

Sverre Tveit

Universitetet i Agder

Rolf Vegar Olsen

Universitetet i Oslo

DOI: <http://dx.doi.org/10.5617/adno.6381>

Eksamens mange roller i sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring

Sammendrag

Hvilke roller har eksamen i norsk grunnopplæring? I Forskrift til opplæringslova framgår det blant annet at eksamenskarakteren skal fastsette kompetansen til eleven ved slutten av opplæringen i faget. Denne artikkelen viser imidlertid at policydokumenter og policyutviklere har en rekke forventninger til hva man skal oppnå ved å ha eksamen og hva man skal kunne benytte eksamensresultater til. I et samspill mellom teori og empiri identifiser artikkelen tre roller for vurdering: å sertifisere, styre og støtte læring og undervisning. Dette begrepsapparatet brukes til å analysere både eksamens formelt definerte formål, og tilsiktede og utilsiktede funksjoner ved eksamenssystemet. Artikkelen baserer seg på tre typer historiske data: (i) Stortingsmeldinger, (ii) Utdanningsdirektoratets årlige rapporter, Utdanningsspeilet, samt (iii) ekspertintervjuer gjennomført i 2013 med en statssekretær og en seniorrådgiver i Kunnskapsdepartementet, og direktøren og en avdelingsdirektør i Utdanningsdirektoratet. I tillegg er det gjort en analyse av gjeldende bestemmelser, slik de kommer til uttrykk i Forskrift til opplæringslova og Rammeverket for eksamen av 2017. Artikkelen avdekker manglende avklaringer av eksamens formål og peker på at dette kan føre til at politikere og andre policyutviklere forvalter og videreutvikler eksamenssystemet på sviktende grunnlag. Videre kan det føre til at Utdanningsdirektoratet, Fylkesmannen og andre tilsynsmyndigheter, skoleeiere, skoleledere, lærere, foreldre og elever bruker eksamensdata på uheldige måter. Artikkelen konkluderer med at lovverk, forskrift og rammeverk for eksamen bør gjennomgås og endres til eksplisitt å uttrykke de rollene eksamen skal fylle og ikke. Slik kan man få et bedre utgangspunkt for å utvikle eksamensprøver og sikre hensiktsmessig bruk av eksamensresultater både nasjonalt og lokalt.

Nøkkelord: eksamen, standpunkt, vurdering, policy, karakterer, formål.

The multiple roles of national exams in the certification, governing and support of learning and instruction in Norwegian secondary education.

Abstract

What are the roles of the national exams in the Norwegian school system? The regulations of the Education Act simply state that exam marks should reflect the student's competency upon completing the education. This article demonstrates that policy documents and policy-makers express a range of expectations in the use of exam results. By integrating theoretical and empirical encounters, the article identifies three roles of educational assessment: to certify, govern and support learning and instruction. This conceptual framework is used to analyse the formally defined purposes of the exams and intended and unintended functions of the examination system. The article is based on three types of historical data: (i) Ministry reports to the Parliament («white papers»), (ii) the annual report of the National Agency for Education and Training («The Education Mirror») and (iii) expert interviews undertaken in 2013 with a state secretary and a government official in the Ministry of Education and Research, and the general director and department director of the National Agency for Education and Training. Additionally, it analyses current regulations of the Education Act and the «The Examination Framework» of 2017. The article illuminates that clarification of the purposes of the exams is insufficient and maintains that this may lead politicians and other policymakers to undertake insufficient approaches to the administration and development of the examination system. Further, it can cause inappropriate use of examination data by the executive agency, school inspectors, municipalities, school leaders, teachers, parents and students. The article concludes that legislation, regulations and the Examination Framework should be reviewed and changed to explicitly state the roles the examination system is expected to fulfil. This way, one may achieve a better foundation for the development of exams and facilitate sound use of exam data both nationally and locally.

Keywords: examinations, teacher judgments, assessment, policy, grades, purposes

Innledning

Denne artikkelen drøfter et tilsynelatende enkelt spørsmål: Hvilke roller har eksamen i norsk grunnsopplæring? Spørsmålet viser seg imidlertid å være vanskelig å få klare svar på, om man leser Opplæringslova, Forskrift til opplæringslova, Rammeverket for eksamen, stortingsmeldinger eller Utdanningsdirektoratets årlige rapportserie, *Utdanningsspeilet*. Disse policydokumentene utgjør grunnlaget for en analyse som beskriver eksamens rolle, fra forarbeidene til Kunnskapsløftet og fram til utgangen av 2017. Dette sammenstilles med intervjuer med sentrale policyutviklere i Kunnskapsdepartementet og Utdanningsdirektoratet gjennomført i 2013. Samlet danner dette grunnlaget for en drøfting av eksamens roller, slik det kommer til uttrykk i policydokumenter og sentrale policyutvikleres redegjørelser for eksamenssystemets innretning.

I artikkelen presenteres først et teoretisk rammeverk for å analysere de ulike rollene til vurderingsinstrumenter som eksamen. Videre presenteres de metodiske tilnærmingene og avveiningene i studien, inkludert en redegjørelse for utvalget av policydokumenter og ekspertinformanter som inngår i analysen. Analysen består av fire deler. Først analyseres to typer historiske tekster: fire stortingsmeldinger i perioden 2003 til 2016 og seks utgaver av Utdanningsdirektoratets årlige rapportserie, *Utdanningsspeilet*, fra perioden 2008 til 2017. Videre analyseres intervjuer med en statssekretær og en seniorrådgiver i Kunnskapsdepartementet og direktøren og en avdelingsdirektør i Utdanningsdirektoratet gjennomført i 2013. Analysens siste del tar for seg gjeldende forskrifter og Rammeverk for eksamen som kom i 2017. Avslutningsvis drøftes studiens funn, med fokus på det manglende samsvaret mellom forskjellige policyutvikleres forståelse av eksamens mange roller, og hvordan dette er formelt definert i Forskrift til opplæringslova og andre styringsdokumenter. Artikkelen avsluttes med å ta til orde for en gjennomgang av hvilke roller eksamen kan og bør ha i norsk grunnsopplæring.

Teoretiske perspektiver på vurderingsinstrumenters ulike roller

I beskrivelser av formål ved vurdering og evaluering er det vanlig å skille mellom *formativ* og *summativ*. Distinksjonen ble først introdusert i USA for mer enn femti år siden, av Michael Scriven (1967), som trakk opp skillelinjer mellom ulike «*roles of evaluation*» i utviklingen av undervisningsprogrammer (s. 41). Distinksjonen ble raskt knyttet til individuelle perspektiver på vurdering av andre forskere, som Bloom, Hastings og Madhaus (1971). De siste to tiårene har begrepet formativ vurdering blitt mye brukt i utdanningsforskning og politikktutforming i

Skandinavia, i stor grad inspirert av Black og Wiliam (1998a, 1998b, 2009) og Assessment Reform Group (2002) i Storbritannia. Distinksjonen blir også ofte omtalt med begrepsparet *vurdering FOR læring* og *vurdering AV læring*. Flere studier har belyst forskning på formativ vurdering og statlige og kommunale satsinger på Vurdering for læring, både i Norge og Sverige (Burner, 2016; Gamlem, 2015; Hopfenbeck, Petour, og Tolo, 2015; Jonsson, Lundahl, og Holmgren, 2015; Throndsen, Hopfenbeck, Lie og Dale, 2009; Tveit, 2017, 2018a, 2018b).

Black og Wiliams (2009) videreutviklede teori om formativ vurdering har vært et fruktbart rammeverk for å få grep på hvordan vurdering kan brukes til å støtte opp om læring og undervisning (Gamlem & Smith, 2013). Black og Wiliams' konseptualisering av formativ vurdering fra 1998 og 2009 inneholder imidlertid ingen definisjon av summativ vurdering, hvilket kan forklare forskeres og policyutvikleres utfordringer med å forstå distinksjonens meningsinnhold (Bennett, 2011). En analyse av bruken av distinksjonen mellom formativ og summativ vurdering i forskningslitteratur (Tveit, 2017, 2018a) viser at vi ved å bruke den todelte distinksjonen risikerer å underslå viktige sider ved lands vurderingspolitikk, ettersom bruken av begrepsparet tilslører ulike måter såkalte summative vurderinger brukes på.

Herman og Baker (2009) peker på at «å spesifikt formulere hvordan vurdering kan orkestreres til å ivareta disse varierte policy-målene innebærer at man må forstå bredden av potensielle formål ved vurdering» (s. 177, vår oversettelse). Newton (2007) identifiserer hele 18 ulike måter vurderinger brukes på. Han poengterer at all bruk av vurdering baserer seg på en «summativ bedømming» (*summative judgment*), og ser således formativ vurdering som bare én av mange måter å bruke summative bedømminger på. Tveit (2017, 2018a) viser at Stobart (2008) og Black (1998) identifiserer tre overordnede formål for vurdering, som kan sies å ha et sammenfallende innhold. Disse forfatterne belyser et viktig skille mellom to bruksområder for såkalte summative vurderinger: individuelle summative vurderinger og bruk av summative vurdering til å styre og holde skolene ansvarlige for resultater.

Ved å integrere perspektivene til Scriven (1967), Sadler (1989), Black og Wiliam (2009), Black (1998), Newton (2007) og Stobart (2008) har Tveit (2017, 2018a) utviklet et teoretisk rammeverk for forskning på ulike roller for vurdering (Tabell 1). Som vist i tabellen defineres her vurdering som en prosess for å *fastsette mål (eller standard) -oppnåelse* og det skilles mellom tre *roller for vurdering*: vurdering brukt til å *sertifisere*, *styre* og *støtte* læring og undervisning. Ved å bruke Scrivens «rolle»-begrep, anerkjennes det at vurderinger kan ha flere både tilsiktede og utilsiktede konsekvenser utover formelt definerte formål. Dette rammeverket brukes for å analysere omtalen av eksamenssystemet i studiens policydokumenter og ekspertintervjuer.

Tabell 1: Klassifisering av roller for vurdering

Prosess	Å fastsette mål (eller standard) -oppnåelse		
Rolle	Å <i>sertifisere</i> læring og undervisning	Å <i>styre</i> læring og undervisning	Å <i>støtte</i> læring og undervisning
Nivå	Elev- og lærernivå (standpunktvurdering, eksamen)	Organisasjonsnivå (skoler, skoleeier, nasjonalstater)	Elev- og lærernivå (vurdering i klasserommet)
Institusjonell praksis	Å identifisere og rapportere sluttvurdering av kompetansenivå (standpunkt- og eksamenskarakter); brukt til sertifisering og seleksjon til videre utdanning og yrkesliv.	Å evaluere (aggregerte) elevdata for måloppnåelse; brukt til å (a) gi grunnlag for beslutningstakeres kvalitetsutviklingstiltak; og (b) kontrollere praktiseringen av læreplaner, forskrifter og retningslinjer	Å identifisere og kommunisere avvik mellom aktuell og forventet måloppnåelse, brukt til å støtte lærings- og undervisningsstrategier
Assosiert med	Summativ vurdering		Formativ vurdering

Metode

Artikkelen baserer seg på tre typer historiske data: Stortingsmeldinger og Utdanningsdirektoratets årlige rapporter, *Utdanningsspeilet*, samt ekspertintervjuer som ble gjennomført i 2013. Datamaterialet presenteres i Tabell 2. I tillegg er det gjort en analyse av gjeldende bestemmelser slik de kommer til uttrykk i Forskrift til opplæringslova¹ og Rammeverk for eksamen², utgitt i 2017.

I analysen har det ikke vært et mål å se på hva de ulike kildene legger *mest* vekt på. Hensikten har vært å vise bredden av roller som kommer til uttrykk når eksamens formål omtales direkte eller indirekte i datamaterialet. Samlet belyser stortingsmeldingene, de årlige Utdanningsspeilet-rapportene og ekspertintervjuene en historisk utvikling knyttet til Kunnskapsdepartementets og Utdanningsdirektoratets forståelse av eksamens roller, samt gjeldende bestemmelser i dagens forskrifter og rammeverk.

¹ <https://lovdata.no/dokument/SF/forskrift/2006-06-23-724>

² <https://www.udir.no/eksamen-og-prover/eksamen/rammeverk-eksamen/>

Tabell 2: Policydokumenter og ekspertintervjuer i den historiske analysen

Myndighetsnivå	Policydokumenter (2001-2017)	Ekspertintervjuer (2013)
Kunnskapsdepartementet	Stortingsmelding 30 (2003-2004)	Statssekretær
	Stortingsmelding 16 (2006-2007)	
	Stortingsmelding 20 (2012-2013)	Seniorrådgiver
	Stortingsmelding 28 (2015-2016)	
Utdanningsdirektoratet	Utdanningsspeilet 2008, 2009, 2011, 2014, 2016 og 2017.	Direktør
		Avdelingsdirektør med ansvar for eksamen

Dokumentanalyse

Dokumentanalysens historiske del ble avgrenset til stortingsmeldinger og den årlige rapportserien til Utdanningsdirektoratet, *Utdanningsspeilet*, fordi disse representerer en statusgjengivelse og formidling av hva myndighetene anser som viktig for henholdsvis den politiske og administrative styringen av utdanningssystemet. Gjennomgangen av stortingsmeldinger starter med forarbeidene til Kunnskapsløftet, Stortingsmelding nr. 30 (2003-2004), og avsluttes med grunnlaget for den pågående fagfornyelsen i grunnopplæringen, Stortingsmelding nr. 28 (2015-2016). I tillegg til de nevnte stortingsmeldingene fant vi eksamenssystemet omtalt kun i to andre stortingsmeldinger: Stortingsmelding nr. 16 (2006-2007) og Stortingsmelding nr. 20 (2012-2013). Som vi skal se inneholder disse dokumentene få, men sentrale perspektiver på eksamenssystemets roller.

I gjennomgangen av Utdanningsdirektoratets årlige rapportserie *Utdanningsspeilet* startet vi med rapporten for 2008 og gikk igjennom rapportene til og med den siste utgivelsen i 2017. Basert på en innledende gjennomlesing ble rapportene fra 2008, 2009, 2011, 2014, 2016 og 2017 valgt ut for nærmere analyse. Disse seks rapportene ble valgt ut fordi de bidrar til å vise fram en bredde i omtalene av ulike roller som eksamen har. Siste del av analysen består av gjeldende bestemmelser, uttrykt i Forskrift til opplæringslova, og Rammeverket for eksamen, utgitt i 2017.

Ekspertintervjuer

Utvelgelsen av kandidater for ekspertintervjustudier er svært annerledes enn for studier som har til hensikt å nominere et representativt utvalg fra en gruppe eller populasjon. Ekspertinformanter nomineres basert på deres ekspertise og ansvarsområder, og det var derfor her hensiktsmessig å intervjuer policyutviklere med lederansvar. For det politiske nivået i Kunnskapsdepartementet ble først henvendelse rettet til statsråden gjennom et brev til avdelingsdirektør i Opplæringsavdelingen med ansvar for eksamensfeltet. Dette resulterte i en intervjuavtale med statssekretæren med ansvar for skole. Representanten for embetsverket ble også rekruttert på bakgrunn av en henvendelse rettet til avdelingsdirektøren i Opplæringsavdelingen, som anbefalte den mest sentrale medarbeideren i

embetsverket knyttet til temaene for intervjuet. For Utdanningsdirektoratet ønsket vi å intervju direktøren og avdelingsdirektøren med ansvar for eksamen. Begge takket ja til å bidra til studien basert på direkte henvendelse. Informantene ble opplyst om at de i kraft av sitt ansvar måtte påregne å bli indirekte identifisert når studiens resultater ble publisert. I denne typen studier kan indirekte identifisering ikke unngås, ettersom datamaterialets validitet (utsagnenes legitimitet) hviler på informantenes formelle posisjoner.

Bogner og Menz (2009) identifiserer tre typer ekspertintervjuer. *Eksplorerende ekspertintervjuer* (i), som brukes i utforskning av relativt ukjente forskningsfelt; *Systematiske ekspertintervjuer* (ii), brukes til å rekonstruere «objektiv kunnskap» i et spesifikt felt; og *teorigenererende ekspertintervjuer* (iii), som legger vekt på å få fram informantens latente spesifikke fortolkende kunnskap som er tilegnet gjennom profesjonell praksis. Denne studien baserer seg på en kombinasjon av eksplorerende (i) og teorigenererende (iii) ekspertintervjuer.

Dexter (1970/2006) anbefaler at intervjuobjektet selv tar en aktiv rolle i å bestemme hva som er relevant å bringe inn i samtalen (s. 18). I disse ekspertintervjuene fikk informantene anledning til å snakke om det de selv fant relevant å bringe inn, gitt den overordnede tematikken. I intervjustudier som denne forsøker vi ikke å gjøre krav på objektivitet, snarere var det viktig at intervjuerens for forståelse er synlig for den som intervjues. Intervjuene hadde til hensikt å kartlegge ekspertenes synspunkter og refleksjoner, inkludert mindre formelt definerte roller eksamenssystemet har. Vi benyttet det Fontana og Frey (2005, s. 696) kaller for *empatisk intervjuing*, for å skape en situasjon hvor informantene var villige til å bidra med utfyllende og ærlige responser. Analysen og framstillingen av intervjudataene med avdelingsdirektøren i Utdanningsdirektoratet innebar en omfattende reduksjon av datamaterialet. Denne informanten ble forelagt utkastet til artikkel der dataene var redusert og framstilt i tilnærmet ferdig utgave, sammen med en fyldigere narrativ framstilling av intervjuet i et eget dokument. Denne valideringen bidro til noen viktige presiseringer.

Begrensninger ved studien

Denne studiens ekspertintervjuer ble gjennomført i 2013. Intervjudataene anses likevel som relevante og viktige, da de belyser langt flere perspektiver fra policyutviklere enn det som kommer fram i policydokumentene. Det har forøvrig ikke blitt gjort omfattende endringer i eksamenssystemet siden 2013, selv om det i 2017 kom et nytt *Rammeverk for eksamen*. På tidspunktet for intervjuene pågikk det dessuten et omfattende utviklingsarbeid i tilknytning til satsingen *Vurdering FOR læring*, i forlengelsen av omfattende forskriftsendringer i 2009 etter innføringen av Kunnskapsløftet i 2006. Følgelig var forutsetningene på denne tiden gode for

bevissthet hos informantene knyttet til ulike roller for vurdering generelt, og eksamen spesielt. Det hadde vært ønskelig å gjennomføre nye ekspertintervjuer for å belyse utviklingen fra 2013 og fram til i dag. Det hadde også vært ønskelig å analysere de årlige eksamensrapportene til Utdanningsdirektoratet. Det har imidlertid ikke vært mulig innenfor rammen av denne studien. Dokumentanalysene og analysen av gjeldende forskrifter og rammeverk bidrar imidlertid til å belyse utviklingen som er skjedd også de siste fem årene i den undersøkte perioden.

Analyse av data

I denne delen analyseres først de fire stortingsmeldingene og de seks årlige rapportene fra Utdanningsdirektoratet. Deretter analyseres intervjuene med policyutviklere i Kunnskapsdepartementet og Utdanningsdirektoratet. Avslutningsvis analyser gjeldende Forskrift til opplæringslova og Rammeverk for eksamen. Alle sitater og tilhørende sidetallshenvisninger i denne analysedelen refererer til det aktuelle policydokumentet som analyseres, også der eksempelvis rapporter er nevnt i dokumentet som analyseres.

Analyse av stortingsmeldinger (2003-2017)

I *Stortingsmelding nr. 30 (2003-2004)* ble eksamen omtalt slik: «Eksamen i videregående opplæring og avgangsprøven i grunnskolen er spesielt kvalitetssikrede prøver som er sentralt eller lokalt/regionalt gitt» (s. 37). Utover denne beskrivelsen av hva eksamen er, og at de kan være utviklet på ulike nivåer i systemet, ga ikke dette dokumentet noen eksplisitt avklaring av hvilke roller eksamen har.

Dette kom litt tydeligere fram i *Stortingsmelding nr. 16 (2006-2007)*: «Eksamen og fag- og svenneprøve er en sluttvurdering som først og fremst skal informere samfunnet, arbeidsgivere og utdanningsinstitusjoner om den kompetansen som eleven, lærlingen eller lære kandidaten har oppnådd» (s. 78). Her kom *sertifiseringsrollen* klart fram.

Stortingsmelding nr. 20 (2012-2013) inneholdt en noe mer omfattende drøfting av hvordan eksamenssystemet fungerer:

Trekkordningen innebærer at elevene ikke skal ha eksamen i alle fag, men at de skal være eksamens-forberedt i de fagene der eksamen er en mulig sluttvurdering ved siden av standpunkt karakteren (s. 65-66).

I denne kom det fram hvordan eksamen fungerer som et incentiv for at elevene arbeider godt i alle fagene, slik at de er forberedt i tilfelle de trekkes ut til eksamen. Dette kan ses på som at eksamen har en rolle i å *styre*

hvordan en sikrer et bredest mulig grunnlag for *sertifiseringen* som gjennomføres ved standpunktvurdering.

Videre tok stortingsmeldingen opp betydningen av sensorskolering, først og fremst med henblikk på å sikre at eksamenskarakterene settes på en rettferdig måte:

Nasjonale myndigheter engasjerer praktiserende lærere til å utarbeide sentralt gitt eksamen. Det er også praktiserende lærere som får i oppdrag å være sensorer. For å sikre en rettferdig vurdering, er det viktig at sensorer vurderer besvarelser så likt som mulig. Det legges derfor vekt på god sensorskolering i alle fag. I Kunnskapsløftet er det utviklet veiledninger med kjennetegn på måloppnåelse til eksamen i alle fag. Veiledningene brukes på sensorskoleringene. Det oppfordres også til at elevene blir gjort kjent med dem før eksamen. Arbeidet med å kvalitetssikre, forvalte og vurdere eksamen er viktig for elevenes rettssikkerhet (s. 66).

Her kommer det også fram at eksamen har en rolle i å *støtte* elevenes læring, gjennom at de blir gjort kjent med veiledningene med kjennetegn på måloppnåelse. Samlet tar stortingsmeldingen opp samspillet mellom eksamens roller i styring, sertifisering og støtte av læring og undervisning.

I *Stortingsmelding nr. 28 (2015-2016)*, gikk Kunnskapsdepartementet mer eksplisitt til verk i å drøfte ulike roller ved eksamenssystemet:

Departementet mener dagens eksamenssystem fyller flere viktige funksjoner. Eksamen er en form for kvalitetssikring for elevene fordi de får en eksternt vurdering av sin fagkompetanse. Fag med sentralt gitt eksamen har identiske oppgavesett for alle som kommer opp i faget. Det er med på å gi elevene et mer likeverdig vitnemål fordi eksamenskarakterene settes på det samme vurderingsgrunnlaget. Eksamensresultatene kan i tillegg bidra til at lærerne og skolen kan kontrollere og utvikle sitt eget arbeid med vurdering. Karakterer som gis til eksamen er en tilbakemelding til skolen om hvordan eksterne sensorer vurderer elevenes eksamensprestasjoner, det kan bidra til at skolen og lærerne justerer egen praksis og oppnår en mer rettferdig vurderingspraksis. Det gjelder både sentralt og lokalt gitt eksamen (s. 62-63).

Meldingen ga altså helt eksplisitt uttrykk for en forventning om at skolene og lærerne bruker eksamensresultatene til å justere praksisen med standpunktvurdering. Dette er nok et uttrykk for at eksamen brukes til å *styre* skolenes arbeid med *sertifisering*. Departementet rettet imidlertid i den samme meldingen en advarende pekefinger mot for enkle sammenligninger av eksamens- og standpunkt karakterer:

Undersøkelser viser at det er en relativt utbredt praksis at skoleeiere og skoler sammenligner skolers standpunkt karakterer med eksamens karakterer i fagene for å vurdere om det er systematiske avvik mellom dem over tid. Det er viktig å understreke at eksamens- og standpunkt karakterer er to ulike uttrykk for

kompetanse. En eksamenskarakter vil være satt på et mer begrenset vurderingsgrunnlag enn en standpunkt karakter. En eksamen representerer én eller veldig få vurderingsformer (skriftlig, praktisk og/eller muntlig) og den er et enkelttilfelle. Nasjonalt ligger de gjennomsnittlige eksamenskarakterene for sentralt gitt eksamen vanligvis noe under de gjennomsnittlige standpunkt karakterene. Hvis en slik sammenlikning skal være hensiktsmessig, er det departementets vurdering at en må se på om forskjeller over tid avviker systematisk fra den nasjonale, gjennomsnittlige differansen mellom standpunkt- og eksamenskarakterer. Varige avvik bør være et varsel til skoleeier og skoleleder om at det er nødvendig å endre eksisterende vurderingspraksis. Slike analyser er imidlertid ikke nok alene, men bør være én av flere kilder til kunnskap om egen praksis til bruk i en skoles utviklingsarbeid (s. 63).

Departementet var her opptatt av å sikre en mer forsvarlig bruk av eksamensresultater, men ga departementet klare føringer om at systematiske forskjeller er å anse som «et varsel» og bør avstedkomme endringer i praksisen for standpunkt vurdering, nok et uttrykk for bruk av eksamen til *styring* av skolens arbeid med *sertifisering*.

Kunnskapsdepartementet var i denne meldingen også mer opptatt av hvordan man skal få stort nok sensor korps til å gjennomføre sentralt gitte eksamener, og anmodet skoleeier om å arbeide systematisk med dette.

En sentral sensur av alle elevene på et årstrinn vil kreve at flere lærere enn i dag må bidra som sensorer. Det betyr at det må vurderes ulike tiltak for å rekruttere flere lærere til å være sensorer ved sentralt gitt eksamen. Det vil føre til at flere lærere vil delta på sensorskolering og få nyttig sensorerfaring og utvikler sin vurderingspraksis. Skoleeier bør aktivt og systematisk bygge opp en sensor kompetanse i sine skoler. Kompetanse og erfaringer bør benyttes i skolebasert kompetanseutvikling i vurdering. Sensorskoleringen som arrangeres i forbindelse med sentralt gitt eksamen er en viktig kilde til utvikling av læreres vurderingskompetanse (s. 64).

Deltakelse i sensor korpsset identifiseres eksplisitt som et viktig tiltak for å heve lærernes kompetanse, hvilket er et uttrykk for at eksamen har en rolle i å *støtte* lærernes undervisning.

Analyse av Utdanningsdirektoratets årlige rapporter (Utdanningspeilet) (2008-2017)

Utviklingstrekkene og temaene som belyses i stortingsmeldingene finner vi igjen i de seks årlige rapportene fra Utdanningsdirektoratet.

Utdanningspeilet 2008 viste til at «Det er utarbeidd vurderingsrettleingar med kjenneteikn på måloppnåing i dei fleste eksamensfaga (...)» som «skal gje sensorane rettleiing i kva som kjenneteiknar dei ulike karakternivåa, og dette kan fremje ein meir rettferdig vurderingspraksis i skulen» (s. 105). Her ble *sertifiseringsrollen* vektlagt. Men Utdanningsdirektoratet skrev også at «Vurderingsrettleiingane skal ha

ein læringsfremjande effekt ved at lærarar kan formidle kjenneteikna til elevane før eksamen» (s. 105). Vurderingsveiledningenes rolle i å *støtte* læring og undervisning ble altså også vektlagt. Utdanningsdirektoratet la videre vekt på hvordan sensorskoleringene bidrar til diskusjoner og felles forståelse av hva som kreves i fagene og hvordan dette er en del av sluttvurderingen til eksamen.

Rapporten viste også til undersøkelser gjort av Statistisk sentralbyrå (Hægeland mfl, 2005) som identifiserte «en tendens til at ein elev som har medelevar med ein fordelaktig sosial bakgrunn, får noko lågare standpunkt karakter» og at det motsatt ser ut til «å vere noko lettare å få ein høgare standpunkt karakter dersom ein har medelevar med ein mindre fordelaktig sosial bakgrunn» (s. 105). Det understrekes at dette «ikkje er tilfelle for eksamens karakterar» (s. 105). I rapporten ble det også vist til analyser av skoleresultatene for 2007 (Gravaas mfl., 2008), som viste samme tendensen. Forskerne kom fram til dette ved å sammenligne elevenes standpunkt- og eksamens karakterer, et uttrykk for bruk av eksamensresultater til å *styre* skolens arbeid med *sertifisering*. Som vi skal se, la Utdanningsdirektoratet etterhvert stor vekt på slike studier i de årlige rapportene de påfølgende årene.

I *Utdanningsspeilet 2009* var Utdanningsdirektoratet også opptatt av forholdet mellom karakterer til standpunkt og eksamen. De viste til en gjennomgående tendens til at «karaktergjennomsnittet i hovedsak er høyere til standpunkt enn til skriftlig eksamen, mens det høyeste karaktergjennomsnittet kommer til muntlig eksamen» (s. 50). De påviste også at «de fagene som ikke har skriftlig eksamen, som for eksempel samfunnsfag, har høyere gjennomsnittlig standpunkt karakter enn fag som matematikk og norsk hovedmål» (s. 50). Dette fokuset på at sentralt gitt eksamen bidrar til å begrense inflasjonen i lærernes karaktersetning er et uttrykk for bruk av eksamensresultater til å *styre* skolens arbeid med *sertifisering*.

I rapporten viste Utdanningsdirektoratet også til et pågående arbeid de fikk oppdrag om fra Kunnskapsdepartementet i 2007, om å sette i gang en rekke tiltak for å styrke vurderingspraksisen i Norge (s. 96). Også i denne årlige rapporten peker Utdanningsdirektoratet på arbeidet med vurderingsveiledninger med kjennetegn på måloppnåelse, og at dette er utarbeidet «i dei fleste eksamensfaga» (s. 104). Som i 2008-utgaven ble det lagt vekt på at vurderingsveiledningene «skal ha ein læringsfremjande effekt ved at lærarar kan formidle kjenneteikna til elevane før eksamen» (s. 104). Følgelig ble det også her lagt vekt på eksamens rolle i å *støtte* læring og undervisning.

Også i *Utdanningsspeilet 2011* ble det henvist til undersøkelser som ser på skolens praksiser for standpunkt karaktersetning, der eksamens karakterer utgjør referansegrunlaget. Utdanningsdirektoratet konkluderte i rapporten

med at det er «betydelig variasjon i hvordan grunnskoler setter standpunktkarakterer» (s. 64). Påpekningen av forskjeller mellom standpunktkarakterer og eksamenskarakterer omtalt i rapporten, er et uttrykk for at eksamen brukes til *styring* av skolenes arbeid med *sertifisering*. Det problematiseres at «skriftlig eksamen på fem timar eller ein munnleg eksamen på ein halv time, aldri kan prøve like breitt» (s. 112) som lærernes standpunktkarakter, et uttrykk for svakheter ved eksamens rolle i å *styre* skolenes arbeid med *sertifisering*. Videre ble det framholdt at «Sidan eksamen blir gjennomført med ekstern sensur, har eksamen eit element av ekstern kvalitetssikring i seg» (s. 112). Her er fokus på at eksamenens eksterne sensur bidrar til en mer kvalitetssikret form for *sertifisering* enn standpunktvurderingene.

Rapporten pekte også på at lærerne ikke får vite eksamenskarakteren til eleven før standpunktkarakteren er satt, «slik at eksamen ikkje har nokon justerande effekt på karakterane til den enkelte eleven» (s. 113). I rapporten viste Utdanningsdirektoratet likevel til en studie av Prøitz og Spord Borgen (2010), som viste «eksempel på at leiinga ved skolen ser på forholdet mellom eksamen og standpunktkarakterar som ei temperaturmåling på om ein driv med rett karaktersetjing ved skolen» (s. 113). Utdanningsdirektoratet konkluderte med at «slik sett kan ein seie at eksamen har ei kalibrerande rolle» (s. 113). Dette var første gang eksamens rolle i å *styre* skolenes arbeid med *sertifisering* ble uttrykt så klart av Utdanningsdirektoratet. Vi merker oss at Utdanningsdirektoratet med henvisning til Prøitz og Spord Borgen (2010) *deskriptivt* omtalte denne «kalibrerende rollen».

I *Utdanningsspeilet 2014* tok Utdanningsdirektoratet opp en ny form for sammenligning av resultater, denne gangen mellom resultater på nasjonale prøver og eksamen. De viste også til nye undersøkelser fra Statistisk sentralbyrå (2013), som viste at «halvparten av elevene som var oppe til eksamen, fikk lavere karakter til eksamen enn i standpunkt» (s. 75). Likevel påviste Utdanningsdirektoratet en bedring i så henseende:

De to siste årene har differansen mellom skriftlig eksamenskarakter og standpunktkarakter blitt mindre. En større andel elever får samme karakter til standpunkt og eksamen. Samtidig er andelen som får høyere eksamenskarakter enn standpunktkarakter, blitt større, og det er færre som får lavere (s. 76).

I rapporten kommer det også fram at «Utdanningsdirektoratet de siste årene [har] arbeidet systematisk med å heve kvaliteten på sensuren, blant annet gjennom årlige sensorskoleringer», hvilket settes i sammenheng med at forskjellen mellom de skriftlige eksamenskarakterene og standpunktkarakterene har blitt mindre (s. 76). Her brukte Utdanningsdirektoratet altså resultater fra både eksamen og nasjonale prøver i *styringen* av utdanningssystemet generelt, og spesielt knyttet til eksamens

rolle i å begrense karakterinflasjon i skolenes *sertifisering*. Det ser ut som det her har skjedd en forskyvning fra at Utdanningsdirektoratet *deskriptivt* omtaler skolenes bruk av eksamensresultater til sammenligning, til at Utdanningsdirektoratet omtaler sammenhengen mellom standpunkt og eksamenskarakterer i *normative* termer der det er et mål at forskjellene er små.

Utdanningsspeilet 2016 brakte inn et nytt element knyttet til eksamensrolle i å *styre* skolenes arbeid med *sertifisering*. En nedgang i snittkarakteren til skriftlig eksamen i matematikk for 10. trinn fra 3,2 til 2,9 ble omtalt, og Utdanningsdirektoratet var opptatt av «å avdekke om det er eksamensoppgavene som er årsaken til at resultatene på eksamen går ned» (s. 84). Utdanningsdirektoratet presenterte en analyse gjort av Matematikksenteret (2016), som slo fast at slike variasjoner i noen grad kan forklares av at det er systematisk variasjon i innhold og vanskegrad fra en eksamen til den neste. Med andre ord ble det antydning at eksamen vanskelig kan brukes for å si noe om hvordan elevenes læringsutbytte endres fra år til år. Her problematiseres altså eksamensrolle i å *styre* skolenes arbeid med *sertifisering*. Dette var starten på et mer kritisk blikk Utdanningsdirektoratet har hatt på skolenes bruk av eksamensresultater.

I *Utdanningsspeilet 2017* ble det gitt en omtale av sluttvurdering, der det trekkes opp et skille mellom hva standpunkt- og eksamenskarakteren måler:

Både standpunkt- og eksamens-karakteren i et fag blir fastsatt på bakgrunn av de samlede kompetansemålene i læreplanen, men de måler ulike sider ved elevens kompetanse:

- Standpunktkarakteren skal vise elevens samlede kompetanse etter endt opplæring, og skal vurderes på et så bredt grunnlag som mulig.
- En eksamenskarakter skal uttrykke elevens samlede kompetanse slik den kommer til uttrykk på eksamen (kapittel 5.2).

Det er grunnlag for å anta at denne presiseringen har bakgrunn i at skoleeiere og skoleledere sammenligner eksamens- og standpunktkarakterene i kontrollen med og *styringen* av skolenes *sertifiseringspraksis*, og at Utdanningsdirektoratet (nå) anså måten dette gjøres på som problematisk.

Samlet viser analysene av Utdanningsdirektoratets årlig rapporter at eksamen, i tillegg til å være en form for *sertifisering* i seg selv, anses å ha en viktig rolle i *styringen* av skolenes arbeid med sertifisering. Vi merker oss en glidning fra en mer *deskriptiv* omtale av skolene og skoleeiers bruk av eksamensresultater til å styre standpunktvurderingen (Utdanningsspeilet 2008, 2009, 2011), til at Utdanningsdirektoratet selv bruker eksamensdata til å gjøre slike sammenligninger og oppfordrer skolene til å gjøre det samme (Utdanningsspeilet, 2014), før man går over til en mer problematiserende omtale av denne bruken av eksamensresultater til å styre skolenes arbeid med standpunktvurdering (Utdanningsspeilet 2016, 2017). Bruk av

eksamensresultater til å styre standpunktverdinger er det mest sentrale temaet i de undersøkte rapportene. I tillegg ble det innimellom rettet fokus på hvordan vurderingsveiledningene kan bidra til å *støtte* elevenes læring og lærernes undervisning.

Analyse av intervjuer med policyutviklere (2013)

Policyutviklerne ble spurt hva de oppfattet at formålet med eksamen er.

Statssekretæren framholdt at eksamen er «én måte å vurdere elevenes måloppnåelse på» og at «det skal tjene til videre opptak». Her var statssekretæren opptatt av eksamensprøvens sertifiseringsrolle. Videre framholdt statssekretæren viktigheten av «en hjelp til skoler, lærere og sensorer på hva som skal måles» og «hva som er kriteriene». Statssekretæren var også opptatt av at hva «som er forventningen» er «gitt og kjent for elevene i deres læringsarbeid» og «at det skapes en kultur for hva som er standarden». Her pekte statssekretæren først på eksamens rolle i å *styre* (klargjøre for skoler, lærere og sensorer hva som skal måles), men også på å *støtte* (at forventningene klargjøres for elevene) læring og undervisning.

Seniorrådgiveren i Kunnskapsdepartementet poengterte at eksamen «er en så sterk del av vår tradisjon og kultur i Norge» og at det er «et eksternt element». Seniorrådgiveren framholdt også at Utdanningsdirektoratet «er veldig dyktige på å drive med sensorskoleringer og å lage sensorveiledninger». Videre beskrev seniorrådgiveren eksamen som et omstridt felt i forbindelse med læreplanreformer og endringer i fagsammensetning. «Det er stortinget som bestemmer her», fastslo seniorrådgiveren, og pekte på at dette blir særlig tydelig om man forsøker å gjøre endringer rundt for eksempel sidemålet. «Ja, det kan jo være sterke grupper. Det kan være mållaget, det kan være Utdanningsforbundet, LO, NHO». Den sistnevnte uttalelsen er uttrykk for en annen form for styring eksamen brukes til, nemlig i å signalisere *prioriteringer mellom fag* i skolen og *prioritering av fagenes innhold*.

Utdanningsdirektoratets direktør framholdt at eksamen representerer en kvalitetssikring, spesielt i videregående opplæring: «Det er en kvalitetssikring i det, at man skal ha en nasjonal vurdering, fordi disse karakterene brukes til opptak videre». Direktøren framholdt at «det skal være likt» og at «det skal være en nasjonal standard».

Videre framholdt direktøren at eksamen «forhåpentligvis» brukes «til å kalibrere annen karaktersetting». Direktøren la vekt på at «i utgangspunktet skal det jo være lik vurdering bak en karakter gitt i Trondheim som i Rogaland eller Finnmark», selv om «det ikke alltid er sånn». Direktøren understreket at «det er vanskelig å få det helt likt, fordi det er mange som vurderer» og pekte på at det «gjennom en nasjonal eksamen og gjennom at mange deltar i sensorkorpset (...) så gjør vel det noe med at du får en kalibrering av karakternivå».

Om direktøren med «kvalitetssikring» henspilte på eksamens rolle i *sertifisering* i seg selv (kvalitetssikrede eksamensprøver) eller eksamens rolle i å *styre* skolenes arbeid med sertifisering (sikre at lærere setter standpunkt karakterer etter samme kvalitetsstandard) er usikkert. Men når direktøren tok opp at eksamen forhåpentligvis brukes til å «kalibrere annen karaktersetting», så henspiller det tydelig på eksamens rolle i å *styre* skolenes *sertifisering*. Vi merker oss at direktøren her sa at eksamen «forhåpentligvis» brukes til kalibrering, altså en normativ tilnærming som signaliserer at skolene skal bruke eksamensdata på denne måten. Dette skiller seg fra Utdanningsdirektoratets skriftlige rapporter, som i årene før (2008, 2009, 2011) kun omtalte skolenes bruk av resultater på denne måten, og sammenfaller med hvordan det ble omtalt i den nærmeste etterfølgende rapporten som tok opp dette temaet (Utdanningsspeilet, 2014).

Utdanningsdirektoratets avdelingsdirektør med ansvar for eksamen ga som forventet den grundigste redegjørelsen for eksamens formål og funksjoner. Avdelingsdirektøren framholdt først at det «formelle formålet» i henhold til Forskrift til opplæringsloven er at sluttvurdering skal «gi uttrykk for den kompetansen eleven har oppnådd på slutten av opplæringen». Avdelingsdirektøren viste til at eksamens karakteren skal på et vitnemål, og således er et verdipapir. Forskjellen på eksamen og standpunkt karakterer ble tydelig uttrykt: «Den karakteren de får på eksamen skal gi uttrykk for hva eleven har vist av kompetanse på den eksamen. Standpunkt sier noe om hele faget, eksamen sier noe om hva elevene har uttrykt av kompetanse på den eksamen». Innledningsvis trakk avdelingsdirektøren altså opp en svært avgrenset forståelse av eksamens formål: *sertifiseringsprøver*.

Avdelingsdirektøren var likevel opptatt av å få fram hvilke andre roller eksamen har, som ikke formelt er definert i Forskrift til opplæringsloven. Med henvisning til måten Utdanningsdirektoratet lager veiledninger for kjennetegn på måloppnåelse ga avdelingsdirektøren uttrykk for at det er «åpenbart» at «systemet med eksamen har flere effekter». Avdelingsdirektøren ga uttrykk for at «det er jo en slags veiledning generelt i det, som kan brukes inn mot både standpunkt og generell vurdering, selv om materialet vi lager er knyttet til eksamen». Videre at «dette er en del av det elevene også bør forberede seg på». Slik uttrykte avdelingsdirektøren tydelig gjennom ett lengre utsagn om eksamens «flere effekter» det komplekse samspillet mellom eksamens roller i sertifisering, styring og støtte av læring og undervisning.

Avdelingsdirektøren var også opptatt av hvordan eksamenssystemet bidrar til kompetanseutvikling gjennom at «3000 lærere hvert år» går igjennom sensorskolering: «Fordi det er lærere som er sensorer, det er lærere som lager oppgavene, det er lærere som blir skolert», så representerer eksamenssystemet «den kilden til kompetanseutvikling, faglig diskusjon, og opparbeidelse av tolkningsfellesskap som lærere i Norge har hatt».

Avdelingsdirektøren pekte på at «veldig mange steder er rektorer opptatt av at de skal ha sensorer i alle fag, fordi dette er noe de har brukt aktivt inn mot kollegiet på skolen, som en kompetanseutvikling, og som en type kollektiv kompetanseutvikling på skolen».

Avdelingsdirektøren var også opptatt av eksamens «kalibrerende funksjon». Avdelingsdirektøren presiserte at «det er jo ingen regel om at det skal være likt», men framholdt at det er «helt tydelig» at eksamen har en kalibrerende funksjon, med henvisning til rapporter fra både Statistisk sentralbyrå og NIFU. Videre pekte avdelingsdirektøren på at de «alltid» har sett av karakterstatistikken at «i fag med sentralt gitt eksamen, så ligger standpunktvurderingen nærmere opp til eksamen, og karakternivået og karakterfordelingen er annerledes i de fagene, enn i de fagene som ikke har sentralt gitt eksamen», hvilket tilsier at den «åpenbart» har en kalibrerende funksjon.

Avdelingsdirektøren ble også spurt om eksamens rolle i skoloring og den kalibrerende funksjonen formelt var definert noe sted som et formål med eksamen, og om Utdanningsdirektoratet forholder seg til disse rollene som definerte formål.

Det kommer litt an på hva du mener med definert (...). Men vi er jo veldig tydelig på at sensorskoloring er en viktig kompetanseheving, og at å være sensor gir en verdifull kompetanse. Men det er jo ikke definert i forskriften på samme måte som vi sier at formålet med eksamen er definert i forskriften.

Av svaret kommer det klart fram at det ikke forelå slike andre definerte formål, til tross for at Utdanningsdirektoratet forholder seg til dette som helt sentrale sider ved eksamenssystemet.

Vi var interessert i å få vite mer om denne forskjellen mellom definerte formål og ytterligere roller eksamenssystemet har. Avdelingsdirektøren ble derfor stilt oppfølgingsspørsmål om hvordan de oppfattet eksamens rolle utover det som er definert i stortingsmeldinger og andre styringsdokumenter. Avdelingsdirektøren bekreftet at det er en del roller ved eksamenssystemet som ikke er definert noe sted, og sa at dette kan være fordi de er nøye med å fremheve hva som er «hovedformålet». Avdelingsdirektøren framholdt imidlertid at det er viktig «å beskrive og utnytte de andre funksjonene på sett og vis».

Avdelingsdirektøren pekte også på at statistiske undersøkelser viser at eksamen har en kalibrerende funksjon, men at dette som et styringsverktøy tidligst kan skje året etter, og ikke på standpunkt karakteren til den enkelte elev samme år. Dette til forskjell fra Sverige, der «nationella prov» har en direkte kalibrerende funksjon for karakterene elevene får av læreren.

Avdelingsdirektøren understrekte hvor stor vekt Utdanningsdirektoratet «i alle sammenhenger» legger på at skolene bør bruke eksamensresultatene til å diskutere egen karaktersettingspraksis:

Karakterstatistikken er et godt utgangspunkt for å diskutere og vurdere sin egen praksis på skolen. Er det store systematiske avvik over tid? Er det noen fag vi har mer avvik enn andre? Hva har dette å si? Hvordan ligger vi i forhold til andre skoler, i forhold til andre fylker? Så det er jo en sånn type funksjon eksamen har, i den forstand at den gir oss en nasjonal statistikk som er noe annet og mer pålitelig enn annen type statistikk, som ikke er nasjonalt forankret i den forstand.

Avdelingsdirektøren var altså tydelig på at eksamensresultater kan brukes lokalt som grunnlag for å reflektere over skolenes praksis for fastsetting av standpunktkarakterer, som vi har omtalt som eksamens rolle i å *styre* skolenes praksiser for *sertifisering*. Videre pekte også avdelingsdirektøren på eksamens rolle i å *støtte* læring og undervisning gjennom kompetanseheving og utarbeiding av veiledninger med kjennetegn for måloppnåelse. Alt dette kommer altså i tillegg til det avdelingsdirektøren identifiserte som det formelt definerte formålet med eksamen, som kan sammenfattes som at eksamen er *en kvalitetssikret sertifiseringsprøve*.

Samlet viser intervjuene med policyutviklere at eksamen ble oppfattet å ha en rekke ulike roller, knyttet særlig til *styring* men også *støtte*, til tross for at det formelle formålet er begrenset til *sertifisering*. I siste del av den empiriske analysen skal vi se på gjeldende forskrifter og rammeverket for eksamen som kom i 2017, fire år etter intervjuene ble gjort.

Analyse av gjeldende forskrifter og rammeverk for eksamen

Gjeldende Forskrift til opplæringslova slår fast at «sluttvurderinga» – standpunktkarakterer og eksamenskarakterer – «skal gi informasjon om kompetansen til eleven (...) ved avslutninga av opplæringa i fag i læreplanverket» (§3-17) og at eksamenskarakteren skal fastsettes «på individuelt grunnlag og gi uttrykk for kompetansen til eleven (...) slik denne kjem fram på eksamen» (§3-25). Forskrifta gir ikke ytterligere opplysninger om hva som er formålet med eksamen.

Rammeverket for eksamen av 2017 beskriver formålet med eksamen som todelt: «Kandidaten skal få anledning til å vise sin kompetanse i samsvar med læreplanen, og eksamenskarakteren skal gi informasjon om kandidatens individuelle kompetanse i faget, slik den ble uttrykt på eksamensdagen».

I rammeverket beskrives arbeidet med organiseringen av eksamen. Her kommer det fram at «oppmennene er med på å skape nasjonalt tolkningsfellesskap gjennom deltakelse på forhåndssensuren». Det kommer her tydeligere fram at eksamenssystemet og opplegget med forhåndssensur skal bidra til å bedre sammenlignbarheten i lærernes vurderinger generelt (standpunkt), og ikke bare på eksamen. Dette kan knyttes til

sertifiseringsrollen, men det er også tydelig at Utdanningsdirektoratet bruker eksamenssystemet til å styre og forbedre lærernes vurderingspraksis generelt. I omtalen av sensurordningene kommer det fram at «det skal være rulling av sensorer i alle fag over tid for å sikre at flere får ta del i sensorskoleringen». Dette er også uttrykk for at eksamen anses å ha en rolle i kompetanseheving og kvalitetsutvikling både av individuelle lærere og av skolene som organisasjoner.

Rammeverket beskriver også hvordan det for hver sentralt gitte eksamen utarbeides eksamensrapporter. Her presiseres det at «eksamen utgjør én vurderingssituasjon i motsetning til standpunktvurderingen som er basert på flere ulike situasjoner». Likevel framholdes det at eksamensrapporten kan «gi skolen informasjon som kan brukes i planleggingen av opplæringen og gi lærere et utgangspunkt for å diskutere hva kompetanse i faget kan være, og kan brukes i samtaler med elevene». Rammeverket gir ingen avklaring på samspillet mellom eksamen og standpunktvurderinger.

Sammenfatning og drøfting av eksamens roller i sertifisering, styring og støtte av læring og undervisning

Vi vil nå sammenfatte funnene av rollebeskrivelser i de fire datakildene tematisk med utgangspunkt i det teoretiske rammeverket knyttet til rollene *sertifisering*, *styring* og *støtte* av læring og undervisning. Avslutningsvis drøftes de overlappende rollene knyttet til «*kalibrering av standpunktvurderingen*», begrepet «*kvalitetssikring*» og eksamenssystemets «*trekkordning*».

Eksamens rolle i å sertifisere læring og undervisning

Å *sertifisere læring og undervisning* kommer til uttrykk på (minst) to måter i datamaterialet:

Eksamen som sluttvurdering (karaktersetting). Denne rollen ved eksamen er velkjent, og reflekterer den offisielle rollen til både eksamen og standpunkt, slik dette kommer til uttrykk i Forskrift til opplæringslova og i Rammeverk for eksamen. Denne rollen reflekteres også i ekspertintervjuene og i policydokumentene som er inkludert i analysen. Både eksamen og standpunkt karakter skal reflektere en vurdering av den faglige kompetansen eleven besitter ved slutten av opplæringen i faget. Særlig intervjuet med avdelingsdirektøren belyser det komplekse samspillet mellom eksamens rolle i sluttvurdering og lærernes karaktersetting (standpunktvurdering). Eksamens rolle i å «kalibrere» standpunktvurderingen drøftes som et eget punkt nedenfor.

Eksamensresultater som grunnlag for sertifisering og seleksjon. Den rollen karakterene fra sluttvurderingen har for sertifisering av kompetanse

som grunnlag for opptak til høyere utdanning og videre yrkesliv omtales ikke eksplisitt i de dokumentene som er omtalt og analysert i denne artikkelen. Men den er utførlig beskrevet i det store antallet forskrifter som regulerer virksomheten i universitets- og høyskolesektoren. Det er dessuten en av eksamens roller som kommer opp i ekspertintervjuene. Rollen er også nevnt eksplisitt i flere punkter i kapittel 6 i Forskrift til opplæringslova, som omtaler rangering av søkere til videregående utdanning.

Eksamens rolle i å styre læring og undervisning

Rollen *å styre læring og undervisning* kommer til uttrykk på (minst) to måter i datamaterialet:

Eksamensdata som grunnlag for beslutningstakers kvalitetsutviklingstiltak. Denne rollen er eksplisitt beskrevet i vårt analytiske rammeverk foran mer generelt som «Å evaluere (aggregerte) elevdata for måloppnåelse; brukt til å gi grunnlag for beslutningstakers kvalitetsutviklingstiltak» (s. 5). Denne rollen kommer til uttrykk i alle de årlige rapportene til Utdanningsdirektoratet, men vektlegges ikke særlig i stortingsmeldingene eller i intervjumaterialet. I en evaluering av det nasjonale kvalitetsvurderingssystemet (NKVS) fastslår også Kvåle, Langfelt og Skov (2009) at eksamensresultater brukes som informasjonsgrunnlag for skolens kvalitetsarbeid, spesielt i videregående skole. Forskrift til opplæringslova kan også oppfattes som implisitt å tillegge eksamen en slik rolle, gjennom å pålegge skoleeiere å «medverke til å etablere administrative system og å innhente statistiske og andre opplysninger som trengst for å vurdere tilstanden og utviklinga innanfor opplæringa» (§2.2). Videregående opplæring har få andre statistiske opplysninger som kan brukes for å vurdere tilstanden knyttet til læringsresultater, og det er derfor naturlig at eksamen blir en kunnskapskilde i dette arbeidet.

Eksamen som kontroll av praktiseringen av læreplaner, forskrifter og retningslinjer. I Kunnskapsdepartementet var både statssekretæren og seniorrådgiveren opptatt av at eksamen bidrar til å klargjøre hva som skal måles. Sistnevnte var særlig opptatt av kontroversene rundt dette. Eksamen er et politikkområde det strides om i Stortinget ved reformer der ulike fag-, språk- og profesjonsinteresser konkurrerer om hva som gjennom eksamenssystemet anerkjennes som viktig kunnskap. En slik tilbakevirkende tilsiktet effekt, kjent som «positiv washback» i den internasjonale litteraturen som drøfter testsystemer, er kjent fra flere land (se Nordenbo mfl., 2009) selv om det empiriske belegget for begrepet er begrenset (Fjørtoft & Sandvik, 2016). Det er påfallende at denne rollen ved eksamen verken uttrykkes i stortingsmeldingene, i Utdanningsdirektoratets rapporter eller av deres direktører. Vi antar det har bakgrunn i at eksamens rolle i dette er åpenbar og tas for gitt. Det vil være en fordel om dette presiseres som et formål ved eksamen da det vil kunne utgjøre et formelt grunnlag for å

vurdere om eksamenssystemets innretning er hensiktsmessig for å ivareta dette formålet, eller om formålet bedre kan ivaretas på andre måter enn i dag.

Eksamens rolle i å støtte læring og undervisning

Rollen å støtte læring og undervisning kommer også til uttrykk i datamaterialet på (minst) to måter:

Eksamens rolle i å støtte elevenes læring. Denne rollen handler om å identifisere og kommunisere avvik mellom aktuell og forventet måloppnåelse til elevene, som grunnlag for valg av og utvikling av læringsstrategier. Vi merker oss at Utdanningspeilet i 2008 framholdt at vurderingsveiledningene til eksamenssensuren «skal ha ein læringsfremjande effekt ved at lærarar kan formidle kjenneteikna til elevane før eksamen» (s. 105). Statssekretæren var også opptatt av eksamens rolle i å klargjøre forventningene for elevene.

Eksamens rolle i å støtte lærernes undervisning. Denne rollen handler om å identifisere og kommunisere avvik mellom aktuell og forventet måloppnåelse til lærerne, som grunnlag for å valg av og utvikling av undervisningsstrategier. Dette ble tatt opp i ekspertintervjuene av 2013 og i policydokumentene de etterfølgende årene. De to siste stortingsmeldingene, og likeledes seniorrådgiveren i Kunnskapsdepartementet og avdelingsdirektøren i Utdanningsdirektoratet, la stor vekt på eksamens rolle i kompetanseheving av sensorene. Hovdhaugen mfl. (2018) dokumenterer gjennom kvalitative intervjuer at skoleledere og lærere oppfatter deltakelse i sensur som svært viktig for den profesjonelle utviklingen til både skolen som organisasjon og den individuelle lærer. Denne rollen eksamenssystemet har i å styrke profesjonens forståelse av hva som forventes i fagene, synes å ha vært underkommunisert i policydokumentene over mange år. Vi merker oss at Utdanningsdirektoratet løftet det fram som et sentralt tema i Rammeverket for eksamen av 2017. Dette bekrefter inntrykket av at forståelsen av eksamens betydning i så henseende har økt de seneste årene. Etter vår mening bør dette lede til at kompetanseheving defineres som et formelt formål ved eksamenssystemet, slik at dets innretning kan evalueres og utvikles med dette som et klart definert siktemål med konsekvenser for politiske avveininger og økonomiske prioriteringer. En slik avklaring vil også tydeliggjøre konflikter mellom denne og andre roller vi finner i dagens eksamenssystem. Dette vil således legge viktige premisser for arbeidet med å videreutvikle eksamenssystemet i fagfornyelsen.

Bruk av eksamensresultater til kalibrering av skolenes standpunktverdinger

Som det framgår av analysedelen er eksamens «kalibrerende funksjon» eller «kalibrerende rolle» et noe uklart fenomen som omtales i policydokumentene. I analysen omtalte vi dette i termer av eksamens rolle i

å *styre* skolenes praksiser for *sertifisering*. Utdanningsdirektoratets analyser i Utdanningsspeilet gir årlige beskrivelser av samsvaret mellom eksamens- og standpunktkarakterer, som viser ulike praksiser for karaktersetting mellom fag på nasjonalt nivå. Avdelingsdirektøren i Utdanningsdirektoratet ga også uttrykk for at dette er en rolle som eksamen kan forventes å ha. I forbindelse med dette framhevet avdelingsdirektøren også at praksis i Norge, hvor lærere setter sine standpunktkarakterer før de får eksamensresultatene til elevene, er motsatt av ordningen i Sverige. Der forventes resultatene på de eksterne prøvene å tillegges betydelig vekt når læreren setter karakter (se Gustafsson & Erickson, 2018). Eksamen i Norge er imidlertid tilsynelatende ment å ha en *indirekte* kalibrerende rolle, ved at skolene sammenligner sine eksamensresultater med standpunktkarakterene og over tid justerer for avvik.

Ser vi på omtalen av dette i policydokumentene over tid ser det ut som det i 2012-2013 skjedde en dreining i Utdanningsdirektoratet forståelse, i tråd med det som kom til uttrykk i ekspertintervjuene. Dette reflekteres i rapporten fra 2014, der det ble lagt større vekt på slike sammenligninger. I den siste stortingsmeldingen i vårt materiale, Stortingsmelding nr. 28 (2015-2016), ser vi en ny dreining, der skoleeiere og skolers bruk av eksamenskarakterene til å evaluere egen karakterpraksis problematiseres. Problemstillingen blir imidlertid ikke fulgt opp i senere stortingsmeldinger, så dette anslaget til en mer omfattende belysning av eksamens roller ser ut til å ha blitt lagt til side.

Avdelingsdirektøren ga også uttrykk for at «eksamenskarakterer og karakterstatistikken er en del av kunnskapsgrunnlaget» og utgangspunkt for å diskutere og vurdere egen praksis på skolen, til tross for at en slik bruk av data ikke er spesifisert i regelverket. Disse underliggende rollene ved eksamenssystemet, og gyldigheten av slik bruk av eksamensdata, er det stort behov for å utforske nærmere, slik Hovdhaugen, Prøitz og Seland (2018) gjør. Kommunerevisjonen i Oslo (2013) sin gjennomgang og sammenligning av standpunktkarakterer og eksamenskarakterer i fagene norsk skriftlig (hovedmål), engelsk skriftlig og matematikk for årene 2006/07 til 2011/12 baserer seg også på en forståelse av eksamensresultater som et korrektiv til standpunktvurderingen, som Hovdhaugen mfl. problematiserer. Deres analyse er den til nå mest omfattende empiriske dokumentasjonen på at slike praksiser er blitt vanlig, og belyser også grundig hvorfor denne bruken av eksamensresultater kan være problematisk.

Eksamen som «kvalitetssikring»

Vi finner også grunn for å diskutere eksamens rolle i «kvalitetssikring» og bruken av dette begrepet i eksamenssammenheng. I Stortingsmelding nr. 30 (2003-2004) ble eksamen omtalt som «spesielt kvalitetssikrede prøver som er sentralt eller lokalt/regionalt gitt» (s. 37). Rammeverket for eksamen av 2017 representerer en spesifisering av hvordan det arbeides med å sikre

kvaliteten til disse prøvene. Her er det selve utformingen og administreringen av prøvene som omtales som kvalitetssikret. En noe annen betydning av begrepet kvalitetssikring finner vi imidlertid i Stortingsmelding nr. 28 (2015-2016) hvor det står at «Eksamen er en form for kvalitetssikring for elevene fordi de får en ekstern vurdering av sin fagkompetanse» (s. 62). Her forbindes kvalitetssikringsbegrepet med den uavhengige vurderingen, og at man i «fag med sentralt gitt eksamen har identiske oppgavesett for alle som kommer opp i faget» (s. 62). Også i Utdanningsspeilet av 2011 er det den eksterne sensuren som forbindes med kvalitetssikring: «Sidan eksamen blir gjennomført med ekstern sensur, har eksamen eit element av ekstern kvalitetssikring i seg» (s. 112).

Forståelsen av eksamen som kvalitetssikring bør imidlertid også ses i sammenheng med punktet over, eksamens «kalibrerende rolle». Av datamaterialet vårt framstår det implisitt at denne kalibreringen oppfattes å bidra til kvalitetssikring, altså at eksamenssystemet representerer en kvalitetssikring av standpunkt-karaktersettingen. Som vi har vært inne på ovenfor er det grunn til å reise spørsmål ved validitetsgrunnlaget for sammenligning av to vidt forskjellige typer vurderingsformer med delvis ulike kriterier for vurderingen. Når politikere og andre policyutviklere omtaler eksamen som en «kvalitetssikring» er det derfor grunn for å spørre: *Hvordan* bidrar eksamen til kvalitetssikring? Og hva er det som kvalitetssikres?

Er svaret at eksamen bidrar til *kvalitetssikring av lærernes standpunkt-vurdering*, så vil vi minne om utfordringene med en slik praksis omtalt over på enkeltlærer- eller skolenivå. Dette *kan* gjøres på forsvarlige og hensiktsmessige måter om man har tilstrekkelig innsikt i de statistiske mulighetene og begrensningene ved aggregerte dataserier for små utvalg av elever. Imidlertid kan slike sammenlikninger være mer meningsfulle for data aggregert til høyere nivåer (større kommuner, fylker og nasjonalt nivå). I så fall kan man eksempelvis undersøke om man over tid observerer betydelige skjevheter mellom standpunkt- og eksamenskarakterer på tvers av fag, mellom skoler med ulike karakteristikk, mellom kjønn etc. Og dersom slike skjevheter observeres, bør man drøfte om slike forskjeller er akseptable og rimelige.

Dersom man med kvalitetssikring mener å undersøke kvaliteten til eksamensprøvene per se, bør man definere hva slags prinsipper, prosedyrer og rammer for kvalitetssikring som bør anvendes. Rammeverket for nasjonale prøver gir klare føringer for dette, men det er etter hva vi kan se ikke gitt i Rammeverket for eksamen av 2017. Eksamen er imidlertid også prøver som det er rimelig å forvente etterlever krav om reliabilitet og validitet. Et mulig svar på hva som her menes med kvalitetssikring kan eksempelvis være at man rutinemessig innfører studier av oppgavers kvalitet fra et psykometrisk rammeverk i tråd med slik dette er formulert i

rammeverk for nasjonale prøver. I tillegg kan det vurderes å innføre mer systematiske rutiner for å estimere effekten (for elevene) av at sensorene gir ulike karakterer til samme besvarelser.

Vi vil stille spørsmål ved om bruken av begrepet *kvalitetssikring* er dekkende for praksisene vi har avdekket i denne artikkelen. Eksamen som *kvalitetsutvikling* kan være en mer dekkende for det vi har observert i denne studien. Bruk av eksamen som grunnlag for kvalitetsutvikling i sektoren er i ferd med å bli utbredt, og det kan se ut som at en slik rolle tas for gitt av aktørene i systemet. Gjennom å definere også dette tydeligere som et formål for eksamen, vil man kunne få økt bevissthet knyttet til muligheter og begrensninger ved en slik bruk. Det er eksempelvis ikke uproblematisk at karakterstatistikk for små grupper som trekkes ut fra den enkelte skole, brukes som kvalitetsindikator. Videre kan avklaring av dette som et formål bidra til å belyse problematisk bruk av resultater fra eksamen som indikator for utvikling over tid for den enkelte skole eller for hele utdanningssystemet. Til dette svinger resultatene for mye (Hovdhaugen mfl., 2014, 2018), noe som med stor sannsynlighet skyldes at oppgavene fra ett år til det neste ikke har lik vanskegrad.

Eksamens trekkordning

Trekkordningen i eksamenssystemet innebærer at elever blir fordelt til å delta i ulike eksamener. I all hovedsak skal dette være basert på tilfeldige utvalg, men lokalt må man også ta hensyn til at lærere har bestemte elevgrupper i flere fag, at de har flere grupper med eksamen på samme dag og at de kan være oppnevnt som ekstern sensor ved andre skoler de samme dagene. Dette fører til store variasjoner, hvilket elevene opplever som urettferdig³. Denne trekkordningen hviler på et prinsipp som er lite omtalt i policydokumentene. Trolig er det så institusjonalisert i vårt utdanningssystem at vi tar det for gitt. Prinsippet går fram av Stortingsmelding nr. 20 (2012-2013): «Trekkordningen innebærer at elevene ikke skal ha eksamen i alle fag, men at de skal være eksamensforberedt i de fagene der eksamen er en mulig sluttvurdering ved siden av standpunktkarakteren» (s. 65-66). Gjennom dette får eksamen også en ikke formalisert, men likefullt reell rolle: Vissheten om at det kan komme en eksamen i faget skal bidra til å holde elevene på «tå hev». Eksamen har således også en rolle som et ytre press – som skal bidra til at elevene arbeider godt i alle fagene i skolen. Denne rollen er etter hva vi kan se svært lite belyst i policydokumenter og forskning, og bør undersøkes nærmere. For å sikre en mer rettferdig trekkordning burde man vurdere å balansere de praktiske hensynene mot noen prinsipper for trekking fastsatt i rammeverket for eksamen.

³ <https://www.utdanningsnytt.no/nyheter/2018/november/elevorganisasjonen-vil-endre-eksamensordningen/>

Avslutning: Behov for avklaring av eksamenssystemets mange roller

I denne studien har vi identifisert mange roller eksamenssystemet har, knyttet til sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring. Samlet sett må vi likevel konkludere med at eksamenssystemets roller ikke har vært utfyllende omtalt i noen av stortingsmeldingene i perioden. Dermed har eksamen heller ikke vært plassert høyt på den politiske agendaen i den skolepolitiske diskursen. Analysen av stortingsmeldingene viser at det politisk har vært svært lav bevissthet om eksamens roller utover *sertifisering* i den undersøkte perioden fra 2003 til 2017. I den siste stortingsmeldingen, Stortingsmelding nr. 28 (2015-2016), problematiseres det imidlertid hvordan eksamensdata brukes av skoleeiere og skoler til å evaluere egen karakterpraksis. Dette er et uttrykk for at styringsrollen tar større plass. Problemstillingen har imidlertid ikke blitt fulgt opp i senere stortingsmeldinger, så dette lille anslaget til en mer omfattende belysning av eksamens roller ser ut til å ha blitt lagt til side.

Analysen av de årlige rapportene til Utdanningsdirektoratet viser at det i den administrative styringen av utdanningssystemet er større fokus på de mange ulike rollene eksamen har. Det kommer tydelig fram at det norske eksamenssystemet opererer i et tett samspill mellom nasjonale myndigheter og lærerprofesjonen. Dermed skiller det seg vesentlig fra mange land der en viktig rolle ved eksamen er ansvarsstyring (accountability) av lærerprofesjonen (Black, 2016). Legitimiteten til eksamenssystemet hviler trolig i stor grad på dette samspillet mellom aktørene i skolen. Det er derfor påfallende og svært uheldig at eksamens mange roller ikke er tydeligere definert som formålsbeskrivelser i lover og forskrifter.

Prinsippene eksamenssystemets trekkordning hviler på er blant de mest institusjonaliserte fenomenene ved eksamenssystemet, som vi også etterlyser at tydeliggjøres i mer formelle former. Denne åpenbare, men lite uttrykte, siden ved eksamenssystemet, er symptomatisk for vår overordnede konklusjon: Det finnes en rekke roller ved grunnopplæringens eksamener, som burde tydeliggjøres og defineres som formelle formål ved eksamenssystemet. Gjennom en slik tydeliggjøring kan eksamen utvikles videre til å bli et verktøy som er bedre tilpasset disse mange rollene. Alternativt kan man også konkludere med at det er vanskelig, og kanskje umulig, for eksamen å skulle ha så mange roller. I så fall vil det å gjøre disse rollene til eksplisitte definerte formål kunne bidra til en nødvendig diskusjon om endringer som bør gjøres i eksamenssystemet for å ivareta de mange hensynene som må tas for å ivareta sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring.

Om forfatterne

Sverre Tveit er universitetslektor ved Universitet i Agder. Hans forskningsinteresser omfatter blant annet policyutvikling og -legitimering, vurdering og evaluering både i høyere utdanning og grunnsopplæringen.

Institusjonstilknytning: Institutt for pedagogikk, Universitetet i Agder, Postboks 422, 4604 Kristiansand.

E-post: sverre.tveit@uia.no

Rolf Vegar Olsen er professor ved Universitet i Oslo. Hans forskningsinteresser omfatter blant annet hvordan vi kan måle elevenes kunnskaper, ferdigheter og holdninger.

Institusjonstilknytning: Centre of Educational Measurement, Universitetet i Oslo, Postboks 1161 Blindern, 0318 Oslo.

E-post: r.v.olsen@cemo.uio.no

Referanser

- Assessment Reform Group (2002). *Assessment for learning: 10 principles: Research-based principles to guide classroom practice*. London: Assessment Reform Group.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. doi:10.1080/0969594X.2010.513678
- Black, P. (1998). *Testing: Friend or foe?* London, England: Falmer Press.
- Black, P. (2016). The Role of Assessment in Pedagogy – and Why Validity Matters. In Wyse, D., Hayward, L. & Pandya, J. (Eds.). *The SAGE Handbook of Curriculum, Pedagogy and Assessment*. London: Sage knowledge.
- Black, P., og Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), s. 7–74. doi:10.1080/0969595980050102
- Black, P., og Wiliam, D. (1998b). Inside the blax box. Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 92(1), 81–90.
- Black, P. og Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(5), s. 5-31. doi: <https://doi.org/10.1007/s11092-008-9068-5>
- Bloom, B. S., Hasting, J. T., og Madaus, G. F. (1971). *Handbook on formative and summative evaluation of student learning*. New York, NY: McGraw-Hill.
- Burner, T. (2016). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research*, 60(6), s. 626–648. doi:10.1080/00313831.2015.1066430
- Bogner, A. og Menz, W. (2009). The Theory-Generating Expert Interview: Epistemological Interest, Forms of Knowledge, Interaction, s. 43-80. In: Bogner, A., Littig, B. and Menz, W (Eds). *Interviewing Experts*. Hampshire: Palgrave Macmillan.
- Dexter, L. A. (2006). *Elite and specialized interviewing*. Colchester: ECPR Press. (Original work published 1970).
- Fjørtoft, H. og Sandvik, L. (2016). *Vurderingskompetanse i skolen*. Oslo: Universitetsforlaget.

- Fontana, A., og Frey, J. H. (2005). The Interview: From Neutral Stance to Political Involvement. In N. K. Denzin & Y. S. Lincoln (Eds.). *The Sage handbook of qualitative research*. (3rd ed), (s. 695-728). Thousand Oaks, CA: Sage.
- Forskrift til opplæringslova (2006). Oslo: Kunnskapsdepartementet, <https://lovdata.no/dokument/SF/forskrift/2006-06-23-724>
- Gamlem, S. M. (2015). Feedback to support learning: Changes in teachers' practice and beliefs. *Teacher Development*, 19(4), s. 461–482. doi:10.1080/13664530.2015.1060254
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), s. 150–169. doi:10.1080/0969594X.2012.749212
- Gravaas, B. C., Hægeland, T., Kirkebøen, L. J., & Steffensen, K. (2008). *Skoleresultater 2007*. Oslo: Statistisk Sentralbyrå.
- Gustafsson, J-E. og Erickson, G. (2018). Nationella prov i Sverige – tradition, utmaning, förändring. *Acta Didactica* 12(4). doi: <http://dx.doi.org/10.5617/adno.6434>
- Herman, J. L., og Baker, E. L. (2009). Assessment policy: Making sense of the babel. In G. Sykes, B. Schneider, og D. Plank (Eds.). *Handbook of Education Policy Research*, pp. . Newbury Park, London: Sage.
- Hopfenbeck, T. N., Flórez Petour, M. T., og Tolo, A. (2015). Balancing tensions in educational policy reforms: Large-scale implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy & Practice*, 22(1), s. 44–60. doi:10.1080/0969594X.2014.996524
- Hovdhaugen, E., Prøitz, T.S. og Seland, I. (2018). Eksamens- og standpunktkarakterer – to sider av samme sak? *Acta Didactica* 12(4). doi: <http://dx.doi.org/10.5617/adno.6276>.
- Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T.S og Vibe, N. (2014). *Karakterer i offentlige og private videregående skoler: En analyse av eksamens- og standpunktkarakter i norsk og matematikk og rutiner for standpunktvurdering i offentlige og private videregående skoler*. Oslo: NIFU.
- Hægeland, T., Kirkebøen, L.J og Raaum, O. (2005): *Skoleresultater 2004. En kartlegging av karakterer frå grunn og videregående skoler i Norge*. SSB-notat 2005/31. Oslo: SSB.
- Jonsson, A., Lundahl, C., og Holmgren, A. (2015). Evaluating a large-scale implementation of assessment for learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, 22(1), 104–121. doi:10.1080/0969594X.2014.970612
- Kommunerevisjonen Oslo kommune (2013). *Standpunktkarakterer i videregående skole – likebehandles elevene?* Rapport 17/2013. Oslo: Kommunerevisjonen.
- Kvåle, G., Langfeldt, G. & Skov, P. (200). *Evaluering af Det nationale kvalitetsvurderingsystemet for grunnsopplæringen (NKVS) – noen aktuelle funn og fortolkninger*. Kristiansand/Aarhus: Agderforskning/DPU Aarhus Universitet.
- Matematikksenteret (2016). *Vurdering av eksamen i matematikk*. Trondheim: Matematikksenteret. Lastet ned 12.10.2018 fra: https://www.matematikksenteret.no/sites/default/files/attachments/resource/vurdering_av_eksamen_i_matematikk.pdf
- Newton, P. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.
- Nordenbo, S. E., Allerup, P., Andersen, H. L., Dolin, J., Korp, H., Larsen, M. S., . . . Østergaard, S. (2009). *Pædagogisk brug af test - Et systematisk review*. København: Danmarks Pædagogiske Universitets Forlag.

- Prøitz, T. S., og Spord Borgen, J. (2010). *Rettferdig standpunktbedømming – det (u)muliges kunst?* Oslo: NIFU STEP.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), s. 119–144.
- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (Vol. 1, s. 39–83). Chicago, IL: Rand McNally, American Educational Research Association Monograph Series on Curriculum Evaluation.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.
- Stortingsmelding nr. 30 (2003–2004). *Kultur for læring*. Oslo: Utdannings- og forskningsdepartementet.
- Stortingsmelding nr. 16 (2006–2007). *Og ingen stod igjen. Tidlig innsats for livslang læring*. Oslo: Kunnskapsdepartementet
- Stortingsmelding nr. 20 (2012–2013). *På rett vei*. Oslo: Kunnskapsdepartementet
- Stortingsmelding nr. 28 (2015–2016). *Fag – fordypning – forståelse - en fornyelse av kunnskapsløftet*. Oslo: Kunnskapsdepartementet.
- Thronsdalen, I., Hopfenbeck, T., Lie, S. og Dale E.L. (2009). *Bedre vurdering for læring. Rapport fra evaluering av modeller for kjennetegn på mål i fag*. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- Tveit, S. (2017). *50 Years' Legacy of Formative and Summative Evaluation and Assessment: A Critical Theoretical Review of Education Policy and Research*. Innlegg presentert ved European Conference on Educational Research, København. Sammendrag hentet fra <https://eera-ecer.de/ecer-programmes/conference/22/contribution/41939/>.
- Tveit, S. (2018a). 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), s. 327–350. doi: <https://doi.org/10.1080/0969594X.2017.1421522>.
- Tveit, S. (2018b). Transnational Trends and Cultures of Educational Assessment: Reception and Resistance of National Testing in Norway and Sweden during the Twentieth Century. In Alarcon, C. & Lawn, M. (eds.). *Assessment Cultures*. Berlin: Peter Lang. doi: <https://doi.org/10.3726/978-3-653-06867-2>.
- Utdanningspeilet (2008). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet.
- Utdanningspeilet (2009). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet.
- Utdanningspeilet (2011). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Rapport hentet fra: https://www.udir.no/globalassets/filer/tall-og-forskning/forskningsrapporter/utdanningspeilet_2011.pdf
- Utdanningspeilet (2014). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Rapport hentet fra: <http://utdanningspeilet.udir.no/2014/wp-content/uploads/2014/06/Utdanningspeilet-2014-utskriftsversjon.pdf>
- Utdanningspeilet (2016). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Rapport hentet fra: http://utdanningspeilet.udir.no/2016/wp-content/uploads/2016/06/Utdanningspeilet_2016.pdf
- Utdanningspeilet (2017). *Tall og analyse av grunnsopplæringen i Norge*. Oslo: Utdanningsdirektoratet. Rapport hentet fra: <http://utdanningspeilet.udir.no/2017/innhold/del-5/5-2-grunnskolepoeng-og-karakterer-etter-10-trinn/>