

Elisabeth Hovdhaugen

NIFU

Tine Sophie Prøitz

Universitetet i Sørøst-Norge

Idunn Seland

NOVA, OsloMet – storbyuniversitetet

DOI: <http://dx.doi.org/10.5617/adno.6276>

Eksamens- og standpunktkarakterer – to sider av samme sak?

Sammendrag

I denne artikkelen undersøker vi hvilken funksjon den norske ordningen med henholdsvis eksamens- og standpunktkarakter har for læreres vurderingspraksis. Analysen bygger på nasjonal karakterstatistikk og intervjuer med lærere som underviser i norsk eller matematikk i videregående skole, og som har erfaring fra å vurdere elevenes sluttkompetanse i disse fagene ved hjelp av begge disse vurderingsordningene. Studien tar utgangspunkt i en utdanningspolitisk utvikling preget av forsterkede forventinger om å se standpunktkarakterer og eksamenskarakterer i sammenheng. Dette er uttrykt som at det bør være et visst samsvar mellom disse vurderingsordningene, til tross for at standpunkt- og eksamenskarakterene blir satt på ulike premisser og springer ut av svært ulike kontekster. Felles for begge vurderingsformer er imidlertid at eleven skal få demonstrert sluttkompetanse i et fag. Vi viser hvordan disse to vurderingsformene veies opp mot hverandre i lærernes egne vurderingspraksiser, og bygger videre på lærernes synspunkter i en diskusjon av i hvilken grad eksamenskarakteren bør anses som egnet til å kvalitetssikre læreres standpunktvurdering.

Nøkkelord: standpunktkarakter, eksamenskarakter, elevs sluttkompetanse, læreres vurderingspraksis

National examination grades and final classroom grades – two of a kind?

Abstract

In this article, we analyse differences between grades based on national examinations, and final grades based on summative classroom assessment at the end of upper secondary education, using data from two subjects: Norwegian language and advanced mathematics. We compare the two types of grades over

time and provide a statistical demonstration of the differences. Through interviews with teachers of Norwegian or mathematics in upper secondary education, we highlight the differences between grading based on summative classroom assessment and being an external examiner. We show how these two assessment methods are weighed against each other in the teachers' own assessment practices, before we finally discuss the extent to which teachers perceive the exam grades as a form of measure for quality assurance of the grade based on summative classroom assessment in the same subject.

Keywords: summative classroom assessment, national examination grades, students' final competence, teachers' grading practices

Innledning

En av de mest sentrale hensiktene med et lands karaktersystem er seleksjon til videre utdanningsnivå og arbeidsliv. For å kunne ivareta denne oppgaven på en god måte er man avhengig av at karaktersystemet har høy legitimitet. Legitimitet kan knyttes til om og i tilfelle i hvilken grad retningslinjer, rutiner, prosedyrer og praksis for karaktersettingen blir oppfattet som rettferdig og likeverdig. En viss grad av stabilitet, i form av både måleinstrumenter og resultater, kan forventes å stå sentralt i en folkelig oppfatning av karaktersystemets reliabilitet. I et system der prosedyrer og/eller praksis for karaktersetting oppfattes som påvirket av tilfeldige, variable eller systematisk urettferdige faktorer, vil tilliten til karaktersettingen følgelig kunne tape legitimitet.

I Meld. St. 28 (2015–2016) *Fag – fordypning – forståelse. En fornyelse av Kunnskapsløftet* skriver regjeringen om bruken av standpunkt-karakterer og eksamens-karakterer som mål på elevenes kompetanse ved avslutningen av henholdsvis grunnskolen og videregående opplæring. Mens eksamen viser elevens samlede kompetanse slik den kommer til uttrykk på eksamensdagen, skal standpunkt-vurdering gi informasjon om kompetansen eleven har oppnådd i faget ved avslutningen av opplæringen. Selv om begge disse vurderingsformene skal avspeile trekk ved det som i norsk utdanningslovgivning er betegnet som elevens «sluttkompetanse», er de på denne måten forskjellige. Dette kommer også til syne i hvem det er som vurderer elevene til henholdsvis standpunkt- og eksamens-karakter i et fag. Standpunkt-karakteren settes av elevenes faglærer og baserer seg på dokumentasjon av elevens ervervede kompetanse i faget over tid. Eksamen kan i norsk skole være utformet på enten lokalt eller nasjonalt nivå, og selve prøveformen kan være skriftlig, muntlig eller praktisk. Eksamenene som denne studien bygger på, er såkalt *sentralt gitt skriftlig eksamen*, som betyr at de er utformet av Utdanningsdirektoratet og gjennomføres over hele landet. Dette innebærer videre at hver eksamensbesvarelse skal sensureres av to oppnevnte sensorer i en ordning der sensor og elev er anonyme for hverandre. Sensuren

skjer i henhold til nasjonalt standardiserte retningslinjer nedfelt i blant annet forskrift til Opplæringsloven. I tillegg utarbeides en sensorveiledning som ofte har eksempler på vurderinger til hver eksamensoppgave.

Alle elever på utdanningsprogram for studiespesialisering skal ved avslutningen av videregående skole (Vg3) ta eksamen i norsk hovedmål. Alle andre eksamener i den norske grunnopplæringen er såkalte trekkfag, der bare en andel av elevmassen som har faget, også tar eksamen. Ifølge Kunnskapsdepartementet fungerer eksamen som «en form for kvalitetssikring for elevene fordi de får en ekstern vurdering av sin fagkompetanse» (Meld. St. 28, 2015–2016, s. 62). I det nasjonale kvalitetsvurderingssystemet for grunnopplæringen inngår eksamenskarakterer videre som ett av flere kvalitetsmål for skolene (Tveit, 2014). Av dette følger at det kan oppstå et ekstra behov hos skolene for å kvalitetssikre den standpunktvurderingen som elevenes lærere har gjort ved å sammenligne denne med eksamenskarakteren. Eksamen blir dermed gitt en standardsettende funksjon for læreres standpunkt-karaktersetting.

Avvik mellom karakterer gitt til standpunkt og eksamen er ikke en ny problemstilling i norsk grunnopplæring (Prøitz & Borgen, 2010; Galloway, Kirkebøen & Rønning, 2011; Opheim, Grøgaard & Næss, 2010). På nasjonalt nivå ligger resultatene fra skriftlig eksamen i norsk hovedmål vanligvis et halvt karakterpoeng under de gjennomsnittlige standpunkt-karakterene i faget (Vibe & Hovdhaugen, 2014). I stortingsmeldingen anbefaler departementet at vedvarende avvik mellom eksamen- og standpunkt-karakter «bør være et varsel til skoleeier og skoleleder om at det er nødvendig å endre eksisterende vurderingspraksis» (Meld. St. 28, 2015–2016, s. 63). Dette er et tydelig eksempel på hvordan vurdering, elevresultater og læringsutbytte har fått en langt mer sentral rolle i norsk utdanningspolitikk etter innføringen av Kunnskapsløftet (Mausethagen, Prøitz & Skedsmo, 2018). Karakterer som resultat av såkalte summative vurderingsformer får i denne sammenheng en mer fremtredende rolle i styringen av opplæringen og som grunnlag for utvikling i skolen, selv om det i meldingen også står at «en eksamenskarakter vil være satt på et mer begrenset vurderingsgrunnlag enn en standpunkt-karakter» (Meld. St. 28, 2015–2016, s. 63). Et relevant spørsmål knyttet til dette er hvordan et slikt signal tas imot av lærere i deres arbeid med karaktersetting, og hvordan de erfarer og fortolker forskjeller mellom standpunkt- og eksamenskarakterer.

I denne artikkelen stiller vi derfor følgende spørsmål: *Hvilken funksjon har eksamenskarakterer og standpunkt-karakterer for læreres vurderingspraksis?* Vi undersøker dette spørsmålet med utgangspunkt i læreres karaktersetting i norsk og matematikk ved elevenes avslutning av studieforberevende utdanningsprogram i åtte videregående skoler. For å besvare hovedproblemstillingen har vi tatt i bruk følgende delspørsmål i vår analyse:

- Hvordan arbeider lærerne med å sette standpunkt-karakterer i norsk og matematikk?

- Hvordan arbeider lærerne med å sensurere eksamen i norsk og matematikk?
- Hvilken betydning tillegger lærerne forskjellen mellom de to vurderingsformene?

Det norske lovverket og nyere forskning om læreres vurdering

En omfattende del av den internasjonale litteraturen om vurdering diskuterer forholdet mellom formative og summative vurderingsformer, deres formål og de praksiser som følger av disse (se for eksempel Taras, 2005; Harlen, 2005; Wiliam & Black, 1996). Det finnes også mange studier om læreres ulike praksiser i summativ vurdering og ulike faktorer som påvirker læreres praksiser (se for eksempel Cheng & Sun, 2015; Tierney, 2015; McMillan, Myran & Workman, 2002). Disse studiene konkluderer gjerne med at det er mangel på indre konsistens mellom ulike læreres karaktersetning og at lærere trenger mer kompetanse i vurdering (se for eksempel Black, Harrison, Hodgen, Marshall & Serret, 2010; Brookhart, 2013). Det er færre bidrag i litteraturen som diskuterer forholdet mellom ulike former for summative vurdering, og som er relevante for vår studie. Studier som diskuterer forhold ved summativ vurdering, gjør gjerne dette knyttet til diskusjoner om karaktersetning i høyere utdanning ved hjelp av tester og kritikk mot å anse karaktersetning som presise målinger av elever eller studenters læring (se for eksempel Bloxham, 2009; Knight, 2002; Yorke, 2011).

Sentralt i nyere studier om læreres karaktersetning står samtidig spørsmål om hvordan myndigheters økte bruk av ulike typer av nasjonale tester eller eksamener for styring, kontrollformål og ansvarliggjøring (accountability) påvirker læreres praksiser for vurdering på nye måter. I Norge er dette et aktuelt spørsmål når forventninger til samsvar mellom standpunkt karakter og eksamens karakter i større grad er tatt i bruk for eksempel i skoleeiers styring av skolens arbeid med skoleutvikling (Aasen, Møller, Rye, Prøitz & Hertzberg, 2012; Mausethagen et al., 2018). Studier viser også at det finnes klare oppfatninger hos nasjonale og lokale myndigheter om at for store avvik mellom standpunkt og eksamen kan tyde på at lærere og skoler vurderer elever enten «for snilt» eller «for strengt» sammenlignet med karakterer gitt til eksamen. Eksamen blir sett som rettesnor og er på denne måten gitt en kalibrerende funksjon for standpunkt karakteren (se for eksempel Prøitz & Borgen, 2010; Hovdhaugen, Seland, Lødding, Prøitz & Vibe, 2014). Hvordan lærere forholder seg til slike forventninger, er imidlertid i mindre grad studert.

Eksamen og standpunkt er det Utdanningsdirektoratet kaller «sluttvurdering». I forskrift til Opplæringsloven¹ § 3-17 heter det at: «Sluttvurderinga skal gi informasjon om kompetansen til eleven, lærlingen, praksisbrev-

¹ <https://lovdata.no/dokument/NL/lov/1998-07-17-61>

kandidaten og lære kandidaten ved avslutninga av opplæringa i fag i læreplanverket, jf. § 3-3.» Videre er det fastsatt ved § 3-18 hvilken informasjon standpunkt karakteren skal gi, og med § 3-3 understrekes det at: «Grunnlaget for vurdering i fag er dei samla kompetansemåla i læreplanane for fag slik dei er fastsette i læreplanverket», jf. § 1-1 og § 1-3. § 3-18 viser også til at: «Standpunkt karakteren må baserast på eit breitt vurderingsgrunnlag som samla viser den kompetansen eleven har i faget», jf. § 3-3 og § 3-16. Med § 3-16 fastsettes også sammenhengen mellom såkalt underveisvurdering og standpunkt karakter i fag ved at vurderingen skal gi muligheter for forbedring og ses i lys av den underveisvurdering som eleven har gjennomgått i løpet av opplæringen i fag som skal føres på vitnemålet.

I forskrift til Opplæringsloven § 3-25 heter det at «Eksamen skal vere i samsvar med læreplanverket» og at «Eksamen skal organiserast slik at eleven eller privatisten kan få vist kompetansen sin i faget. Eksamens karakteren skal fastsetjast på individuelt grunnlag og gi uttrykk for kompetansen til eleven eller privatisten slik denne kjem fram på eksamen.» Gjennomføring av eksamen er også fastsatt i regelverket med opptil fem timer for skriftlig eksamen og opptil 30 minutter for muntlig eksamen. Regelverket angir som vist at det er klare forskjeller i det som ligger til grunn for fastsettelse av standpunkt karakterer og eksamens karakterer, selv om begge vurderingsformer skal vurdere eleven ut fra kompetansemålene for fag (§ 3-3).

På den annen side er det praksiser og tradisjoner knyttet til standpunkt og eksamen i ulike fag, som gjør at det kanskje kan forventes sterkere sammenheng mellom standpunkt og eksamen i enkelte fag, som for eksempel i norsk. Ikke minst har omfattende bruk av eksempel oppgaver med tilhørende sensorveiledninger som anbefales brukt i undervisning, underveisvurdering og som grunnlag for standpunkt karaktersetting, gjort at avvik eller samsvar mellom karakterer i faget norsk kan handle om hvor nære ulike skoler er til disse tradisjonene og praksisene (Hovdhaugen et al., 2014). Det er dermed et spørsmål om i hvilken grad dette er noe som kan slå ulikt ut i læreres prosedyrer for karaktersetting i ulike skoler.

I tillegg har studier av læreres praksiser for standpunkt vurdering vist at standpunkt ofte settes på et mer sammensatt grunnlag enn fag alene, noe som kan være i strid med forskrift til Opplæringsloven. Her understrekes det at vurderingen skal baseres på elevens kompetanse i fag og ikke på for eksempel oppmøte, innsats eller holdninger (Prøitz, 2013; Prøitz & Borgen, 2010; Hovdhaugen et al., 2014; Galloway et al., 2011). At lærere setter karakterer med utgangspunkt i andre egenskaper hos eleven enn fag, er en velkjent problemstilling i internasjonal forskning om vurdering. Ofte forklares en slik praksis med et manglende samsvar mellom læreres og skolemyndigheters oppfatninger om hva som utgjør et godt grunnlag for vurdering av prestasjoner (Angelo, 1998; Brookhart, 1991, 2013; Manke & Loyd, 1990; Stiggins, Frisbie & Griswold, 1989; Eggen, 2004; Resh, 2009). Det er for eksempel dokumentert at

innsats og aktivitet tas med i læreres karaktervurdering (Throndsen, Hopfenbeck, Lie & Dale, 2009; Prøitz & Borgen, 2010), og at elevgruppers sosiale bakgrunn også kan ha en påvirkning på læreres karaktersetting (Hægeland, Kirkebøen, Raaum & Salvanes, 2005). Wiborg, Arnesen, Grøgaard, Støren og Opheim (2011) fant at foreldres utdanning har større innvirkning på standpunkt-karakterer enn på eksamenskarakterer. Mye av forskningslitteraturen har lenge pekt på behovet for å øke vurderingskompetansen hos lærere for å bøte på praksis som ikke er i tråd med anbefalinger og regelverk (Brookhart, 2013). Prøitz og Borgen (2010) fant imidlertid at lærere i stor grad forsøkte å etterleve nye reguleringer for standpunkt-vurdering og ny læreplan som følge av implementeringen av Kunnskapsløftet, men at flere strevde med å få det til innenfor fagenes rammer. Nyere forskning viser også til at ulike kontekster krever ulike tilnærminger til karaktersetting, og antyder at det ikke alltid er mer kompetanse om vurdering lærere og skoler trenger. Isteden kan det være et behov for mer støtte og veiledning i arbeid med karaktersetting (Brookhart, 2013; Wyatt-Smith & Klenowski, 2013).

Som det går frem av denne gjennomgangen, er lærernes praksis for karaktersetting ulik, og oppfatningene om hvorvidt karakterene for standpunkt og eksamen ideelt sett skal være like eller hvor store avvik som er akseptable, kan som følge av dette variere. I det videre presenterer vi studiens metodiske tilnærming og datamateriale før vi presenterer selve analysen og studiens resultater. Vi har valgt å presentere resultatene strukturert etter de problemstillingene vi har søkt å besvare, før vi diskuterer våre funn opp mot vurderingslitteraturen vi har presentert.

Metode og data

Artikkelen tar utgangspunkt i to typer data: kvantitative data om karakterer i norsk hovedmål og realfagsmatematikk, og kvalitative data basert på intervjuer med lærere i norsk og matematikk. Alle data er samlet inn gjennom prosjektet *Karakterpraksis i offentlige og private videregående skoler*, som ble gjennomført i 2013–14 på oppdrag av Utdanningsdirektoratet. Prosjektet er dokumentert i en NIFU-rapport (Hovdhaugen et al., 2014).

Utgangspunktet for problemstillingen er den dokumenterte forskjellen mellom standpunktkarakter og eksamenskarakter. Vi anvender de kvantitative dataene for å belyse denne forskjellen, hvor vi også finner tegn til større stabilitet i den ene vurderingsformen enn i den andre. Både forskjellen mellom eksamens- og standpunktkarakter, og stabiliteten over tid i standpunktkarakterene begrunnet vår forskningsinteresse i å intervjuer lærere om deres vurderingspraksis.

Kvantitative data

I artikkelen brukes karakterdata fra Utdanningsdirektoratet, der det finnes informasjon om både standpunktkarakter og eksamenskarakter for elever på Vg3, i fagene norsk hovedmål og realfagsmatematikk (Matematikk R2)². Vi har data for fire årskull med avgangselever: elevene som startet sin videregående opplæring i årene 2007/08 til 2010/11, og dermed var avgangselever i perioden 2009/10 til 2012/13. Presentasjonen av data gjøres grafisk i diagrammer som viser gjennomsnittlig standpunkt- og eksamenskarakter i de to fagene. Her sammenligner vi dermed gjennomsnittlig standpunktkarakter med gjennomsnittlig eksamenskarakter for samme gruppe elever, og alle som er inkludert i data har dermed både standpunktkarakter og eksamenskarakter i faget det året.. Fra år til år er elevgruppene ulike, men siden vi har data for alle som har tatt eksamen og fått avsluttende standpunktkarakter i faget, har slike variasjoner lite å si.

Kvalitative data

Våren 2014 gjennomførte vi gruppeintervjuer med til sammen 16 matematikklærere og 16 norsklærere på åtte videregående skoler i det sentrale Østlandsområdet. Dette er lærere som har erfaring med å sette standpunktkarakter i enten matematikk eller norsk, men som ikke nødvendigvis underviste i faget eller skulle sette standpunkt den våren vi intervjuet på skolen. I den videre analysen refererer vi for enkelhets skyld til disse som henholdsvis «norsklærere» og «matematikklærere». På hver skole deltok lærere fra begge fag i gruppeintervjuet, hvor den primære hensikten var å få innsikt i lærernes rutiner for standpunktvurdering. Gjennom intervjuene ville vi belyse forholdet mellom disse rutinene og de nasjonale påleggene for vurdering i grunnopplæringen.

De åtte skolene var på forhånd valgt ut fordi de enten hadde et stort avvik (større enn 1,0) mellom eksamens- og standpunktkarakterer (fire skoler) eller fordi de hadde et lite avvik (mindre enn 0,4) mellom eksamens- og standpunktkarakterer (fire skoler) over en tidsperiode på fire år. På vår forespørsel hadde rektor ved hver skole rekruttert lærerne til gruppeintervjuet. Rektor var ikke på forhånd gjort kjent med årsaken til at skolen var valgt ut til intervju, altså at skolen var valgt ut fordi den hadde enten stort eller lite avvik mellom standpunkt- og eksamenskarakter. På den annen side hører det til rektors oppgave å kjenne til hvordan skolen hun leder, ligger an på slike parametere. Vi kan derfor ikke utelukke at lærergruppen vi intervjuet, var forberedt på å snakke om forskjellen mellom elevenes standpunkt- og eksamenskarakterer på bakgrunn av situasjonen ved skolen.

² Matematikk R2 er det mest avanserte kurset i matematikk som tilbys i videregående opplæring og gir ekstrapoeng ved søkning til høyere utdanning.

Uavhengig av dette fikk vi ved begynnelsen av hvert gruppeintervju opplysninger fra lærerne som bekrefter at utvalget kan fremvise god variasjon med hensyn til informantenes erfaring, ansiennitet, utdanning og fra hvilke fag de hadde undervisningsbakgrunn i tillegg til norsk eller matematikk. På hver av disse skolene ble det gjennom intervjuene klart at flere av lærerne hadde erfaring og skoloring som sensor, og disse lærerne fortalte – gjerne uoppfordret – om hvilken verdi de tilla denne erfaringen når de som faglærer skulle sette standpunktarakter på elever. I gruppeintervjuene bidro alle lærerne, både med og uten sensorerfaring, til å sammenligne disse to formene for sluttvurdering og til å belyse hvilke profesjonelle krav, fordringer og normer som påvirker lærerne på veien mot å kunne dokumentere elevenes kompetanse ved avslutningen av videregående opplæring.

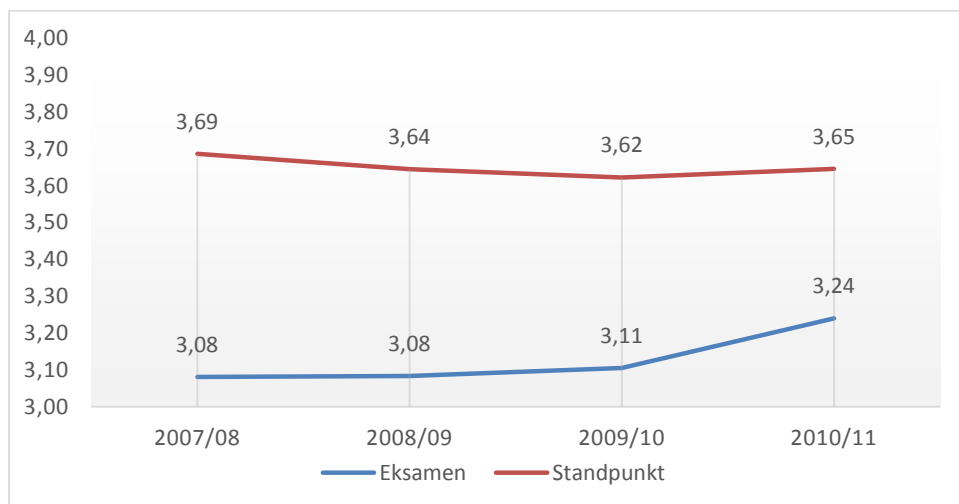
Datainnsamlingen hadde form av fokusgruppeintervju hvor deltakerne (mellom tre og fem lærere på hver skole) snakket mest med hverandre, og hvor forskeren opptrådte som en moderator med forholdsvis få spørsmål som kunne bidra til å styre samtalen i en bestemt retning. Til dette formålet brukte vi en semi-strukturert intervjuguide. I intervjuguiden hadde vi et sett med forhåndsdefinerte spørsmål som vi ønsket å få svar på, men vel så viktig var det Halkier (2012) beskriver som samspillet mellom informantene og den informasjonen dette kunne gi oss gjennom intervjusituasjonen.

Denne artikkelen bygger på uttalelser fra samtlige gruppeintervjuer med de til sammen 32 lærerne. At lærere med bakgrunn fra begge disse undervisningsfagene deltok i felles intervjuer, formet selve intervjuene og har i seg selv bidratt til å bygge opp vår forståelse av hvordan lærere arbeider med og resonnerer om vurdering både i henholdsvis norsk og matematikk og uavhengig av fag. Det var vårt primære inntrykk at lærerne generelt ønsker å snakke om vurdering, og at dette er særdeles viktig for dem i utøvelsen av faget og yrket. Selv om mange lærere er del av et fagfelleskap hvor man lager og retter prøver i fellesskap eller for hverandre, var det vårt inntrykk at lærere mangler anledninger til å diskutere vurdering som fenomen, krav og praksis med kolleger på et overordnet og ikke bare på et situasjonsbestemt nivå. Vi oppfattet at gruppeintervjuet representerte en slik situasjon for lærerne som deltok, og at denne opplevelsen hos lærerne preget diskusjonene slik at de kan fortelle oss om gruppens *sosialitet* (Halkier, 2012). Dette betegner de fortolkninger, samhandling og normer som enten er felles eller som divergerer innad i gruppen, og som avdekkes gjennom samtalen.

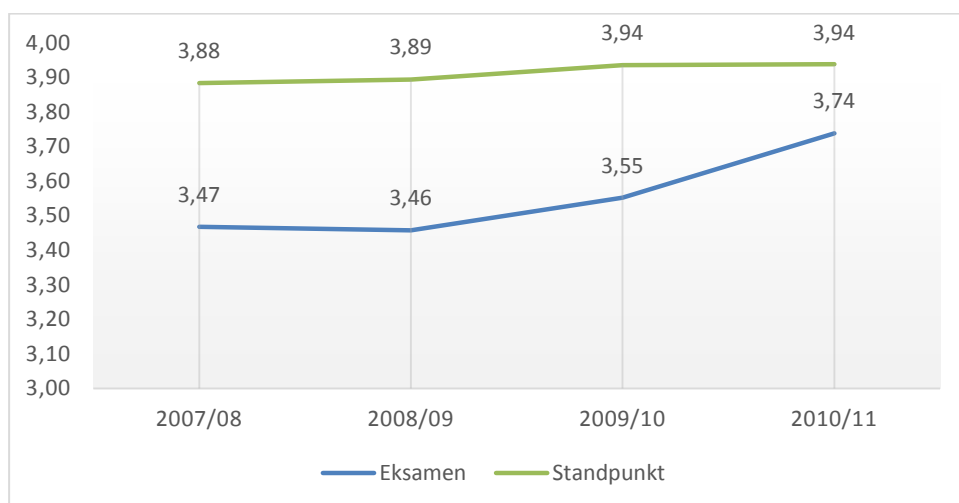
Analyse og resultater

Vi begynner med å se nærmere på hvordan de to typene karakterer ser ut i norsk hovedmål og i realfagsmatematikk, ved å sammenligne gjennomsnittlig eksamenskarakter og standpunktarakter for hele kullet med elever som tar studieforberevende utdanningsprogrammer og har faget i Vg3. Dette er

presentert i figur 1 og 2. Felles for de to figurene er at gjennomsnittlig standpunkt karakter er forholdsvis stabil over tid, mens eksamens karakteren derimot ikke er like stabil. Større variasjon i gjennomsnittlig eksamens karakter enn i gjennomsnittlig standpunkt karakter, både i norsk og i matematikk, indikerer også at dette målet er mindre stabilt.



Figur 1. Gjennomsnittlig eksamens karakter og standpunkt karakter i norsk hovedmål for elever i Vg3 som startet i videregående opplæring i perioden 2007/08 til 2010/11.



Figur 2. Gjennomsnittlig eksamens karakter og standpunkt karakter i realfagsmatematikk (R2) for elever i Vg3 som startet i videregående opplæring i perioden 2007/08 til 2010/11.

De to figurene viser altså at eksamens karakterene er mindre stabile enn det standpunkt karakterene er, målt over det samme tidsrommet. Lærernes refleksjoner over forskjellene mellom disse to sluttvurderingsformene kan bidra til å forklare hvorfor det oppstår avvik mellom eksamens- og standpunkt karakterer.

Når lærere setter standpunkt karakter

Vi spurte først lærerne om hvordan de går frem når de gir elevene standpunkt karakterer. Alle lærerne hadde erfaring med dette, og sa at de la Vurderingsforskriften³ til grunn i dette arbeidet. Vurderingsforskriften pålegger lærerne å la standpunkt karakteren bygge på slutt kompetansen i faget. Slik skal elevens progresjon og faglige utvikling vurderes ved avslutningen av opplæringen gi mer uttelling for standpunkt karakteren enn prestasjonene da eleven var på et presumptivt lavere faglig nivå tidligere i opplæringsløpet. Intervjuene viste samtidig at lærerne i norsk og matematikk tolket dette noe ulikt. En matematikklærer sa for eksempel:

Jeg synes ikke at en 2'er i matematikk i september skal bety noe som helst når jeg setter karakteren i mai.

En norsklærer sa i samme intervju:

Vi vil legge mest vekt på det som skjer i 2. termin mot slutten, men det er jo vurderingssituasjoner ellers i året og i 1. termin som man må se på.

Forskjellen mellom de to utsagnene skyldes at ved avslutningen av de studiespesialiserende utdanningsprogrammene skal eleven testes i kompetansemål for hele tre år med opplæring i norskfaget. Matematikkfaget i videregående skole har til sammenligning avsluttende årsheter med tilhørende kompetansemål for hvert skoleår. Disse utsagnene representerer likevel ytterpunkter i lærernes tolkning av «slutt kompetanse». I intervjumaterialet er det også lærere i matematikk som heller mot at elevens tidligere prestasjoner kan trekkes inn i vurdering til standpunkt karakter.

For å få dokumentert elevens slutt kompetanse i faget holder alle skolene i vårt utvalg heldagsprøver, årsprøver eller såkalte eksamensforberedende prøver i norsk og matematikk i andre halvdel av vårsemesteret i avgangsåret. Prøvene lages av elevenes faglærere, eventuelt i samarbeid mellom flere lærere med ansvar for samme fag ved skolen. Faglærerne setter dermed standard for hva elevene skal prøves i for å dokumentere slutt kompetansen.

Det å lage heldagsprøver og vurdere elevens resultater på bakgrunn av slike prøver er derfor noe lærerne tar svært alvorlig. Mange av dem velger å bruke tidligere gitte eksamensoppgaver tilgjengelige på Utdanningsdirektoratets nettsider til dette formålet. En matematikklærer sa om dette: «Jeg liker det best, for da får elevene et inntrykk av hva de skal testes i til slutt.» Bruk av tidligere gitte eksamensoppgaver er også relativt vanlig som ledd i undervisningen i vårsemesteret, dels som en forberedelse til eksamen men også som del av det løpende arbeidet med kompetansemålene. Med disse tidligere gitte eksamensoppgavene følger det ofte også forslag til løsninger, vurderingskriterier og sensorveiledninger. Disse veiledningene brukes av lærerne i varierende grad:

³ https://lovdata.no/dokument/SF/forskrift/2006-06-23-724/KAPITTEL_4#KAPITTEL_4

Jeg har lest dem. Men da mest i en situasjon hvor jeg skal vurdere til eksamen, jeg har ikke brukt dem mye i min egen vurdering.

(norsklærer)

Jeg ser på dem i forbindelse med karakterskala, (...) så det er mest for å se at jeg ligger i nærheten av det samme.

(matematikklærer)

Tidligere gitte eksamener får på denne måten status som en form for nasjonal standard for noen læreres standpunktvurdering, selv om de vektlegger dette i ulik grad. Gjennom intervjuene fikk vi vite at til tross for at lærerne på denne måten bruker en avsluttende heldagsprøve til å dokumentere elevens standpunkt-karakter, vil de samtidig trekke med seg kunnskap og erfaringer om elevens utvikling og tidligere prestasjoner gjennom skoleåret. En matematikklærer sa om dette:

Den siste prøven blir den viktigste, for den viser jo totalkompetansen, men så er det lov til å ha en dårlig dag. Du kan feile på den siste prøven og likevel få en god karakter hvis det viser seg at du har hatt et jevnt og høyt nivå det siste året.

Samtidig fortalte lærerne at de mange samtalsituasjonene de har med elevene gjennom skoleåret, kan være med på å styrke lærerens oppfatning av hva eleven kan og hvilken faglig utvikling som har funnet sted. Til dette regnet lærerne de såkalte elevsamtalene, vurderingssamtalene eller halvårssamtalene som de er pålagt å gjennomføre, men også løpende samtaler med elevene i og utenfor undervisningen.

Gjennom intervjuene ble vi også gjort kjent med at selv om heldagsprøven skal være dokumentasjon på elevenes sluttkompetanse, er ikke denne prøven nødvendigvis «endelig» på alle skoler. På en av skolene i utvalget vårt kunne elevene søke om å få ta en ny heldagsprøve hvis de ikke var fornøyde med resultatet. Ved alle skolene ville lærere også forsøke å sette opp nye vurderingssituasjoner for elever som hadde gjort det så dårlig på heldagsprøven at de ville få stryk på vitnemålet. Vi fikk inntrykk i intervjuene av at her var mange lærere villige til å strekke seg langt for å hjelpe eleven til å demonstrere det nødvendige minimum av kompetanse for å unngå stryk. En matematikklærer sa om dette:

Hvis resultatene har vært dårlige gjennom hele året og han stryker på heldagsprøven, så stryker han. Men har det vært litt variasjon, så kaller jeg ham inn til en samtale for å gi ham en 2er.

På flere av skolene kunne elever også få såkalte «vippeprøver», hvis de sto mellom to karakterer. Dette gjaldt også elever som lå i de øvre deler av karakter-skalaen. Intervjumaterialet viser at dette er praksiser som varierer blant lærere både mellom skoler og innenfor samme skole – enkelte ganger også mellom lærere i samme fag på én skole.

Denne samlede innsatsen, i tillegg til de formaliserte og lovpålagte elevsamtalene som skal finne sted mellom lærer og elev to ganger i løpet av hvert skoleår, er til for at eleven skal få demonstrere og få dokumentere sin kompetanse mot slutten av opplæringen. Slik vi har sett av det foregående, betyr dette at det også underveis i semesteret legges stor vekt på vurdering, fordi dette òg kan få betydning for dokumentasjonen av elevens sluttkompetanse.

Når lærere er sensor på eksamen

De fleste lærerne i utvalget vårt hadde erfaring som sensor, selv om erfaringene var spredt på sentralt gitt skriftlig eksamen, lokalt gitt skriftlig eksamen eller muntlig eksamen. Dette dreide seg om mange ulike fag som lærerne hadde undervist i, men mer enn halvparten av dem hadde sensorerfaring fra sentralt gitt skriftlig eksamen i enten norsk eller matematikk. Felles for disse erfaringene var den verdien lærerne tillat sensormøter eller formell sensorskolering som lærerne hadde hatt sammen med kolleger før eksamen. Lærerne fortalte i intervju at de satte stor pris på og hadde stort utbytte av å delta i slike møter og skoleringer, fordi de der fikk justere sine egne vurderinger av elevens prestasjoner etter faste vurderingskriterier (eksamensveiledninger). En norsklærer uttrykte dette slik:

Den erfaringen de har, de som har vært sensor til sentralt gitt eksamen, de tar jo med seg den kompetansen inn i kollegafellesskapet. Så dette er viktig, ikke bare i egen vurderingspraksis, men også fordi kunnskap og kompetanse kan tilflyte andre. Mange av oss har lang erfaring, men noen er ferske og skal sette standpunkt karakter for første gang. Det er viktig for oss å følge dem opp.

Sensormøtene og vurderingskriteriene som sensorene bruker, representerer for lærerne en rettesnor i vurderingsarbeidet, som flere av dem forteller at de savner når de selv skal sette karakter på elevenes arbeider gjennom skoleåret og ved standpunkt vurdering. En norsklærer beskrev på denne bakgrunnen sensormøtene som «den beste dagen i året» da han fikk møte kolleger og diskutere faglige problemstillinger med mål om ensartet vurderingspraksis. Det kan være verdt å nevne her at alle lærerne svarte nei på vårt direkte spørsmål om de hadde fått formell opplæring i vurdering og karaktersetning som del av utdanningen eller på den enkelte skole. Opplæringen de hadde fått, var uformell og overført som kunnskap gjennom samarbeid i fagfellesskapet.

Lærerne som hadde erfaring som sensor, brukte uttrykket «sensorbriller» når de fortalte hvordan de vurderte en anonym eksamensbesvarelse som de hadde foran seg. Med dette mente de at de tok utgangspunkt i vurderingskriteriene for prøven og gjorde vurderingen utelukkende basert på disse. Fremgangsmåten står i kontrast til hvordan lærerne forteller at de setter standpunkt karakter, hvor de beskriver hvordan de gjerne benytter seg av (lokalt utviklede) vurderingskriterier, men hvor de samtidig tar hensyn til egen kjennskap til eleven og hva eleven har vist at hun mestrer gjennom året. Sensor har ingen slike opplysninger om eksamenskandidaten, og vil dermed sette en objektiv karakter på elevens

besvarelse. Resultatet er ofte en strengere vurdering av eksamensbesvarelsen enn hva eleven kunne tenkes å ha fått hvis det var faglærer som rettet en tilsvarende eksamensforberedende prøve som skulle gi grunnlag for standpunkt-karakter. «Jeg er strengere som sensor, tror jeg, enn jeg er som faglærer,» sa en norsklærer om dette. En matematikklærer betegnet eksamen som et «sjansespill» for eleven, fordi lærerens standpunkt-vurdering ville bygge på et mer helhetlig inntrykk. En annen matematikklærer uttalte ved samme anledning:

Eksamen er jo en spesiell måte å teste elever på. En elev kan jo være god i matte selv om han ikke er flink til å sitte og jobbe i timer i strekk med matematikkoppgaver.

Som vi har sett i det foregående, vil faglærer ofte bestrebe seg på å sikre at eleven ikke får strykkarakter ved standpunkt-vurdering. Sensor har på den annen side ingen tilsvarende bindinger til eleven og dennes fremtidsutsikter i videre utdanning eller jobb, og vil derfor ifølge lærerne sette strykkarakter hvis vurderingskriteriene tilsier dette. Hvor faglærer kvier seg for å bruke bunnen av karakterskalaen, bruker altså sensor også karakteren 1. På den annen side vil sensor ofte, med grunnlag i vurderingskriteriene, være tilbakeholden med å gi toppkarakter til eksamen. En norsklærer sa om dette:

Når jeg selv sensurerer, ser jeg at det er litt lenger mellom 6'erne enn hva det kanskje er i [mine egne] enkeltvurderinger i løpet av året. I en bunke på 200 [eksamens] besvarelser er det ikke mange 6'ere.

Den samme læreren hadde hele tiden arbeidet parallelt som faglærer med ansvar for standpunkt-vurdering. Hun sa om dette, her om elever som ikke får toppkarakter, men som likevel får uttelling på standpunkt-karakteren som «belønning» for hardt arbeid:

Helt ærlig, du ønsker virkelig at eleven din skal få den 4'eren, du vil gjerne se den utviklingen. Og så tenker man gjerne at man gir den høyeste av de to karakterene mot slutten.

Disse to utsagnene fra samme lærer illustrerer godt hvordan lærerne arbeider ulikt med eksamensvurdering og standpunkt-vurdering, og hvilke ulike hensyn som spiller inn i de to situasjonene.

Lærernes syn på avvik mellom standpunkt- og eksamens-karakter

Så langt i denne artikkelen har vi vist hvordan lærerne bruker to ulike tilnæringsmåter for å sette henholdsvis standpunkt- og eksamens-karakter. Bare ved én av de åtte skolene vi besøkte i forbindelse med disse intervjuene, var det lærere som *avviste* dette som grunnlag for at de to karakterene bør være forskjellige, de mente at grunnlaget for eksamen og standpunkt skulle være likt. Disse lærerne, som begge underviste i matematikk, anså eksamenssensuren som den viktigste dagen i året fordi den ga dem som lærere en justering av egen vurderingspraksis. Disse lærerne sa seg fornøyde hvis sensuren lå tettest mulig

opp til deres egen standpunktvurdering av elevene. Ved alle de andre skolene aksepterte lærerne et avvik mellom standpunkt- og eksamenskarakter.

Hvor stort avvik som ble tolerert, varierte imidlertid mellom skolene. På én skole aksepterte kollegiet et avvik på en hel karakter, mens på en annen skole hadde lærerne en norm om at avviket ikke burde være mer enn 0,5 i karaktergjennomsnitt for avgangskullet. Lærernes aksept for slike avvik var konsistente med kriteriene som vi la til grunn for vårt utvalg. Slik hadde lærere ved skoler som hadde lite avvik mellom de to vurderingsformene over tid, også de strengeste holdningene i dette spørsmålet. Alle lærerne vi snakket med, aksepterte imidlertid avvik på individnivå. Lærere ved skoler som hadde større avvik mellom de to vurderingsformene over tid, sa imidlertid også at de ville være mer vaksomme overfor større avvik som gjaldt hele klasser eller årstrinn eller som vedvarte over tid. Dette forteller oss at lærerne på alle de åtte skolene tenkte på hva slike avvik kunne bety for deres egen praksis. «Man må jo måle seg opp mot noe,» sa en matematikklærer, og fortsatte:

For meg er det veldig nyttig å se [sensuren]. Jeg blir veldig glad hvis elevene mine gjør det omtrent like bra på eksamen som det jeg har gitt dem [i standpunkt]. Jeg tenker, ok, da er jeg i hvert fall sånn høvelig på rett spor.

Dette illustrerer for oss at selv om lærere fastholder at det her er snakk om to ulike vurderingsformer, forholder de seg likevel til sensuren som en tilbakemelding om kvaliteten på deres egen vurderingspraksis. Andre lærere inntok samtidig en pragmatisk holdning til sensuren, eller de ga uttrykk for at de ønsket en slik pragmatisme. En norsklærer sa:

Standpunkt måler jo mye bredere, det måler jo ideelt sett alle kompetansemål, men eksamen måler jo bare den ene dagen og de oppgavene som ble gitt den dagen. Samtidig bør ikke spriket være for stort, i alle fall ikke over tid.

Ved fem av de åtte skolene ble avvik mellom eksamen og standpunktkarakter brukt som ett av flere mål på kvalitet, og ved to av skolene hadde skoleledelsen et mål om lavest mulig avvik mellom disse vurderingsformene. Ved særlig én av disse skolene var norsklærerne svært misfornøyde med denne målsettingen, og en av dem sa i intervju: «Dersom standpunktkarakter skal sammenfalle med eksamen, kan standpunkt like gjerne sløyfes.» Denne norsklæreren og kollegene hennes fortalte at de selv hadde forsøkt å arbeide med standpunktvurdering som om heldagsprøven var en eksamensbesvarelse levert av eleven. I denne situasjonen så de for seg at elevens heldagsprøve skulle vurderes som om den var levert av en for læreren ukjent elev, som om det var en eksamensbesvarelse. Det de oppdaget da de på denne måten tok på seg «sensorbrillene», var at vurderingen de gjorde ikke var i tråd med det de oppfattet som sitt eget faglærerideal. Dette idealet beskriver de samme lærerne som å la kjennskap til elevens styrker og svakheter samt faglige progresjon komme eleven til gode i vurderingen, ikke minst ved at læreren bestreber seg på å forstå hva eleven vil

uttrykke. En norsklærer på en annen skole uttrykte det samme som å se etter «den røde tråden» i elevens besvarelse, selv om denne tråden kanskje ikke var synlig med første blikk.

Lærerne vi intervjuet nevnte også forskjellen i klagebehandling for eksamens- og standpunkt karakteren. Som sensor retter læreren en bunke eksamensbesvarelser, og ved klage oppnevnes en ny kommisjon. Hver enkelt sensor er derfor ferdig med vurderingen når karakter er satt. Klage på standpunkt karakter bringer derimot faglærer inn i vurderingen på nytt, og faglærer må derfor sørge for å ha godt dokumentasjonsgrunnlag.

Diskusjon

Vi innledet analysene med å vise forskjellen mellom standpunkt- og eksamens karakter, og at det over tid er større stabilitet i standpunkt karakterene enn i eksamens karakterene.

Vi har sett at lærere ser på det å sette standpunkt og det å sette eksamens karakter som to ulike vurderingsformer. Mens eksamen av lærerne regnes for å være en «stikkprøve» på kunnskap som eleven kan gjenkalle og nedtegne i løpet av noen timer, er standpunkt karakteren et resultat av lærerens kjennskap til eleven og elevens progresjon over tid. Dette gjelder for lærerne selv om de sier at også ved standpunkt vurdering er det elevens sluttkompetanse som skal telle, og de mener av samme årsak at dette gir en riktigere vurdering av slik sluttkompetanse enn hva eksamens karakteren representerer. Faglærer har imidlertid insentiver til å behandle elever differensielt i standpunkt vurderingen. Dette skjer blant annet på grunn av klageordningens innretning, hvor faglærer selv må svare for vurderingen. Sensor er ikke underlagt en slik klageordning, og vil ifølge lærerutsagnene i våre intervjuer tendere mot å behandle alle elever mer universelt, prinsipielt og strengere, målt mot de definerte kriteriene og nasjonale standardene gitt i sensuren for eksamen. Elevens faglærer vil imidlertid, og med støtte i regelverket, mene at det tilligger faglærers oppgave og skjønn å vurdere standpunkt karakteren som en gjenspeiling av elevens progresjon gjennom opplæringen.

Slik mener alle norsklærerne vi har intervjuet, at det er naturlig at eksamens- og standpunkt karakter avviker. Matematikklærerne er mindre kategoriske i så henseende. I analysen har vi sett hvordan matematikklærerne er mer tilbøyelige enn norsklærerne til å se på eksamenssensuren for å få en føling med hvordan deres vurderingspraksis ligger an, sammenlignet med fagfellers vurdering av elevens kompetanse på eksamen. Vi har også vist eksempler i denne artikkelen på at lærere i begge fag bruker eksamensveiledninger og elevens besvarelser på tidligere gitte eksamensoppgaver for å sette standpunkt karakter, og at faglærere verdsetter egen eller kollegers sensorerfaringer for å få justert sin egen vurderingspraksis. Alt dette er eksempler på at eksamen har en innvirkning på

læreres standpunktvurdering og undervisning og forberedelse av elevene til eksamen, samtidig som det ikke rokker ved lærernes oppfatning om at standpunkt og eksamen er to forskjellige ting.

Våre funn, i form av både de stabile ulikhetene mellom standpunkt karakter og eksamens karakter og læreres vurderingspraksiser forbundet med disse to vurderingsformene, reiser spørsmålet om eksamen faktisk kan fungere som rettesnor for læreres standpunkt karaktersetning og dermed ha en form for kalibrerende funksjon. Stabiliteten i standpunkt karakterene kan fortolkes på minst to måter. Dels kan den ses som et uttrykk for at det finnes en stabilitet knyttet til det faglig profesjonelle skjønnnet lærere utviser, og som står i kontrast til at eksamensoppgaver av naturlige årsaker må være ulike fra år til år. Dels kan den også ses som et uttrykk for at nettopp eksamen og sensorskolering har en samordnende funksjon som bidrar til å underbygge og sikre et slikt faglig profesjonelt fellesskap mellom lærere, men uten at det betyr at vi skal anse eksamen som standarden. Videre reiser våre funn også spørsmålet om eksamen bør ha rollen som rettesnor og om det er behov for mer rom og felles arenaer for lærere til å diskutere standpunkt karaktersetning nasjonalt. Variasjonen over tid i eksamens karakter, for eksempel grunnet endringer i vanskegrad eller omlegging av eksamensopplegg gjør det også rimelig å spørre om det er hensiktsmessig å bruke avvik som en indikator på kvalitet i læreres karaktersetning.

Hvis det er slik at standpunkt- og eksamens karakter bør være like sett fra skoleeier eller de nasjonale myndighetenes side, vil vi peke på at vi i denne artikkelen har vist at elevenes innflytelse over de to vurderingssituasjonene er svært forskjellig. I analysen av hvordan lærerne setter standpunkt karakterer, har vi gitt flere eksempler på hvordan denne vurderingen foregår i en slags dialog med eleven, selv om denne dialogen i noen grad kan skje uten aktiv medvirkning fra eleven. Et eksempel på dette er når faglærer setter opp nye vurderingssituasjoner for elever som står i fare for å stryke, men også når faglærer går «i dialog» med elevens tidligere innleveringer og prøveresultater for å finne mer solid belegg for en standpunkt karakter. Mer aktiv dialog inntreffer når elever ber om eller lærer foreslår såkalt vippeprøve, og når lærer og elev gjennomfører såkalte elevsamtaler eller vurderingssamtaler om karakterer i løpet av skoleåret.

Vi har i denne artikkelen vist at ingen slik dialog skjer mellom den anonyme eleven og sensor, som retter eksamen uten kjennskap til elevens øvrige prestasjoner, og dennes eventuelle årsaker til fravær eller sykdom gjennom skoleåret. Sammenlignet med standpunkt vurdering har eleven ingen stemme i vurderingen av eksamens karakteren, med unntak av en formell klageadgang (hvor det oppnevnes ny kommisjon). Et paradoks ved forventningen om samsvar mellom eksamens- og standpunkt karakter er at ut fra vår analyse kan dette kun oppnås dersom faglærer med Brookharts (2013) ord opptrer som «dommer» i enhver vurderingssituasjon. Dette vil også måtte innebære å ta bort det vi har kalt for dialogen (den tause så vel som den uttalte) om læring og progresjon mellom faglærer og elev i vurderingen. Dette vil tilsvare det å sette standpunkt karakter

med «sensorbriller», slik to av lærerne i vårt materiale sier at de har prøvd. Lærerne opplevde imidlertid dette som å fjerne seg fra det profesjonelle faglige skjønnet som de mente måtte ligge til grunn for en valid og rettferdig standpunkt-vurdering.

Gitt de forskjellene som denne artikkelen har avdekket mellom de to vurderingsformene, kan det være grunn til å dempe forventningene om at et sammenfall i karakterer betyr høy kvalitet i læreres praksis for standpunkt-vurdering. Det finnes også grunn til å tro at lærere, som utmerket godt kjenner forskjellen mellom disse to vurderingsformene, vil vite og kunne tilpasse seg eksamensvurderingsformen i standpunkt-vurderingen, men uten at dette reflekteres i deres øvrige profesjonelle praksis som lærer, for eksempel ved endringer i undervisningen. Tilpasningene mellom de to karakterene vil da være rent administrative sett fra lærerens ståsted, og verdien av å ha to ulike vurderingsformer vil falle bort.

Konklusjon

Spørsmålet om hva som er god eller riktig praksis ved vurdering i form av standpunkt- og eksamenskarakter, synes tett forbundet med hvilken status og funksjon disse vurderingsformene er gitt i det norske opplæringssystemet. Våre funn i denne studien peker på at det finnes flere svakheter ved ideen om at eksamen kan fungere som kvalitetssikring av standpunktkarakterene. Dels handler dette om at det dreier seg om to svært ulike vurderingsformer som rent praktisk skiller seg tydelig fra hverandre, men som også har klare ulikheter slik de er juridisk definert. Videre omfattes de to vurderingsordningene av helt forskjellige premisser i selve karaktersettingen. Den ene skjer på grunnlag av en lærer-elev-relasjon preget av en form for dialog og lærerens inngående kjennskap til eleven som person, mens den andre gjennomføres med gjensidig anonymitet og et såkalt sensorblikk. De to vurderingsordningene har styrker hver for seg som det er flere grunner til å forsvare, men også svakheter som kan kritiseres. Samtidig er et stadig tilbakevendende spørsmål om det norske utdanningssystemet trenger begge ordninger for å sikre legitimitet i karaktersettingen. Vår studie tyder på at begge ordninger bidrar til å ivareta viktige vurderingsteoretiske spørsmål hver for seg, og at det er gode begrunnelser for begge ordninger, men at ideen om at den ene kan kvalitetssikre den andre er langt mer problematisk. Det er således et spørsmål om det isteden bør arbeides for å finne alternative måter å kvalitetssikre standpunkt-vurderingen og eksamen på – hver for seg og basert på de to vurderingsordningenes ulike premisser.

Takk

Datainnsamlingen til dette prosjektet er finansiert av Utdanningsdirektoratet gjennom prosjektet *Karakterpraksis i offentlige og private videregående skoler*.

Om forfatterne

Elisabeth Hovdhaugen er forsker ved NIFU. Hennes forskningsinteresser omfatter blant annet rekruttering til og frafall i høyere utdanning, studentgjennomstrømning og studenters tidsbruk og karakterer.

Institusjonstilknytning: NIFU Nordisk institutt for studier av innovasjon, forskning og utdanning, Postboks 2815 Tøyen, 0608 Oslo.

E-post: elisabeth.hovdhaugen@nifu.no

Tine Prøitz er professor ved Universitetet i Sørøst-Norge. Hennes forskningsinteresser omfatter blant annet sluttvurdering, læringsutbytte, utdanningspolitikk og -reform.

Institusjonstilknytning: Institutt for pedagogikk, Universitetet i Sørøst-Norge, Postboks 4, 3199 Borre.

E-post: tine.proitz@usn.no

Idunn Seland er forsker ved NOVA. Hennes forskningsinteresser omfatter blant annet skolepolitikk og utdanning, særlig demokrati- og religionsundervisningen i skolen og spørsmål knyttet til private skoler.

Institusjonstilknytning: Velferdsforskningsinstituttet NOVA, OsloMet – storbyuniversitetet, Postboks 4 St. Olavs plass, 0130 Oslo.

E-post: idunn.seland@oslomet.no

Referanser

Angelo, T. (1998). *Classroom Assessment and Research: An Update on Uses, Approaches, and Research Findings*. San Francisco: Jossey-Bass.

Black, P, Harrison, C., Hodgen, J., Marshall, B. & Serret, S. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215–232.

Bloxham, S. (2009). Marking and moderation in the UK: False assumptions and wasted resources. *Assessment and Evaluation in Higher Education*, 34(2), 209–220.

Brookhart, S. M. (1991). Letter: Grading Practices and Validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.

Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69–90.

Cheng, L. & Sun, Y. (2015). Teachers' grading decision making: Multiple influencing factors and methods. *Language Assessment Quarterly*, 12(2), 213–233.

- Eggen, A. E. (2004). *Alfa and Omega in Student Assessment: Exploring Identities of Secondary School Science Teachers*. Doktoravhandling, Universitetet i Oslo.
- Galloway, T. A., Kirkebøen, L. J. & Rønning, M. (2011). *Karakterpraksis i grunnskoler. Sammenheng mellom standpunkt- og eksamenskarakter*. Rapport 4/2011. Oslo/Kongsvinger: Statistisk sentralbyrå.
- Halkier, B. (2012). Fokusgrupper. I S. Brinkmann & L. Tanggaard (red.), *Kvalitative metoder. Empiri og teoriutvikling* (s. 133–152). Oslo: Gyldendal Akademisk.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning–tensions and synergies. *Curriculum Journal*, 16(2), 207–223.
- Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T. S. & Vibe, N. (2014). *Karakterer i offentlige og private videregående skoler: En analyse av eksamens- og standpunkt-karakter i norsk og matematikk og rutiner for standpunktvurdering i offentlige og private videregående skoler*. Rapport 24/2014. Oslo: NIFU.
- Hægeland, T., Kirkebøen L. J., Raaum, O. & Salvanes, K. G. (2005). Familiebakgrunn, skoleressurser og avgangskarakterer i norsk grunnskole. I *Utdanning 2005 – deltakelse og kompetanse*. Statistiske analyser, 74, s. 34–52. Oslo/Kongsvinger: Statistisk sentralbyrå.
- Knight, P. T. (2002). Summative assessment in higher education: practices in disarray. *Studies in Higher Education*, 27(3), 275–286.
- Manke, M. P. & Loyd, B. H. (1990). *An investigation of non-achievement related factors influencing teachers' grading practices*. Boston: National Council on Measurement in Education.
- Mausethagen, S., Prøitz, T. S. & Skedsmo, G. (2018). *Elevresultater mellom kontroll og utvikling*. Oslo: Fagbokforlaget.
- McMillan, J. H., Myran, S. & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, 95(4), 203–213.
- Meld. St. 28 (2015–2016). *Fag – Fordypning – Forståelse. En fornyelse av Kunnskapsløftet*. Oslo: Kunnskapsdepartementet.
- Opheim, V., Grøgaard, J. B. & Næss, T. (2010). *De gamle er eldst? Betydning av skoleressurser, undervisningsformer og læringsmiljø for elevenes prestasjoner på 5., 8. og 10. trinn i grunnopplæringen*. NIFU STEP-rapport 34/2010. Oslo: NIFU STEP.
- Prøitz, T. S. & Borgen, J. (2010). *Rettferdig standpunktvurdering. Det (u)muliges kunst?* NIFU STEP-rapport 16/2010. Oslo: NIFU STEP.
- Prøitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, 4(3), 555–572.
- Resh, N. (2009). Justice in Grades Allocation: Teachers' Perspective. *Social Psychology of Education*, 12(3), 315–325.
- Stiggins, R. J., Frisbie, D. A. & Griswold, P. A. (1989). Inside High School Grading Practices: Building a Research Agenda. *Educational Measurement: Issues and Practice*, 8(2), 5–14.
- Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British journal of educational studies*, 53(4), 466–478.
- Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, 8, 5–30.
- Thronsen, I., Hopfenbeck, T. N., Lie, S. & Dale, E. L. (2009). *Bedre vurdering for læring. Rapport fra «Evaluering av modeller for kjennetegn på måloppnåelse i fag»*. Universitetet i Oslo.
- Tveit, S. (2014). Educational assessment in Norway. *Assessment in Education: Principles, Policy and Practice*, 21(2), 221–237.
- Vibe, N. & Hovdhaugen, E. (2014). Vedlegg 3 Karakterforskjeller – delrapport levert til Udir i desember 2013. I E. Hovdhaugen et al. (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.* Rapport 24/2014. Oslo: NIFU.
- Wiborg, Ø., Arnesen, C. Å., Grøgaard, J. B., Støren, L. A. & Opheim, V. (2011). *Elevers prestasjonsutvikling – hvor mye betyr skolen og familien? Andre delrapport fra prosjektet 'Ressurser og resultater'.* Rapport 35/2011. Oslo: NIFU.
- Wiliam, D. & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), 537–548.
- Wyatt-Smith, C. & Klenowski, V. (2013). Explicit, latent and meta-criteria: Types of criteria at play in professional judgement practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 35–52.
- Yorke, M. (2011). Summative assessment: dealing with the 'measurement fallacy'. *Studies in Higher Education*, 36(3), 251–273.
- Aasen, P., Møller, J., Rye, E., Prøitz, T. S. & Hertzberg, F. (2012). *Kunnskapsløftet som styringsreform – et løft eller et løfte? Forvaltningsnivåenes og institusjonenes rolle i implementeringen av reformen.* Oslo: NIFU.