som sentralt tema knyttet til bruk av prøver i skolen (Brandlistuen mfl, 2016). Kavli (2008) finner at lærerne ønsker kompetanse innen tilpasset opplæring og fagdidaktikk/metode.

Skedsmo og Mausethagen (2015) påpeker behov for økt kompetanse blant lærerne i hvordan prøvene kan bidra til forbedret praksis. Dette gjelder både lesing og regning. Tavas finner at lærerne mangler kompetanse til å følge opp elevene i den videre leseopplæringen. Nortvedt og Rohatgi (2018) påpeker mangel på ressurser i skolen til oppfølging, samt lærerens fagdidaktiske kompetanse i regning. Ulike kompetanse fører til forskjeller i lærernes vurderingspraksis (Nortvedt mfl., 2016).

Litteraturliste

- Aasen, P., & Sandberg, N. (2010). Hvem vet best? Om styringen av grunnopplæringen under Kunnskapsløftet. *Acta Didactica Norge*, 4(1), (Art. 18, 20 sider).
- Allerup, P., Kovac, V., Kvåle, G., Langfeldt, G., & Skov, P. (2009). Evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnopplæringen. *FoU Rapport*, 8, 2009.
- Almaas, T. S. (2017). *Hvordan brukes elevresultater til læring i skolen?* ,
- Almestrand, K. (2012). *Hva gjør de gode leserne?: En studie av elevers strategiske lesing av sammensatte tekster i nasjonale prøver på slutten av barneskolen.*
- Augedal, P. K. (2006). *Målstyring og desentralisering—en øvelse i motsigelser?: vil det nasjonale kvalitetsvurderingssystemet sikre kvalitet i skolen innenfor dagens styringssystem?* (Master), University of Oslo, Oslo.
- Bakken, A., & Elstad, J. I. (2012). For store forventninger. *Kunnskapsløftet og ulikhetene i grunnskolekarakterer*, 7, 2012.
- Berge, K. L. (2005). Skrivning som grunnleggende ferdighet og som nasjonal prøve—ideologi og strategier. *Det nye norskfaget*, 161-188.
- Berger, M.-O. (2008). *Nasjonale prøver 2007: arbeidet skolene etter intensjonen?: hvordan takler skolene presset?* ,
- Björnsson, J. K. (2018). Om lenkefeil og ekvivaleringsmetoder på nasjonale prøver: Evaluering av endring over tid. *Acta Didactica Norge*, 12(4), Art. 16, 24 sider.
- Blömeke, S., & Olsen, R. V. (2018). På vei mot et sammenhengende nasjonalt kvalitetsvurderingssystem. *Acta Didactica Norge*, 12(4), Art. 1, 21 sider.
- Bonesrønning, H., & Iversen, J. V. (2008). Suksessfaktorer i grunnskolen: Analyse av nasjonale prøver 2007. *Trondheim, Norway: Senter for økonomisk forskning.*
- Brandlistuen, R. E., Schjølberg, S., Tambs, K., & Aase, H. (2016). Læringsmiljø, skoleferdigheter og utviklingsvansker.
- Bråthen, A. (2011). *Hva kjennetegner 2. klassingers leseferdigheter slik dette blir kartlagt i nasjonale kartleggingsprøver?: en analyse.*
- Chavannes, I. W., Engesveen, H., & Strand, E.-M. (2011). *Nasjonale prøver som grunnlag for skoleutvikling og kontroll.*
- Døsrønningen, S. (2007). *Hvordan ansvarliggjøres ledere i skolen?: en sammenligning av to kommuner og deres bruk av vurderingssystemer og strategier for styring.*
- Engeland, Ø., & Langfeldt, G. (2009). Forholdet mellom stat og kommune i styring av norsk utdanningspolitikk 1970–2008. *Acta Didactica Norge*, 3(1), (Art. 9, 16 sider).
- Engen, L., Begnum, A. C., Heyen, T., & Solheim, R. G. (2007). Leseferdighet på 2. årstrinn, våren 2007. In: Delrapport, Nasjonalt senter for leseopplæring og leseforskning

- Eriksen, A., & Roe, A. (2013). Den nasjonale prøven i lesing på 8. og 9. trinn, 2012. *Rapport basert på populasjonsdata*. Oslo: Institutt for lærerutdanning og skoleforskning. Universitetet i Oslo.
- Eriksen, G. V. (2011). *Hvordan har nasjonale prøver påvirket til endring av praksis på to norske skoler?: skoleledelse i endringstid*.
- Evensen, L. S., Berge, K. L., Thygesen, R., Matre, S., & Solheim, R. (2016). Standards as a tool for teaching and assessing cross-curricular writing. *The Curriculum Journal*, 27(2), 229-245. doi:10.1080/09585176.2015.1134338
- Falk, E. (2006). *Nasjonal prøve i IMX våren 2005: et lærerperspektiv*. Høgskolen i Agder,
- Gjelsvik, J. Ø. (2012). *Kartlegging-og så kva?: Kva blir gjort for elever som kjem under kritisk grense i leseferdigheit i 2. klasse?* ,
- Grandemo, L. I. Hva kjennetegner ledelsens tenkning ved skoler som over tid scorer høyt på nasjonale prøver? *Norsk pedagogisk tidsskrift*, 101(01), 19-30.
- Grant & Booth (2009). A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26, pp.91–108
- Grytten, K. (2015). *Korleis erfarer eit utval rektorar i Sogn og Fjordane bruken av nasjonale prøver?* , The University of Bergen,
- Hatlevik, O. E., & Radtke, I. (2017). Standard Setting in a Formative Assessment of Digital Responsibility Among Norwegian Eighth Graders. In *Standard Setting in Education* (pp. 205-223): Springer.
- Henriksen, M. B., Eira, K. I., Keskitalo, J. H., & Øzerk, K. (2018). Nasjonale prøver i lesing på samisk-på hvilke vilkår? *Acta Didactica Norge*, 12(4), Art. 4, 20 sider.
- Holmen, H., & Lyngsnes, K. (2016). Continuous Collective Development as a Road to Success in Primary School. *International Journal of Learning, Teaching and Educational Research*, 14(2).
- Holthe, M. A. (2007). *Ordforråd og syntaktisk kompleksitet i argumenterende elevtekster: en studie av et utvalg tekster skrevet av tiendeklassinger til den nasjonale prøven i skriving våren 2005*.
- Hovet, K. H. & Sætre, I.R. (2014). " Eg blir både ein overvakar og hjelpar" *Kartleggingsprøver som verkty for styring og leiing*.
- Hultin, H. (2017). *Endring i digital kompetanse hos elever på småskoletrinnet*. Høgskolen i Oslo og Akershus, Oslo.
- Huse, H. B. (2006). *Evaluering som resultatvaluering: en diskursanalyse av beslutninger knyttet til innføringen av nasjonale prøver i norsk grunnopplæring*.
- Høie, Y. N. (2011). *Skolevandring i en kommune: Fører praktisering av Skolevandring til endring i læringstrykk og synliggjøres dette i resultatene på nasjonale prøver?* ,
- Høihilder, E. K., & Engh, K. R. (2006). Elevvurdering på grunnskolens sjuende trinn i Vestfold. In: Tønsberg: Høgskolen i Vestfold.

- Jenssen, E. S. (2011). Et skolelederperspektiv på tilpasset opplæring. *Norsk pedagogisk tidsskrift*, 95(03), 172-184.
- Jetne, Ø. (2017). *Å lese i samfunnsfag. Samfunnsfaglig lesing i nasjonale leseprøver for 8. trinn (2007–2016)*.
- Johansen, M. A. (2017). *Nasjonale prøver i lesing. Hvordan tar lærerne resultatene i bruk?*, UiT Norges arktiske universitet,
- Johnsen, H. (2010). *Hvordan opplever rektor rollen som aktør, i skjæringspunktet mellom ytre krav og indre liv i organisasjonen?: nasjonale prøver i rene ungdomsskoler*.
- Krumsvik, R.J. (2016). *En doktorgradsutdanning i endring*. Bergen: Fagbokforlaget
- Landmark, E. (2018). *Nasjonale prøver som kvalitativt styringsverktøy for alle elever-et systembidrag til kvalitet i opplæringen*.
- Langfeldt, G., & Lauvdal, T. (2006). Evaluering av det nasjonale kvalitetsvurderingssystemet. *Delrapport pr. 1*, 2006.
- Lie, S., Hopfenbeck, T. N., Ibsen, E., & Turmo, A. (2005). *Nasjonale prøver på ny prøve: rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005*. Oslo: Universitetet i Oslo (Hentet herfra <https://www.duo.uio.no/handle/10852/32300>)
- Ljønes, G. K. H. (2012). *Nasjonale prøver, hva så?: Undersøkelse av 5. trinnelevers lesekompetanse i lys av nasjonal leseprøve*. Oslo: Universitetet i Oslo
- Mausethagen, S. (2013). *Reshaping teacher professionalism. An analysis of how teachers construct and negotiate professionalism under increasing accountability*: Høgskolen i Oslo og Akershus.
- Midtbø, T., Rossow, A., & Sagbakken, B. (2018). Måling av sensorreliabilitet ved vurdering av norskprøve i skriftlig framstilling. *Acta Didactica Norge*, 12(4), Art. 12, 25 sider.
- Myhre, S. A. (2014). *Nasjonale prøver i regning: en kvalitativ studie av hvordan nasjonale prøver i regning blir brukt i skolen* NTNU, Trondheim.
- Møller, J., Ottesen, E., & Hertzberg, F. (2010). Møtet mellom skolens profesjonsforståelse og Kunnskapsløftet som styringsreform. *Acta Didactica Norge*, 4(1), (Art. 15, 23 sider).
- Nes, H. (2013). *Nasjonale prøver på mellomtrinnet: Lærernes elevforberedelser i et accountability-perspektiv*.
- Nilsen, R.L. (2014). *Tilpasset opplæring i lesing. En kvalitativ studie av hvordan kartleggingsprøven i lesing påvirker pedagogisk praksis i en videregående skole* (Master's thesis, UiT Norges arktiske universitet).
- Nortvedt, G. A. (2006). Kompetanse i matematikk kontra grunnleggende ferdigheter i å kunne regne—Nasjonale prøver i ny støpeskje? *Norsk pedagogisk tidsskrift*, 90(05), 382-390.
- Nortvedt, G. A. (2011). Coping strategies applied to comprehend multistep arithmetic word problems by students with above-average numeracy skills and below-average reading skills. *The Journal of Mathematical Behavior*, 30(3), 255-269.

- 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*, 12(4), Art-8.
- Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: responding to issues regarding methodology, policy, and equity. *ZDM*, 50(4), 555-570.
- Nortvedt, G. A., & Rohatgi, A. (2018). Rapport fra skolelederundersøkelsen. Om Osloskolenes bruk av kartleggingsprøver i regning for småtrinnet.
- 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.
- OECD (1989). Ekspertvurdering fra OECD. OECD-vurdering av norsk utdanningspolitikk. Oslo: Aschehoug.
- Olsen, R. V., Hopfenbeck, T. N., & Lillejord, S. (2013). Elevenes lærings situasjon etter Kunnskapsløftet. *Norsk pedagogisk tidsskrift*, 97(06), 355-369.
- Ottem, E., Frost, J., & Veia, G. (2014). Kartlegging av elevers ferdigheter i språk og lesing- underliggende forskjeller i prøvenes krav til relasjonell tenkning. *Psykologi i kommunen*, 1, 9-16.
- Rangnes, H., & Gourvenec, A. F. (2018). Læreres samtale om pedagogisk bruk av flerspråklige elevers prøveresultater. *Acta Didactica Norge*, 12(4), Art. 6, 21 sider.
- Ravlo, G. (2009). Vurdering med nasjonale prøver. *Tangenten*, 1.
- Roe, A., & Vagle, W. (2013). Kjønnsforskjeller i lesing – et dybdedykk i resultatene fra nasjonale prøver på åttende trinn fra 2007 til 2011. *Norsk pedagogisk tidsskrift*, 96(6), 425-441.
- Sandset, E. (2015). *Elevers resonneringer og refleksjoner i arbeidet med oppgaver fra nasjonale prøver 2014-en høyttenkningsstudie*.
- Seland, I., Hovdhaugen, E., & Vibe, N. (2015). Mellom resultatstyring og profesjonsverdier. *Nordisk Administrativt Tidsskrift*, 3(92), 44-59.
- Seland, I., Vibe, N., & Hovdhaugen, E. J. H. f. h. b. b. n. x. h. (2013). *Evaluerings av nasjonale prøver som system (NIFU rapport 4/2013)*. Retrieved from
- Skar, G. B. U., & Aasen, A. J. (2018). Å måle skriveferdighet som grunnleggende ferdighet. *Acta Didactica Norge*, 12(4), Art. 10, 29 sider.
- Skedsmo, G., & Mausesthaugen, S. (2015). Spenningen mellom kontroll og læring. Bruk av nye vurderingsverktøy i utdanningssektoren. *Bedre skole*, 16-213.
- Skov, P. (2010). Bruken av nasjonale prøver: en evaluering. *Bedre skole*(2), 84-89.
- Smith, K. (2009). Grisen blir ikke feitere jo mer vi veier den. *Tangenten: Tidsskrift for matematikk i grunnskolen*, 1(1), 38.
- Solli, A.-M. H. & Isaksen, T. H. (2014). «Resultatene er kommet! Hva nå?» *Nasjonale prøver-fra fokus på resultat til utvikling av kvalitet*. UiT Norges arktiske universitet,

- Stensen, M. H. (2005). *Måling av leseforståelse i den nasjonale leseprøven for 10. og 11. trinn: en teoretisk drøfting av hvilke aspekter ved begrepet leseforståelse som måles i den nasjonale prøven i lesing for 10. klasse og grunnkurs for videregående opplæring 2005*.
- Stornes, T., Tvedt, M. S., & Bru, E. (2013). Best med test?—Et motivasjonsteoretisk perspektiv på bruk av normative tester i grunnskolen. *Norsk pedagogisk tidsskrift*, 97(04-05), 315-325.
- St. Meld. 30. (2003–04). Kultur for læring. Oslo: Utdannings- og forskningsdepartementet.
- Strøm, K. (2006). *Nasjonale prøver i matematikk-til hvilken nytte?: en studie av de nasjonale prøvene i matematikk for 7. klasse våren 2005*. (Master), University of Oslo, Oslo.
- Tavas, T. (2006). *Den nasjonale leseprøven på godt og vondt: Om den viderekomne lesekompetansen og den nasjonale prøven i lesing på ungdomstrinnet*.
- Tonne, I., & Pihl, J. (2013). Andrespråkseleven, leseprøver og litteraturbasert leseopplæring.
- Tveit, S. (2018). Ambitious and ambiguous: shifting purposes of national testing in the legitimation of assessment policies in Norway and Sweden (2000–2017). *Assessment in Education: Principles, Policy & Practice*, 25(3), 327-350.
- Vassli, O. (2014). *Skolen vet best-Hvorfor oppnår enkelte kommuner fremgang på nasjonale prøver?* ,
- Vestheim, O. P. (2018). Nasjonale prøver—hemmende styringsverktøy eller lokale redskap for praksisutvikling? *Acta Didactica Norge*, 12(4), Art. 3, 16 sider.
- Volckmar, N. (2014). Et blå-blått utdanningspolitisk skifte?-En studie av den utdanningspolitiske retorikken i partiprogrammene til stortingsvalget 2013 og Sundvollen-plattformen. *Norsk pedagogisk tidsskrift*, 98(06), 482-494.
- Walgermo, B. R., Uppstad, P. H., Lundetræ, K., Tønnessen, F. E., & Solheim, O. J. (2018). Kartleggingsprøver i lesing-tid for nytenking? *Acta Didactica Norge*, 12(4), Art. 7, 21 sider.

Vedlegg 1: Publikasjoner som er utelatt etter gjennomlesning

Alseth, B. r., Thronsen, I., & Turmo, A. (2009). Rapport fra kartleggingsprøver i tallforståelse og regneferdighet for 2. årstrinn og Vg1. *Acta Didactica*, 2, 2009.

Elstad, E., Nortvedt, G. A., & Turmo, A. (2009). *The Norwegian assessment system: An accountability perspective*. Paper presented at the Cadmo.

Elstad, E., Turmo, A., & Guttersrud, Ø. (2011). Achievement tests in the Norwegian school system: Types and uses. *Achievement Tests: Types, Interpretations and Uses*, Nova Science Publishers, Hauppauge, NY, 109-124.

Fasting, R. B., Thygesen, R., Berge, K. L., Evensen, L. S., & Vagle, W. (2009). National assessment of writing proficiency among Norwegian students in compulsory schools. *Scandinavian Journal of Educational Research*, 53(6), 617-637.

Gabrielsen, E., & Solheim, R. G. (2013). Over kneiken. *Leseferdighet på*, 4.

Hansen, H. H. (2015). *Arbeidet med resultatene av nasjonale prøver-på vei mot de nasjonale målene?* ,

Huang, L. (2009). Social capital and student achievement in Norwegian secondary schools. *Learning and Individual Differences*, 19(2), 320-325. doi:<https://doi.org/10.1016/j.lindif.2008.11.004>

Lindvig, Y., Wærness, J. I., & Fosnes, A. R. (uå). Bruk av farger og dialogkart i evidensinformert skoleutvikling.

Lyngsnes, K., & Vestheim, O. P. (2015). Praksis på skoler med gode resultater på nasjonale prøver.

Naigaga, D. A., Pettersen, K. S., Guttersud, Ø., & Henjum, S. J. N. S. i. S. E. (2019). Assessing adolescent self-efficacy in 'body and health'-Exploring the psychometric properties of the SEBH scale. *Nordic Studies in Science Education*, 15(2), 145-158.

Naigaga, D. A., Pettersen, K. S., Henjum, S., & Guttersrud, Ø. (2018). Assessing adolescents' perceived proficiency in critically evaluating nutrition information. *International Journal of Behavioral Nutrition Physical Activity*, 15(1), 61.

Nilsen, K. S. (2008). *Kartleggingsverktøy: en pedagogisk analyse av noen aktuelle verktøy som brukes i Norge i forhold til kartlegging av norskspråklige ferdigheter hos språklige minoriteter*.

Nordahl, I. S. (2018). *Nasjonale prøver og TIMSS som læreplan? En komparativ og temasentrert studie av LK06, TIMSS og nasjonale prøver*. UiT Norges arktiske universitet,

Nordahl, T., Kostøl, A. K., & Mausestaden, S. (2009). *Skoler med liten og stor forekomst av atferdsproblemer: En kvantitativ og kvalitativ analyse av forskjeller og likheter mellom skolene*. Hamar: Høgskolen i Hedmark.

Nordgård, R., & Strøm, R. K. O. (2014). *Nasjonale prøver i regning-som grunnlag for kunnskap og refleksjon rundt utviklingsarbeid*. (Master), University of Oslo, Oslo.

- Nortvedt, G. A. (2009). *The relationship between reading comprehension and numeracy among Norwegian grade 8 students.*
- Nortvedt, G. A. (2012). Bruk av tekstoppgaver på matematikktester og prøver: et kort review. *Kvalitet i norsk skole. Internasjonale og nasjonale undersøkelser av læringsutbytte og undervisning*, 212-222.
- Nortvedt, G. A. (2013). Are girls or boys better at mathematics? A commentary on the game of reporting gender differences. *Proceedings of the International Groups for the Psychology of Mathematics Education*, 385-392.
- Nortvedt, G. A., & Siqveland, A. (2018). Are beginning calculus and engineering students adequately prepared for higher education? An assessment of students' basic mathematical knowledge. *International Journal of Mathematical Education in Science and Technology*, 50(3), 325-343.
- Olsen, O. E. W., & Johnsen, E. S. (2018). *Har økt fysisk aktivitet i skolen større effekt på akademisk prestasjon hos elever fra familier med lav sosioøkonomisk status?*, Høgskulen på Vestlandet,
- Olsen, R. V., Tveit, S., & Björnsson, J. K. (2018). Nasjonale prøver og eksamener i norsk og svensk grunnopplæring. *Acta Didactica Norge*, 12(4), Art. 0, 6 sider.
- Raaheim, A., Mathiassen, K., Moen, V., Lona, I., Gynnild, V., Bunæs, B. R., & Hasle, E. T. (2019). Digital assessment – how does it challenge local practices and national law? A Norwegian case study. *European Journal of Higher Education*, 9(2), 219-231. doi:10.1080/21568235.2018.1541420
- Ravlo, G. (2010). Hvilken sammenheng er det mellom elevers prestasjoner på en test i naturfag og formatet til oppgavene?: En sammenligning av resultater oppnådd på lukkede oppgaver (SR) og tilsvarende oppgaver i åpent format (CD).
- Skaftun, A., & Solheim, O. J. (2013). Den nasjonale prøven i lesing på 5. trinn 2009.
- Skar, G. B. U., & Aasen, A. J. (2016). Risikotrekk og skjulte kvaliteter i elevtekstar. *Nordic Journal of Literacy Research*, 2.
- Skedsmo, G., & Mausehagen, S. (2017). Nye styringsformer i utdanningssektoren-spenninger mellom resultatstyring og faglig-profesjonelt ansvar. *Norsk pedagogisk tidsskrift*, 101(02), 169-179.
- Svensen, H. N. (2018). *Bekymring og kartlegging. En kvalitativ studie av lærere, kartleggingsprøver og bekymring i 1-4 trinn.* The University of Bergen,
- Solheim, O. J., Skaftun, A., & Walgermo, B. R. (2012). Den nasjonale prøven i lesing på 5. trinn 2012.
- Stokke Karlsøen, H.-H. (2018). *Selvregulert læring-hvilke kvaliteter kan identifiseres ved en prosessorientert prøve i matematikk på ungdomsskoletrinnet.* Høgskulen på Vestlandet/Western Norway University of Applied Sciences,
- Sørensen, I. (2015). Læreren og læremidlene: om læremiddelpraksis ved skoler som har skåret høyt på nasjonale prøver.
- Turmo, A., Guttersrud, Ø., Elstad, E., & Olsen, R. V. (2009). The impact of attending after-school care schemes on science achievement in primary school: A Norwegian study. *International Journal of Educational Research*, 48(5), 331-341.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy & Practice*, 21(2), 221-237. doi:10.1080/0969594X.2013.830079

Vestheim, O. P. (2017). *Nasjonale prøver-fra "tvangstrøye" til verktøy i utvikling av skolens praksis: En casestudie av praksis på skoler som over tid har oppnådd gode resultater på nasjonale prøver*. (PhD), NTNU, Trondheim.

Vinje, B. (2018). *How can we develop multiple choice items that support assessment for learning?* Paper presented at the The 19th Annual Conference of the Association for Educational Assessment – Europe, Arnhem-Nijmegen, the Netherlands.