

Kunnskapsgrunnlag om standpunktvurdering

Gustaf B. Skar¹ & Therese N. Hopfenbeck^{1,2}

¹Skrivesenteret, ILU, NTNU

²Department of Education, University of Oxford

Forord

Denne rapporten er utarbeidet med utgangspunkt i et oppdrag gitt til NTNU fra Utdanningsdirektoratet, mai 2021. Vi vil takke Kristin Andersland i Utdanningsdirektoratet for godt samarbeid angående avklaring av oppdragets formål og tilbakemeldinger på første utkast. Målet for oppdraget var å sammenfatte forskning på standpunktkarakterer i Norge og Norden, og deretter belyse hvilke tiltak som har vært gjennomført i land vi sammenligner oss med, for å forbedre praksis rundt standpunkt.

Rapporten gir en metodisk oversikt over litteratursøk, eksklusjon- og inklusjonskriterier for kunnskapsoppsummeringen relatert til standpunktvurdering og koding av litteratur og analyser.

Basert på funn fra Norge og Norden, har vi gitt anbefalinger for fremtidig arbeid med standpunkt i Norge. Vi vil understreke at det er viktig å se våre anbefalinger angående fremtidig arbeid med standpunktvurdering i sammenheng med den historiske og lokale konteksten norsk læreplan er en del av.

Trondheim, Oxford, september 2021,

Forfatterne

Sammendrag

Denne rapporten svarer på tre forskningsspørsmål: 1. Hvilken forskningsbasert kunnskap finnes om læreres og skolelederes praksis med standpunktvurdering i Norden? 2. Hvilken forskningsbasert kunnskap finnes om kvaliteten på læreres standpunktvurdering og skolelederes bruk av standpunktvurdering i Norden? 3. Hvilke eksempler på tiltak for å øke kvaliteten til standpunktvurdering finner vi i et utvalg land det er naturlig å sammenligne seg med?

Rapporten gir en metodisk oversikt over hvilke eksklusjons- og inklusjonskriterier for litteratursøk som kunnskapsoppsummeringen bygger på, samt koding og analyser av litteraturen. Sammenlagt bygger denne kunnskapsoversikten på 42 undersøkelser og i alt har 933 lærere, 1818 elever og 35 rektorer deltatt i disse undersøkelsene. Videre har henholdsvis 305 dokumenter og 3 357 8000 karakterer/prøveresultater blitt analysert.

I de undersøkelsene som inngår i oversikten, fremkommer det at lærere synes å ta standpunktvurdering på stort alvor og at standpunkt, ifølge selvrappoteringsen, bærer preg av systematikk. Videre har undersøkelser vist at lærergitte karakterer, slik som standpunkt, bedre predikerer senere skolefremgang enn resultater fra enkeltprøver. Samtidig viser denne kunnskapsoversikten at vi fremdeles har lite kunnskap om hva de ulike aktørene gjør i praksis. De fleste studiene har heller undersøkt meninger, holdninger og opplevelser til elever og lærere. En viktig konklusjon er dermed at vi har lite forskningsbasert kunnskap om læreres og skolelederes praksis og om kvaliteten på læreres standpunktvurdering.

Introduksjon: Kvalitet i vurdering av læring

De siste tjue årene har det vært en rekke statlige satsinger på formative vurderingspraksiser i Norge, blant annet gjennom pilotprosjektet *Evaluering av modeller for kjennetegn på måloppnåelse i fag*, (Thronsdén mfl., 2009) og *Vurdering for Læring* (Gamlem & Smith, 2013; Sandvik & Buland, 2014; Udir, 2019). Til sammenligning har det vært noe mindre fokus på summative vurderingspraksiser, *Vurdering av Læring*. Mens *Vurdering for Læring* (VfL) i en norsk kontekst har blitt sett i sammenheng med underveisvurdering, elevsamtaler, klasseromsdialoger, tilbakemeldinger fra elev til elev, eller fra lærere, og hvor tilbakemeldinger fra elever har ført til endringer i undervisning, er vurdering av læring en sluttvurdering, slik som nasjonale prøver, muntlig og skriftlig eksamen og *standpunktkarakterer* (Hopfenbeck, 2016). Det er viktig å erkjenne at for elever og lærere er det ikke alltid klare skiller mellom de to vurderingspraksisene, noe også forskere internasjonalt er enige om (Bennett, 2011). Fra elevperspektivet vil en standpunktkarakter derimot oppleves som endelig, en summativ vurdering, da den representerer en sluttvurdering som vanligvis ikke kan endres. I mange sammenhenger vil den også kunne ha avgjørende betydning for fremtidig skolevalg, da elever er avhengig av et høyt grunnskolepoeng for å komme videre på enkelte studier. Slik sett vil kvaliteten på læreres evne til å sette pålitelig standpunktkarakter oppleves som svært viktig for den enkelte elev.

Forskning på læreres standpunktkarakterer har gjennomgående vist at de har en tendens til å sette høyere karakterer ved standpunkt, enn det elever får ved eksamen. I enkelte land som avlyste eksamen under den globale pandemien i 2020 og benyttet seg av lærergitte karakterer, ble det framsatt kritikk av manglende kvalitet i vurderingsarbeidet og det som kalles «karakterinflasjon» (Adams, 2021). Samtidig er det viktig å erkjenne at eksamen representerer noe annet enn standpunkt, og dermed vil noen ulikheter ikke nødvendigvis bety at lærere har satt feil standpunktkarakter. Elever vil likevel kunne oppleve bekymring dersom det er mistanke om systematiske feil ved standpunktkarakterer. I Norge har særlig elevorganisasjonen vært opptatt av dette, og i perioden 2001–2007 jobbet de spesielt for å forbedre kvaliteten på standpunktkarakterer (Tveit, 2007). Elevorganisasjonen rettet spesielt oppmerksomheten mot det de oppfattet som ulike vurderingskulturer mellom skoler

og det de opplevde som urettferdig karaktersetting for elever. I en undersøkelse fra Eksamen- sekretariatet i 1999, ble 42 representativt utvalgte skoler med allmennfaglig studie- retning og totalt 6000 karakterer analysert. Det viste seg at elevene i snitt gjorde det bedre på muntlig eksamen og dårligere på skriftlig eksamen enn ved standpunkt. For elevene var det mest dramatiske de påviste forskjellene mellom skolene, et gjennomsnittlig fall i karakteren fra standpunkt til eksamen som varierte fra 0.3–1.0 karakterenhet (Kristensen, 1999). Siden dette ble avdekket, har det blitt påvist tegn til bedring. Blant annet har man dokumentert en nedgang i antall klager på standpunktkarakterer i perioden 2010–2015 (Utdanningsdirektoratet, 2015) noe Utdanningsdirektoratet hevder skyldes den økte sat- singen på vurderingskompetanse hos lærere. Vi har derimot få forskningsstudier som for- teller oss noe om standpunkt og vurdering av læring i Norden, og formålet med denne kunnskapsoversikten er å bidra til dette feltet ved å sammenfatte forskningsstudier som har blitt publisert etter debattene som blant annet Elevorganisasjonen fremmet.

Hva sier forskriften?

Forskrift til opplæringslova (2006) utgjør den offisielle og juridiske beskrivelsen av hvordan standpunktvurdering skal foregå. (Til forskjell fra eksamen, finnes ikke et rammeverk for standpunktvurdering.) I kapittel 3 i Forskrift til opplæringslova¹ står det følgende om standpunktvurdering²:

Ein standpunktkarakter skal vere uttrykk for den samla kompetansen eleven har i faget ved av- slutninga av opplæringa.

Eleven skal vere kjend med kva det blir lagt vekt på i fastsetjinga av hennar eller hans stand- punktkarakter. Eleven skal ha fått høve til å vise kompetansen sin på fleire og varierte måtar. Kompetanse som eleven har vist i løpet av opplæringa, er ein del av vurderinga når stand- punktkarakteren skal fastsetjast.

Standpunktkarakterar i fag med sentralt gitt eksamen skal fastsetjast seinast dagen før felles- sensurmøtet.

Standpunktkarakterar i fag med lokalt gitt eksamen skal fastsetjast seinast dagen før skolen gjennomfører den første eksamenen i faget på det aktuelle trinnet innanfor utdanningspro- grammet.

¹ Hele kapittel 3 om individuell vurdering i grunnskolen og i videregående opplæring hvor standpunktvur- dering er omtalt ble endra ved forskrift 1 juli 2009 nr. 964 (i kraft 1 aug 2009) og endra ved forskrift 29 juni 2020 nr. 1474 (i kraft 1 aug 2020).

² Den samme teksten finnes også i Forskrift til friskolelova (2006).

Faglæreren set standpunktkarakter. Dersom rektor er i tvil om reglane for fastsetjing av standpunktkarakter er følgde, kan rektor krevje at faglærer gjere ei ny fagleg vurdering før karakterane blir fastsette og førte.

Rektor har ansvaret for at det blir gjort enkeltvedtak om ikkje å gi standpunktkarakter i fag. For å gjere enkeltvedtak om å ikkje gi karakter, skal eleven først vere varsla, jf. § 3-8.

De to første avsnittene er av særlig interesse for denne kunnskapsoversikten, fordi de samtidig peker på praksis i forbindelse med standpunkt og kvaliteten til standpunktvurdering. I første avsnitt fremkommer at en standpunktkarakter skal være mulig å tolke som en indikasjon på elevens samlede faglige kompetanse.³ Dette kan sies å være et kvalitetskriterium, knyttet til validitet; brukere av karakterer skal kunne stole på at det tallet som utgjør karakteren faktisk *betyr* noe i relasjon til den faglige kompetanse eleven har vist. I andre avsnitt peker forskriften på at underlaget for standpunktvurdering skal være bredt og variert og at standpunktkarakteren skal speile kompetansen til eleven slik den har kommet til uttrykk under *hele* opplæringen (og ikke i en bestemt fase, som f.eks. slutten). Dette tolker vi til å bety at en god vurderingspraksis innebærer at elever får flere anledninger og måter å vise kompetansen sin på.

Det finnes naturlig nok en mengde aspekter ved (setting av) standpunkt som forskriften ikke sier noe om. Et åpenbart aspekt, som kan relateres til det forskriften sier om demonstrert kompetanse i hele opplæringen, er hvordan motstridende bevis på elevs kompetanse skal tolkes; skal elever som viser lavere faglig kompetanse mot slutten av opplæringen enn midt i opplæringen få karakter som gjenspeiler det «lavere» eller «høyere» nivået? Forskriften sier heller ikke noe om hvordan lærere kan undersøke eller forsikre seg om at underlaget for standpunkt er godt nok. En overveldende forskningsfunn peker for eksempel på at elevs faktiske kompetanse kommer til uttrykk først ved gjentatte vurderinger av samme kompetanse (Kane et al., 1999). Forskriften sier heller ikke noe om hvordan oppnåelse av ulike deler av den faglige kompetanse skal vektes i forhold til hverandre.

³ Dette uttrykkes også eksplisitt i Forskrift til opplæringslova og Forskrift til friskolelova, Kap 3, § 3-14: "Sluttvurderinga skal gi informasjon om kompetansen til eleven, lærlingen, lære kandidaten eller praksisbrevkandidaten ved avslutninga av opplæringa i fag. Alle elevar, også elevar med individuell opplæringsplan, skal vurderast etter kompetansemåla i læreplanen, jf. § 3-3."

Selv om forskriften overlater mange spørsmål ubesvarte, er den nyttig for den kommende diskusjonen i denne rapporten.

- Har forskningen produsert kunnskap som kan brukes for å konkludere noe om i hvilken utstrekning lærere tilbyr tilstrekkelig variasjon i vurderingssituasjoner?
- Har forskningen produsert kunnskap som kan si noe om i hvilken grad eller hvor ofte lærere tilbyr slike vurderingssituasjoner?
- Hva sier forskningen om i hvilken grad grunnlaget for standpunkt speiler den samlede kompetansen til eleven?

Forskningsspørsmål

I denne rapporten svarer vi på tre forskningsspørsmål, som er utarbeidet sammen med oppdragsgiver (Utdanningsdirektoratet). De tre spørsmålene er:

1. Hvilken forskningsbasert kunnskap finnes om lærers og skoleleders praksis med standpunktvurdering i Norden?
2. Hvilken forskningsbasert kunnskap finnes om kvaliteten på lærers standpunktvurdering og skoleleders bruk av standpunktvurdering i Norden?
3. Hvilke eksempler på tiltak for å øke kvaliteten til standpunktvurdering finner vi i et utvalg land det er naturlig å sammenligne seg med?

Metode

I dette kapitlet beskrives hvordan vi gikk frem for å identifisere undersøkelser relatert til standpunktvurdering. Som det vil fremgå, kan dette kunnskapsgrunnlaget klassifiseres som en systematisk, men ikke uttømmende «oversikt» (jamfør Grant & Booth, 2009). Det betyr at vi har gått systematisk til verks for å finne relevant litteratur, men at vi ikke har benyttet strategier som sikrer uttømmende liste over undersøkelser. Valg av søkestrategi baserte seg på driftstiden for prosjektet som av prosjektmedarbeiderne og oppdragsgiver ble vurdert til å være for kort for å gjennomføre et uttømmende søk. En slik tilnærming svekker selvsagt muligheten til å trekke definitive slutninger, men som det vil komme frem av beskrivelsen av søkestrategien, er det sannsynlig at vi har identifisert en vesentlig del av

den forskning som finnes på området standpunktvurdering som er relevant for den norske konteksten.

Søkestrategi

Vi benyttet en «manuell» søkestrategi i tidsskrifter og databaser som med høy sannsynlighet ville inneholde relevante treff (se mere i «Fase 2: søk»). De manuelle søkene foregikk ved at et utvalg søkeord ble brukt på det enkelte tidsskriftets egen søkefunksjon, for eksempel ved å søke på ordet «summativ» i tidsskriftet *Nordic Journal of Literacy Research*. I tillegg benyttet vi oss av en «snøballmetode» (dvs. skanning av litteraturlister i innhentede artikler) for å fange relevant litteratur som ikke inngikk i trefflister. Vi avgrenset søkene til tidsrommet januar 2005–mai 2021.

Fase 1: Søketermer, kilder og søk

Søkene foregikk i mai 2021 i databasene til enkelte tidsskrifter og i regelrette databaser. Vi gjennomførte søk i følgende tidsskrifter:

- *Acta Didactica Norge*
- *Acta Didactica Norden*
- *Educare*
- *Education Inquiry* (EI)
- *Nordand*
- *Nordic Journal of Literacy Research*
- *Nordic Studies in Education*
- *Nordisk tidsskrift for utdanning og praksis* (tidl. *FoU i Praksis*)
- *Sakprosa*
- *Scandinavian Journal of Educational Research* (SJER)

To av tidsskriftene (EI og SJER) hadde «avanserte» søkefunksjoner, mens øvrige hadde enklere søkefunksjoner. I disse søkte vi på ett og ett begrep (se tabell 1), mens vi i EI og SJER kunne bruke mere avanserte søkestrenger.

I tillegg til tidsskriftene søkte vi i databasene *DiVA Portal* og *Google scholar*. I svenske DiVA brukte vi på søkeordet «betyg». I Google scholar brukte vi «standpunktvurdering» og «standpunkt karakter». Av pragmatiske årsaker konsentrerte vi oss om treff høyt

opp på trefflisten i forbindelse med søk i Google scholar. I tabell 1 nedenfor framkommer hvilke søkeord som ble brukt til de ulike kildene.

Tabell 1

Søkestermer og antall treff

Kilder	Søkeord	Innledende bruttoliste	Bruttoliste etter første skanning	Nettoliste
Tidsskrift Enkel db	Betyg	14	14	5
	Grade	101	101	5
	Karakter	50	50	4
	Standpunkt	8	8	0
	Summativ	10	10	0
Tidsskrift Avansert db	[All: grade] AND [All: assessment] AND [in *Journal*] AND [Publication Date: (01/01/2005 TO 5/31/2021)]	598	73	16
Databaser	Betyg	26	26	8
	Standpunkt vurdering	355	3	3
	Standpunkt karakter	843	0	0
Snøball	-	3	0	3
Summer	Søkeord	2008	285	44

NB. «db» = database.

Fase 2: Utvalg av publisert forskning

I alt genererte søkene 2008 treff. Etter første skanning var endelig bruttoliste 285 artikler.

Før inkludering i endelig materiale ble de 285 artiklene gransket ut fra følgende kriterier:

Inklusjonskriterier

- Forskingen skal være publisert i tidsrommet 2005–2021.
- Forskingen skal i utgangspunktet være publisert i fagfelleverderte tidsskrift; unntak kan gjøres for doktorgradsavhandlinger eller rapporter som vurderes å være av vesentlig interesse for kunnskapsgrunnet.
- En vesentlig del av den rapporterte forskningen skal omhandle standpunkt vurdering eller summativ vurdering.
- Forskingen skal relatere til de nordiske landene.

Eksklusjonskriterier

- Forskning hvor standpunktvurdering eller summativ vurdering behandles som et kvalitetssikret kriterium (f.eks. undersøkelser av karakterfordeling mellom jenter og gutter).
- Forskning hvor standpunktvurdering eller summativ vurdering nevnes, men hvor vurdering ikke *per se* undersøkes (f.eks. i forskning hvor vurdering inngår som en opprømsing om aspekter av lærers arbeid, uten nærmere undersøkelse).

Bruken av disse kriteriene resulterte i et utvalg på 44 undersøkelser (se tabell 1). Av dem var 2 undersøkelser «dubletter», det vil si artikler (hhv. Hovdhaugen et al., 2018 og Prøitz, 2013) som gjengav resultater som allerede var blitt mer utfyllende presentert (hhv. Hovdhaugen et al. 2014 og Prøitz & Borgen, 2010). Mens artiklene var fagfellevurderte, var de to sistnevnte rapporter som ikke er fagfellevurdert. Av pragmatiske grunner valgte vi likevel å analysere rapportene fordi de var betydelig mer omfattende og de utdypet forskningsresultatene. Tilsammen bygger denne kunnskapsoversikten altså på 42 undersøkelser.

Materialets karakter

Av de undersøkelsene som ble inkludert i materialet, hadde flest studert ungdomstrinnet ($n = 21$), deretter vgs ($n = 13$) og barneskolen ($n = 12$). To undersøkelser var trinnuavhengige. I alt hadde 933 lærere, 1818 elever og 35 rektorer deltatt i undersøkelsene. Videre hadde henholdsvis 305 dokumenter og 3 357 8000 karakterer/prøveresultater blitt analysert.

Det var vanligst med faguavhengige undersøkelser ($n = 24$) og undersøkelser av vurdering i morsmålsfaget ($n = 8$). Naturfaget var representert tre ganger, mens følgende fag var representert én gang: kroppsøving, fysikk, engelsk, matematikk, samfunnsfag og praktisk-estetiske fag. Se Appendix for en oversikt over hvilke undersøkelser som inngår i materialet.

Selv om materialet favner over store mengder elever og lærere i forskjellige aldersgrupper og forskjellige fag, vil det være mange kombinasjoner av fag, trinn og deltakere som ikke dekkes av de undersøkelsene som inngår i denne kunnskapsoversikten.

Koding av utvalg av publisert forskning

Hensikten med kunnskapsoversikten gjorde at vi valgte å bruke kjernebegrepene i forskningsspørsmålene, *praksis* og *kvalitet i standpunkt*, som primære kategorier. Vi kodet alle artikler som beskrev praksis uten å vurdere dem som praksis-undersøkelser. Praksiskategoriene delte vi videre opp i to koder: praksis og erfaring med praksis for å skille mellom undersøkelser av praksis og undersøkelser av gjengivelser av oppfatninger av praksis.

Vi kodet alle artikler som undersøkte standpunktvurdering ved å vurdere den, eller et aspekt av den, som kvalitetsundersøkelser. Med vurdering av standpunkt menes her alle typer av undersøkelser som kan relateres til validitet eller reliabilitet i standpunktvurdering. Eksempel på dette kan være undersøkelser hvor en studerer hvor godt standpunkt karakter kan "forutsi" skolefremgang sammenlignet med eksamenskarakter. Et annet eksempel kan være undersøkelse av læreres samstemmighet i forbindelse med vurdering av elevarbeid.

I forbindelse med koding av materialet laget vi også en kode for undersøkelser som foretok systemanalyser (f.eks. ved å undersøke konsekvenser av ulike typer vurderingssystem). De kategorier vi brukte, er ikke nødvendigvis gjensidig utelukkende, siden en undersøkelse for eksempel kan ha som formål både å beskrive en praksis og å vurdere den. Vi fant ingen artikkel som var nødt til å kodes i mer enn én kategori.

I alt brukte vi disse fire kategoriene for å kode artiklene: *praksis*, *erfaring av praksis*, *systemanalyse* og *kvalitet i standpunktvurdering*. Dette er fire brede kategorier som korresponderer til forskningsspørsmålene og som vil brukes til organiseringen av resultatkapitlet. Under hver kategori har vi gruppert undersøkelser på en måte som vi håper oppfattes som leservennlig, vel bevisst om at det finnes mange organisasjonsprinsipp som kunne vært like aktuelle.

Resultat

Forskningsspørsmål 1: Hvilken forskningsbasert kunnskap finnes om læreres og skolelederes praksis med standpunktvurdering i Norden?

Praksis

Vi identifiserte fire artikler som rapporterte undersøkelser av praksis. To av disse kan sies å ha tatt et elevperspektiv ettersom undersøkelsene dreide seg om ukeprøver (Flaten & Volan, 2018) og elevers erfaring med å få standpunkt karakter (Olovsson, 2014). Flaten og Volan (2018) undersøkte 42 ukeprøver på barneskolen og fant at lærere ofte stilte spørsmål som hadde en "kontrollerende" funksjon, dvs. var utformet for å kontrollere om elever lært seg bestemte, avgrensede aspekter (som for eksempel rettskriving). Olovsson (2014) fulgte opplæringen i to 5. trinnsklasser i Sverige og konkluderte med at vurderingspraksisen i disse klasserommene var forskjellig, blant annet når det kommer til transparens og hva som var fokus for vurderingen. Én av årsakene til dette, mener forfatteren, kunne ha vært eksterne faktorer som påtvunget vs. ikke-påtvunget bruk av læringsplattformer.

De to resterende praksis-undersøkelsene studerte henholdsvis kunst- og håndverklæreres standpunktvurdering (Lutnæs, 2011) og samtaler mellom mentorer og nyutdannede lærere, som underviste i fag de ikke hadde formell kompetanse i (Føinum, 2016). Lutnæs (2011) fulgte syv KH-lærere i forbindelse med at disse samlet informasjon til og gjennomført standpunktvurdering. Lutnæs (2011) konstaterte at lærerne hadde et «velfungerende språklig repertoar», at de vektla visuelle produkter i vurderingen, men at de strevde med å konkretisere hva som bidro til originalitet i elevarbeidene.

Selv om fokus for Føinum (2016) ikke var formell standpunktvurdering eller vurdering som lå til grunn for standpunkt, viste analysen av den ene læreren at han var usikker i vurderingssituasjonen i engelskfaget: «I utdrag 1 kommer det frem hvor ukomfortabel Roar er når han arbeider med vurdering i engelsk. Han bruker mye tid, er bekymret for hvorvidt han klarer å være konsekvent, og om han greier å gi elevene konstruktive tilbakemeldinger» (s. 300).

Erfaring av praksis

Vi identifiserte 23 publikasjoner som presenterte undersøkelser av erfaring av praksis. Av disse omhandlet halvparten læreres og skolelederes erfaring og den andre halvparten elevers erfaring av lærers praksis med vurdering. Av tretten "lærerundersøkelser" var seks norske, seks svenske og en finsk.

Norske undersøkelser av læreres erfaring. Av de norske undersøkelsene er Prøitz og Borgen (2010) den mest omfattende. Forskerne intervjuet 41 lærere og 23 skoleledere på ungdomstrinn og videregående skole om praksis i forbindelse med standpunkt-vurdering. Noen av de mest sentrale funnene var at informantene hevdet at standpunkt i hovedsak var et uttrykk for oppnådd kompetanse, noe som kan tolkes som at standpunkt-vurderingen gjenspeiler elevers kompetanse heller enn andre faktorer, samtidig som det for alle fag unntatt kroppsøving ble meldt om at elever med sterke prestasjoner fikk vurdering i deltakelse som del av ordenskarakteren og elever med svakere prestasjoner fikk vurdering av deltakelse som en del av standpunkt-vurderingen i faget. Videre rapporterte de at standpunkt ble oppfattet å ha to funksjoner: som informasjon og som motivasjon.

Informantene rapporterte også at standpunkt-vurderingen bygde på et omfattende grunnlag, at skriftlige arbeider til eleven er det viktigste grunnlaget på tvers av fag, samtidig som de meldte at det var delvis uklart om standpunkt skal speile kompetanse etter endt opplæring og/eller elevers faglige utvikling. Forfatterne kunne konstaterte en del variasjon i hvordan lærere forsto "rettferdig vurdering", selv om alle informanter var opptatt av nett-opp rettferdighet. Et fellestrekk blant informantene var ellers at systemer for vurdering (kriteriesett, poengsystemer og lignende) bidro til en økt følelse av rettferdighet. Forfatterne rapporterte også at ledelsen hadde kjennskap til bestemmelser for standpunkt-vurdering.

Oltedal et al. (2015) undersøkte erfaring med blant annet standpunkt-vurdering til 28 lærer fra praktisk-estetiske fag. Resultatene indikerte at lærere, slik som de som ble intervjuet av Prøitz og Borgen, belønnet høy innsats med økt karakter i faget, men så ut til å la manglende innsats og oppførsel gå ut over ordenskarakteren. Videre mente informantene at de

selv, elever og lærere forventet mye av elevene og at det var nær sagt umulig å få lavere karakter enn 3 i de praktisk-estetiske fagene, fordi alle elever hadde kompetanse nok til å vise noe.

I Birkelund og Midthaugen (2019) rapporteres en undersøkelse av nyutdannede kroppsøvlingslærere på ungdomstrinn og vgs. Standpunktvurdering var ikke et stort tema i undersøkelsen, men lærere i undersøkelsen fortalte at det er vanskelig å sette karakterer, særlig fordi det kunne innebære en *vanskelig relasjon til foresatte*. De nyutdannede lærerne erfarte også at eldre kollegaer hadde andre strategier for å sette karakterer (f.eks. tester), og de nyutdannede fikk få muligheter til å introdusere nye måter å tenke og gjøre vurdering på.

I en norsk-svensk undersøkelse (Brante & Lund, 2017) ble 18 lærere, hvorav 17 var lærere i henholdsvis norsk og svensk, intervjuet om opplæring i kildebruk og kritisk lesing. Selv om undersøkelsene ikke var direkte relatert til standpunktvurdering, fremkom det at de svenske lærerne hadde en «mekanisk» innstilling og relaterte kilder til karakterer (dvs. analyserte hvordan opplæringen kunne tilpasses karakterkrav), mens de norske lærerne var mer opptatt av hva kildebruk innebærer i norskfaget.

Til slutt skal nevnes to relaterte undersøkelser. Eriksen og Elstad (2019) intervjuet lærere på skoler som hadde innført karakterfri opplæring og mente i sin hovedkonklusjon at «norsk-lærere i videregående skoler i overveiende grad har positive erfaringer med karakterfritt semester: Karakteroppmerksomheten avtar med denne ordningen, og bidrar til å øke elevens oppmerksomhet om læringsframgang basert på lærerens tilbakemeldinger» (s. 4).

I Gillespie og Burner (2019) ble fire rektorer fra skoler som innførte karakterfritt semester intervjuet. Rektorene fortalte at man gjennomførte karakterfri skole på bakgrunn av forskning. Rektorene mente også at karaktersystemet i Norge er utfordrende fordi elever fokuserer mest på gjennomsnittlig standpunkt, og at det er vanskelig å få alle parter involvert i forsøk med karakterfri skole. Som en viktig suksessfaktor for hvorfor innføringen av karakterfri skole hadde lyktes, framhevet undersøkelsen bruken av begreper om vurdering, for

eksempel ble begrepet «vurderingssituasjon» brukt heller enn «test». Andre suksessfaktorer som ble rapportert, var lærerinvolvering, tidlig start med karakterfri skole, og prosesser som starter fra «bunnen» av heller enn fra «toppen».

Svenske og finske undersøkelser av lærers erfaring. De seks svenske undersøkelsene spente over flere tematikker. Jönsson og Klapp (2020) og Wetterstrand et al. (2017) var relaterte til svenske nasjonale prøver, som har en eksamenslignende funksjon, heller enn en kartleggende og læringsstøttende, slik som i Norge. Jönsson og Klapp (2020) undersøkte hvordan lærere forsvarte det å gi et annet *betyg* (dvs. standpunktkarakter) enn det som tilsvarte resultatet på nasjonale prøver. Dette kan sammenlignes med spørsmålet om hvordan norske lærere forklarer hvorfor standpunktkarakter skiller seg fra eksamenskarakter, selv om nasjonale prøver ikke vurderes eksternt. De intervjuede Lærerne presenterte tre grunner til at det er forskjeller mellom betyg og resultat på nasjonale prøver (NP): 1) at resultater på NP-resultater kan være misvisende fordi prøvene er feilkonstruerte; 2) mangler blir rettet opp, ved at læreren oppdager «mangler» i elevenes kunnskap i forbindelse med NP og gir tilpasset opplæring for å rette opp disse manglene, slik at elevene presterer bedre fram mot fastsettelse av «betyg», og 3) ekstern påvirkning, for eksempel ved å gi en ufortjent god karakter av redsel for foresattes reaksjon.

Wetterstrand et al. (2017) undersøkte læreres innstilling til at den svenske myndigheten Skolinspektionen kontrollerer («omrättar») læreres vurdering av elevers nasjonale prøver. Resultatene indikerte at lærerne oppfattet økt fokus på «omrättning» i massemedia kunne få negative konsekvenser, samtidig som de sa at de likte nasjonale prøver og skulle ønske seg tilbakemelding på individnivå om hvordan de selv klarer seg i «omrättningen».

Øvrige svenske undersøkelser hadde ikke felles tematikk, men var nedslag i ulike aspekter av standpunktvurdering. Strandler (2016) observerte og intervjuet lærere før og etter innføring av standpunktvurdering på 6. trinn i Sverige. Lærerne rapporterte om økt fokus på å dokumentere elevers kunnskap og økt fokus på å iscenesette situasjoner som innebar at elevene fikk vise hva de kunne. Olovsson og Näsström (2018) presenterte en undersøkelse

av læreres svar på om de ga elever samlekarakter i samfunnsfag og naturfag eller karakterer i hvert enkelt av de fagene som sammen utgjør Samfunnsfag(ene) og Naturfag(ene). Resultatene viste at begge måtene å gi karakter på forekom. Olovsson (2020) studerte hvordan den nye svenske læreplanen påvirket opplæringen og kunne notere at lærerne meldte at den nye læreplanen hadde konsekvenser for hvordan de bedrev opplæringen. Blant annet meldte en lærer at læreplanens krav om én standpunktkarakter i respektive samfunnsfag gjorde at hun underviste fagene hver for seg i stedet for å undervise i ett samfunnsfag. Alm og Colnerud (2015) hadde studert læreres erfaring med urettferdig vurdering og fant at lærere i høy utstrekking rapporterte at de hadde vært utsatt for urettferdig vurdering i egen skolegang.

Til slutt skal den finske undersøkelsen av Huhtala et al. (2019) nevnes. Denne artikkelen rapporterte om en tradisjonspreget opplærings- og vurderingspraksis. Det er uklart hvor overførbare resultatene er til norsk skole, gitt at de finske og norske skolesystemene skiller seg fra hverandre i større grad enn svensk og norsk skole.

Undersøkelser av elevers erfaring. Samtlige undersøkelser av elevers erfaring med standpunktvurdering var svenske. To av undersøkelsene studerte erfaringer til elever på 6. trinn, som er det første skoleåret med karakterer i Sverige. Löfgren et al. (2019) noterte en variasjon i hvor mye elever engasjerte seg i karakterer. Mens noen svarte at karakterer ikke var særlig viktig, svarte andre at karakterer kunne ha mye å si for fremtiden. Tanner (2016) intervjuet elever i samme aldersgruppe og mente at elevene viste tegn på å "gjøre policy", det vil si å ha inntatt rollen som elever i et skolesystem hvor en får karakterer. Elevene rapporterte også om å være trøtte av stadig å bli vurdert.

To andre undersøkelser var Sivenbring (2019) og Torkildsen og Erickson (2016), som begge intervjuet elever om blant annet hvor godt elevene forsto tilbakemelding i form av standpunktvurdering. I undersøkelsen fremkom det for eksempel at elevene rapporterte om å ha vanskeligheter med å skjønne motiveringer for karakterer på prøver, og de hadde vansker med å skjønne hva som må til for å få en gitt standpunktkarakter (Sivenbring, 2019). Også Torkildsen og Erickson (2016) rapporterte at elevene oppga at de slet med å skjønne

hva som lå til grunn for standpunktvurdering. I Knetka (2017), til sist, ble elevers testmotivasjon undersøkt. Elever rapporterte om å være mer motiverte for å gjøre sitt beste i forbindelse med prøver som fikk store konsekvenser. Forskeren så også at om eleven hadde investert like mye innsats i prøver med lite konsekvenser, kunne eleven fått en høyere skår. Dette har implikasjoner for standpunktvurdering, fordi lærerens innramming kan påvirke hvor mye en elev investerer av innsats i en vurderingssituasjon i klasserommet.

Resterende fem svenske undersøkelser var mer perifere med tanke på tematikken for denne kunnskapsoversikten, men likevel relaterte til spørsmålet om praksis. Felles for disse undersøkelsene var at standpunktvurdering var noe som fremsto som viktig for elever: som en måte å "være svensk" på (Andersson & Lunneblad, 2020), som et tegn på egen kompetanse (Harling, 2018), som viktig for foresatte (Löfgren & Aman, 2020) og som noe som kan kreve spesielle strategier, for eksempel å ikke stille spørsmål i timene og derved risikere å blottlegge manglende kunnskap (Månsson et al., 2021). En siste undersøkelse i kategorien elevers erfaringer er Tholin (2010), som hadde studert elevstemmer i spørsmålet om standpunktkarakter. Materialet var presseklipp fra årene 1932 til 1990 og i 2009. Tholin (2010) dokumenterte at elever var "usynlige" i starten av materialet, for så å komme med kritiske innspill til karakter i 1970-årene. I 2009 var elever flest positive til karakter.

Oppsummering av praksis-undersøkelser

Som det fremgår av reviewen av praksisundersøkelsene har disse produsert lite utdypende kunnskap om hva lærere og rektorer *gjør* i forbindelse med vurdering. Undersøkelsen til Flaten og Vollan (2018) sier noe om hva lærere synes å ha vektlagt, Olovsson (2014) noe om hvordan elever oppfatter vurdering, mens Lutnæs (2011) og Føinum (2016) sier noe om enkelte læreres praksis og utfordringer.

Et fellestrekk for disse undersøkelsene er at de baserer seg på små, ikke-representative utvalg. Dette forringer på ingen måte resultatene, men gjør det vanskelig å trekke konklusjoner om praksis på generelt grunnlag. At én lærer er bekymret eller at syv lærere sliter med å definere originalitet, sier lite om praksis med standpunktvurdering generelt.

De norske studier som dekte erfaring av praksis kan sies å ha vist at lærere synes å ta standpunkt-vurdering på stort alvor og at standpunkt ifølge selvrapporteringen bærer preg av systematikk. Disse undersøkelsene vitner også om usikkerhet om hva som skal inngå i standpunkt-vurderingen, hvordan en skal skaffe «bevis» på måloppnåelse og også variasjon på tvers av lærere og fag. Undersøkelsene om karakterfritt semester er interessante fordi de delvis svarer på utfordringen om omfattende dokumentasjonskrav til læreren. Samtidig er det viktig å huske at selv den mest omfattende undersøkelsen (Prøitz & Borgen, 2010) bygger på et svært lite materiale, at den ikke har undersøkt praksis, men erfaring av praksis og at den dessuten er mer enn ti år gammel. Dette forringer ikke resultatene, men er en påminning om behovet for undersøkelser som enten dekker et større og mere representativt utvalg og/eller som mer inngående studerer den faktiske praksisen med standpunkt-vurdering på tvers av fag.

Selv om hverken det svenske eller finske skolesystemet er identisk med norske, er det svenske så likt at resultatene fra undersøkelsene om læreres selvrapporterte erfaring kan ha en viss interesse. Det gjelder særlig undersøkelsene til Jönsson og Klapp (2020) og Wetterstad et al. (2017), som på ulike måter berørte nasjonale prøver. Også i Norge forekommer diskusjoner om relasjonen mellom standpunkt-vurdering og eksamenskarakter (Hovdgauhen et al., 2014) og også i Norge har man foretatt undersøkelser for å sjekke vurderersamsvar på eksamen (Björnsson & Skar, 2021). Hvis de svenske resultatene har hold også i Norge kan dette ha implikasjoner for myndighetenes kommunikasjon om relasjonen standpunkt–eksamen og om tiltak for å sikre rettferdig karaktersetning.

Undersøkelsene relatert til elevers erfaring kan si noe om hvordan elever oppfatter og opplever lærers praksis. Gitt den tematiske spredningen og de små utvalgene elever er funnene ikke uten videre generaliserbare. Oppsummert kan en likevel si at felles for disse undersøkelsene er at elever stort sett, så vel som lærere, virker å ta standpunkt-vurdering på alvor. De indikerer også at det kan være viktig å diskutere med elever hva en standpunkt-vurdering faktisk representerer. For eksempel skulle eleven som tenker at standpunkt-vurdering på 6. trinn (i den svenske skolen) er viktig for fremtiden, kunne tilbys et annet perspektiv på nettopp den innsikten.

Forskningsspørsmål 2: Hvilken forskningsbasert kunnskap finnes om kvaliteten på lærers standpunktvurdering og skolelederes bruk av standpunktvurdering i Norden?

Vi identifiserte 13 undersøkelser relatert til kvaliteten på læreres standpunktvurdering. En gruppe undersøkelser var relatert til karaktersetting og bruk av vurderingsskalaer. En annen gruppe undersøkelser hadde studert prediksjonskraften i standpunkt karakter. Vi identifiserte også to systemanalyser, ved siden av de 13 kvalitetsstudiene.

Karaktersetting og bruk av vurderingsskalaer

Relasjonen standpunktvurdering og ekstern vurdering. To undersøkelser studerte relasjoner mellom standpunktvurdering og ekstern vurdering. Gitt at den ene låses som et kriterium (oftest ekstern vurdering som for eksempel eksamen), kan forskere i denne typen undersøkelser studere faktorer som ser ut til å systematisk samvariere (korrelere) med sprik mellom standpunktvurdering og ekstern vurdering. Vi fant to undersøkelser av dette slag, én norsk og én svensk.

Den norske undersøkelsen (Hovdhaugen et al., 2014) dreide seg om to separate spørsmål. Først om det var mulig å notere en systematisk forskjell i avvik mellom standpunkt- og eksamenskarakter for offentlige og private videregående skoler i årene 2010–2013. Forskerne konkluderte med at en slik forskjell var til stede og at avviket mellom standpunkt- og eksamenskarakterer var noe større i private allmenne videregående skoler.⁴ Slike systematiske avvik kan tolkes som systematisk overvurdering på private skoler, men like gjerne som systematisk undervurdering på offentlige skoler. (Uansett bygger tolkingen på at eksamen er et gyldig kriterium.) Forskerne noterte også at det som på elevnivå best predikerte størrelsen på avstanden mellom standpunkt- og eksamenskarakter var høye karakterer, slik at elever med høye standpunktkarakterer hadde størst risiko for å få lavere karakter på eksamen.

⁴ De private allmenne skolene er skoler som tilbyr studieforberedende programmer og der flere av skolene ble godkjent på grunnlag av Friskoleloven før omlegging til ny privatskolelov i 2005.

Det andre spørsmålet til Hovdhaugen et al. (2014) dreide seg om læreres og elevers erfaring av standpunktkarakterer på henholdsvis offentlige og private skoler. Åtte skoler inngikk i undersøkelsen. Forskerne konkluderte at eierform ikke så ut til å kunne forklare variasjon i det selvrapporterte arbeidet med standpunkt, men at det heller gikk en skillelinje mellom de fag som var representert i undersøkelsen – norsk og matematikk – hvor matematikklærerne ga uttrykk for en mer «aritmetisk» tilnærming til vurdering (dvs. teller poeng og gir karakter i henhold til fastsatt poenggrense) og norsklærerne en mer «analytisk» tilnærming (dvs. vurderer elevsvar i lyset av vurderingskriterier). Et annet skille mellom matematikk- og norsklærere var innstillingen til sluttprøver. I en oppsummering skrev forskerne: «Mens matematikklærerne mener at en sluttvurdering er mulig fordi faget matematikk på Vg3 studiespesialiserende utdanningsprogram er ettårig, viser norsklærerne til at norskfaget på videregående avsluttes først etter tre år og at en sluttvurdering ikke kan dekke alle fagets kompetansemål. Norsklærerne ser følgelig behovet for å «lukke» kompetansemål i løpet av Vg3, mens matematikklærerne i prinsippet ser det som mulig å kunne holde samtlige kompetansemål i faget «åpne» frem til og med siste vurderingssituasjon i skoleåret» (Hovhaugen et al., 2014, s. 48).

Elever ble også intervjuet av Hovdhaugen et al. (2014) og i likhet med lærere var det umulig for forskerne å konkludere med systematiske forskjeller relatert til eierform. Forskerne fant imidlertid andre resultater som var av interesse. For det første oppfattet noen elever vurderingen i norsk som diffus og personavhengig. For det andre, og i kontrast til dette, mente forskerne at elever likevel stort sett ga uttrykk for at de var godt kjent med kriterier. En mulig forklaring til det paradoksale funnet var ifølge forskerne at den diffuse og personavhengige vurderingen i norsk ikke nødvendigvis var representativ.

En svensk undersøkelse som også studerte relasjonen mellom standpunktkarakter og resultat på en eksternt lagd (men ikke vurdert) prøve, var Lekholm (2011). Forskeren noterte at skoler med foresatte med lav utdanning så ut til å bruke et kompensatorisk prinsipp ved å gi elevene høyere standpunktkarakterer enn hva resultat på nasjonale prøver tilsa at de burde ha.

Bruk av vurderingsskalaer. Fire artikler rapporterte undersøkelser av bruk av vurderingsskalaer. Tre var svenske, én var finsk. I Blomqvist (2018) fikk en liten gruppe lærere snakke om og komme til enighet om karakterer på elevtekster. Deretter fikk samme lærere fylle ut individuelle karakterskjemaer. Samtalene indikerte god samstemmighet, men analysen av individuelle skjemaer indikerte et betydelig sprik i vurdering. Disse resultatene indikerte med andre ord at lærere etter samtaler kan oppfatte å være nokså samstemte, uten at det har empirisk hold i den faktiske vurderingen.

Tengberg og Skar (2016) undersøkte samstemmighet i vurdering av åpne oppgaver i lesedelen av nasjonale prøver i svensk. Analysen viste en betydelig variasjon mellom lærere og at elevresultater ikke uten videre var sammenlignbare.

Jönsson et al. (2021), gjennomførte en eksperimentell undersøkelse med fokus på om analytisk eller holistisk vurdering ville gi høyest samstemmighet i vurdering. Resultatene indikerte høyere samstemmighet ved analytisk vurdering.

I den finske undersøkelsen (Harju-Luukkainen et al., 2016) studerte forskerne relasjonen mellom PISA-resultater og standpunktvurdering i morsmålsfaget. Forskerne mente at mange lærere i Finland undervurderte sine elever, og at en årsak til det kunne ha vært at kriteriene i læreplanen ikke var klare nok, eller at lærere ikke var gode nok til å vurdere.

Prediksjonskraft i standpunktkarakter

I en serie svenske storskalaundersøkelser ble den prediktive validiteten til standpunktkarakterer studert (Cliffordson, 2008; Cliffordson & Askling, 2006; Lyrén, 2008; Thorsen, 2014; Thorsen & Cliffordson, 2012; Svensson, 2004). Felles for disse undersøkelsene var spørsmål om hva som best kunne «forutse» skolefremgang: standpunktkarakter eller resultat på standardiserte prøver (som f.eks. eksamen). I samtlige undersøkelser viste det seg at standpunktkarakterer var den beste prediktoren. Funnene indikerte derved at standpunktkarakter målte noe annet enn hva som er mulig på en enkelt vurderingshendelse som en stan-

dardisert prøve. Cliffordson (2008) mente at den høye prediktive validiteten til standpunkt karakter var ventet, gitt internasjonal forskning på feltet, og at en forklaring nettopp kan være at lærere i sin standpunkt vurdering inkluderte annet enn kun fagkunnskap, som for eksempel selvregulering, motivasjon etc. Disse forklaringene er, imidlertid, ikke systematisk undersøkt i disse svenske studiene. En finsk studie (Mullola et al. 2014) undersøkte imidlertid dette, og fikk støtte for hypotesen at lærere inkluderte mer enn kun fagkunnskap i standpunkt vurderingen.

Systemanalyse

I litteraturundersøkelsen fant vi to artikler som kunne sies å være systemanalyser. Slike undersøkelser studerer ikke en gitt standpunkt karakter, men heller trekk ved eller konsekvenser av ulike typer av vurderingssystemer. Lysne (2006) er av typen beskrivende systemanalyse, og forskeren tegnet historikken til vurderingssystemer i de nordiske land. Undersøkelsen bidrar ikke med kunnskap om kvaliteten til standpunkt vurdering per se, men er et viktig dokument for kunnskap om standpunkt vurdering. Arensmeier (2020), derimot, studerte konsekvenser av tre ulike karaktersystemer i Sverige. Analysen viste at innføring av et kriteriebasert system i midten av 1990-årene resulterte i at mange elever, som i det forrige norm-relaterte systemet hadde vært "lavtpresterende", nå fikk karakteren ikke bestått. Teknisk var det slik at laveste karakter i det norm-relaterte systemet var 1, mens laveste karakter i det nye systemet var Ikke godkänd (IG) og IG = 0. Konsekvensen av den nye måten å telle kompetanse på var at mange elever ikke lenger hadde krav på plass i den videregående opplæringen, siden den krevde et gitt antall bestått-karakterer.

Oppsummering av kvalitetsundersøkelser

Oppsummert kan en si at forskningen om kvalitet på standpunkt vurdering gjennom Hovdhaugen et al. (2014) har gitt kunnskapsbidrag om relasjonen standpunkt-eksamen (med større systematiske avvik i private allmenne skoler enn i offentlige skoler), og om lærer- og eleverfaringer ved private og offentlige skoler. Ellers har (den fortrinnsvis svenske) forskningen gjort nedslag i bruk av vurderingsskalaer og i prediksjonskraften til standpunkt karakterer. Undersøkelsene av bruk av vurderingsskalaer var småskala og ikke generaliserbare. Likevel peker særlig funnene i Blomqvist (2018) på et område som kan være

verdt å undersøke nøye også i Norge. Funnene til Blomqvist (2018) er jo en sterk indikasjon på at lærere kan oppleve å bli oppfattet til å være samstemte, selv når dette ikke er tilfelle.

Undersøkelsene om prediksjonskraften til standpunktkarakterer har entydig vist at lærerkarakterer er bedre på å forutsi senere skolefremgang, enn resultater fra enkeltprøver. Fordi dette er et funn som også går igjen internasjonalt (Brookhart et al., 2016), er det grunn til å tro at det samme gjelder i Norge. En forklaring som blant andre Brookhart et al (2016) har ført frem, er at den relative suksessen til standpunkt i forhold til enkeltprøver skyldes at standpunkt vurderer mer enn kun fagkompetanse.⁵

Forskningsspørsmål 3: Hvilke eksempler på tiltak for å øke kvaliteten til standpunkt-vurdering finner vi i et utvalg land det er naturlig å sammenligne seg med?

Problemstillinger rundt standpunktkarakterer er et internasjonalt kjent fenomen, men med ulike perspektiver, avhengig av hvilket land vi studerer. Allerede i 2006, påpekte Black og Wiliam internasjonale studier som viste manglende reliabilitet ved eksamen, og myndigheters vegring for å gå inn i kritikk av eksisterende vurderingssystemer.

Slik sett var Norge tidlig ute med åpenhet rundt manglene ved standpunktvurdering og dens reliabilitet (manglende pålitelighet). Som vi påpekte innledningsvis, finnes det studier som dokumenterte manglende sammenheng mellom standpunktkarakterer og eksamen, med gjennomgående lavere resultat ved skriftlig eksamener, dokumentert av Eksamenssekretariatet, Læringscenteret og KUF allerede fra tidlig 1990 -tall (Tveit, 2007). På samme måte som man i Norge har arbeidet for en bedre vurderingspraksis i skolen, har man eksempler på tiltak også utenfor Norden.

Dersom man ser til land det er naturlig å sammenligne seg med, finner man at ulike tiltak har vært satt inn for å forbedre læreres vurderingskompetanse, og da spesielt med tanke på summative vurderinger.

⁵ Brookhart et al. (2016) skriver: «This quality of graded achievement as a multidimensional measure of success in school may be what makes grades better predictors of future success in school than tested achievement» (s. 834).

I England har man i flere tiår sett manglende tillit til lærere sin evne til å vurdere, og myndighetene har derfor valgt et system hvor eksternt produserte tester avholdes og deretter vurderes av eksterne *markers*. Dette er ofte tidligere *lærere* som har fått opplæring og dermed deltar i vurderingsarbeidet.

Selv om systemet i England har endret seg noe fra 1990 med antall tester, andel prosenter av det som omtales som «coursework» som lærere vurderer, og selve testene som er i bruk, er det dessverre gjennomgående et syn at lærere ikke kan gi standpunkt karakterer eller summative vurderinger i fag på en reliabel og valid måte; *they can't be trusted*.

Utfordringen er som blant annet Gardner (2007) har påpekt, et system hvor lærere får lav status, fordi kontroll over vurderingen er tatt fra dem. Det resulterer i manglende vurderingskompetanse, fordi lærere ikke praktiserer vurdering, de utvikler ikke vurderingsoppgaver og er ikke deltakere i praksiser som f.eks. *mediering av vurdering*. I profesjoner som medisin og jus, er dette motsatt, noe som også gjør at disse profesjonene har høyere anseelse.

Et av de foreslåtte tiltakene fra England, er at man bruker *standardisering* og *moderering*. Dette kom blant annet som et resultat av en studie ved King's College i London, hvor Black ledet en eksplorerende, kvalitativ intervensjonsstudie ved tre skoler, hvor forskerteamet over tre år samarbeidet med matematikk og engelsklærere i tre skoler med undervisning av elever aldersgruppen 12 – 16 år. Studien, som er kjent som *The King's Oxfordshire Summative Assessment Project (KOSAP)*, undersøkte blant annet hvilke tiltak som må settes inn for å fremme kvaliteten på læreres summative vurderinger.

Prosjektet varte over to år og resulterte i råd om hvordan læreres summative vurderinger kunne forbedres. Blant annet var tiltakene rettet inn mot at lærere skulle få en felles forståelse av vurderingskriterier som ble benyttet. Forskerteamet konkluderte med at både *Moderation* og *Standardisation* kan brukes. Standardisation vil si at man trener på å sette ka-

rakterer ved å øve seg med bruk av eksempler, dette kan gjøres internt på en skole. Moderation er mer krevende, da man blir bedt om å vurdere autentiske elev eksempler (anonymisert) fra egen og andres skoler. I KOSAP studien kom det frem at lærere opplevde moderation som svært givende, da det ble avdekket ulike vurderingskulturer mellom skolene som ble drøftet slik at et tolkningsfellesskap kunne utvikles. Samtidig opplevde mange dette som krevende, da egen vurderingskompetanse kan bli utfordret. Slik kan moderering bli sett som et eksempel på tiltak som kan styrke læreres evne til å gi pålitelige standpunkt karakterer.

I deler av Australia, har man også hatt erfaring med moderation, beskrevet i Black et al. (2011). Delstaten New South Wales og Queensland har praktisert moderation i stor skala, med utviklet materialet fra myndighetene i samarbeid med skoler og lærere. Dette har ført til økt tillit til læreres summative vurderingskompetanse, og standpunkt karakterer har ofte vært basert på læreres vurdering/moderation og eksternt vurderte tester (Stanley et al. 2009).

Lignende tiltak har funnet sted i Skottland, hvor man har hatt fokus på å styrke læreres vurderingskompetanse gjennom tilgang til vurderingseksemplar som *National Assessment 5-14 Bank* (www.aiflna.net). Slik har lærere hatt mulighet til å sammenligne egen vurdering med nasjonale retningslinjer, i tillegg til samtaler med kolleger om vurdering. I Skottland har man videre praktisert moderation ved at et utvalg vurderingseksemplar trekkes fra skoler, og vurderes i forhold til nasjonale standarder. I motsetning til England, blir ikke skoler holdt ansvarlige for resultatet av vurderingen, men resultatet brukes i stedet som informasjon for å forbedre vurderingssystemet totalt sett.

Diskusjon

Som vi har påpekt i denne kunnskapsoppsummeringen, har undersøkelser vist at lærer-gitte karakterer, slik som standpunkt, predikerer bedre senere skolefremgang, enn resultater fra enkeltprøver. Dette funnet bør i seg selv styrke læreres anseelse hva vurderingskompetanse angår. Vi kan også argumentere for at standpunkt dermed har større validitet enn eksamen. På den annen side, mangler vi nyere storskala undersøkelser i Norge som

kan etterprøve at standpunkt predikerer skolefremgang. Vi ser dette som spesielt viktig, da vi har lignende funn internasjonalt, men fra nyere studier. Brookhart et al., (2016) har argumentert for at en mulig forklaring på at standpunkt predikerer skolefremgang bedre, skyldes at standpunkt vurderer mer enn kun fagkompetanse.⁶

I et review av internasjonal litteratur om læreres summative vurderingskompetanse fra Harlen et al. (2004), ble det konkludert at et avgjørende tiltak for økt reliabilitet i summativ vurdering, er trening hvor lærere inkluderes i å identifisere vurderingskriterier slik at de kan ha eierskap og forståelse over språket som brukes. I tillegg ble det anbefalt at man setter av tid til arbeidet, men og diskuterer hvordan man unngår bias i vurderingen når man kjenner elevene. Dersom man sammenligner med Norge, vil lærere som har vært sensorer ofte ha gjennomgått lignende praksiser, ved skolering, moderering ved sensurering, og diskusjon rundt eksempeloppgaver og karakterer (Tveit & Olsen, 2018). Utfordringen er at dette ikke er tilfelle for majoriteten av lærere, men kun de som melder seg som sensorer ved eksamen eller nasjonale prøver. I rapporter til Utdanningsdirektoratet har kurs om sensorskolering for nasjonale prøver og eksamensvurdering fått positive tilbakemeldinger fra fylkesmenn gjentagende år. Det er grunn til å undersøke mer systematisk de ulike former for tiltak som er satt i gang i Norge siden tusenårsskiftet med tanke på å fortsette med det som fungerer, og justere det som eventuelt ikke viser effekt. Dette kan synes spesielt viktig da denne kunnskapsoversikten viser at vi fremdeles har lite kunnskap om hva de ulike aktører *gjør i praksis*, de fleste studiene har undersøkt meninger, holdninger og opplevelser fra elever og lærere.

Konklusjon og noen avsluttende refleksjoner

Vi har valgt å fremheve UK, da England og Skottland har hatt spesiell innflytelse på utviklingen av vurderingsarbeidet i Norge, gjennom de ulike reformene fra 2005–2019 (Hopfenbeck et al. 2013, 2015). Samtidig er det viktig med en avsluttende kommentar om kontekstens betydning for de ulike tiltak. UK har fremdeles et mye sterkere *accountability-system*,

⁶ Brookhart et al. (2016) skriver: «This quality of graded achievement as a multidimensional measure of success in school may be what makes grades better predictors of future success in school than tested achievement» (s. 834).

enn Norge, og resultater av standpunktkarakterer blir sett på som *high stake*: skoler i England kan miste økonomisk støtte dersom de oppnår for lave resultater, eller de kan bli satt under administrasjon. Man har valgt en linje hvor reliabilitet i eksamensresultater/endelig karakter, er viktigere enn validitet av tester som gis. Eksterne tester i alle fag, som ved f.eks. GCES (General Certificate of Education) utvikles av eksamens organisasjoner som AQA, og vurderes eksternt. Slik blir læreres vurdering og undervisning mindre sammenkoblet, enn i en norsk skolekontekst hvor lærere underviser og lager egne prøver/halvårs tentamener, basert på lokale skoleplaner/læreplanen.

Dersom man ønsker en mer reliabel standpunktvurdering, kan standardisering av tester gjennom skoleåret være et mulig valg. Samtidig vil en da miste noe av det mange har vurdert som positivt i den norske skole, nemlig valide vurderingsoppgaver som tester det de lokale skolene har undervist. I tillegg vil en engelsk modell sannsynligvis øke presset på elevene, da avgangselever i England ved avsluttende GCES, tar mellom 20 og 28 tester over 4 uker, da de har flere tester i de fleste fag.

En mulig utfordring for fremtiden kan ligge i å balansere ønsket om reliable standpunktkarakterer med valide vurderingsoppgaver som fremmer motivasjon og ny læring. Her kan Norge være i en gunstig situasjon, med lang tradisjon for tett samarbeid mellom forskere, universitetsmiljø og lærere og skoleledere i skolene. Med tanke på studiene vi har sammenfattet, og manglende storskalaundersøkelser som har dokumentert praksis rundt standpunktkaraktersetting, kan det og være nærliggende å etterlyse forskning nettopp på dette området.

Litteratur⁷

- *Alm, F., & Colnerud, G. (2015). Teachers' Experiences of Unfair Grading. *Educational Assessment*, 20(2), 132–150. <https://doi.org/10.1080/10627197.2015.1028620>
- *Andersson, S., & Lunneblad, J. (2020). Elevskapets utmaningar. *Educare*, 14(4), 136–157. <https://doi.org/10.24834/educare.2020.4.6>
- *Arensmeier, C. (2020). Three Decades of School Failure in Swedish Compulsory School. *Scandinavian Journal of Educational Research*, 1–14. <https://doi.org/10.1080/00313831.2020.1833235>
- Bennett, R. (2011) Formative assessment: a critical review, *Assessment in Education: Principles, Policy & Practice*, 18:1, 5-25. <https://doi.org/10.1080/0969594X.2010.513678>
- *Birkelund, I., & Midthaugen, P. (2019). Ny som kroppsøvlingslærer - hvordan oppleves det første året i yrket? *Nordisk tidsskrift for utdanning og praksis*, 13(1), 25–43. <https://doi.org/10.23865/up.v13.1835>
- Black, P., C. Harrison, J. Hodgen, B. Marshall, and N. Serret. 2010. Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice* 17, no. 2: 217–34.
- Black, P., and D. Wiliam. 2007. Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement* 5, no. 1: 1–53.
- *Blomqvist, P. (2018). Är vi överens om det? Samspel och samstämmighet i svensklärares bedömningssamtal om gymnasieelevers skrivande. *Nordic Journal of Literacy Research*, 4(1–21). <https://doi.org/10.23865/njlr.v4.1050>
- *Brante, E. W., & Lund, E. S. (2017). Undervisning i en sammansatt textvärld: En intervjustudie med svenska och norska gymnasielärare om undervisning i kritisk läsning och kritisk värdering av källinformation. *Nordic Journal of Literacy Research*, 3(2). <https://doi.org/10.23865/njlr.v3.671>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A Century of Grading Research. *Review of Educational Research*, 86(4), 803–848. <https://doi.org/10.3102/0034654316672069>
- *Cliffordson, C. (2008). Differential Prediction of Study Success Across Academic Programs in the Swedish Context: The Validity of Grades and Tests as Selection Instruments for Higher Education. *Educational Assessment*, 13(1), 56–75. <https://doi.org/10.1080/10627190801968240>
- *Cliffordson, C., & Askling, B. (2006). Different Grounds for Admission: Its effects on recruitment and achievement in medical education. *Scandinavian Journal of Educational Research*, 50(1), 45–62. <https://doi.org/10.1080/00313830500372026>
- *Eriksen, H., & Elstad, E. (2019). Norsk læreres erfaringer med karakterfritt semester: rom for forbedringsinnsats hos eleven og større fokus på læringsfremmende tilbakemelding fra læreren? *Nordisk tidsskrift for utdanning og praksis*, 13(2), 4–20. <https://doi.org/10.23865/up.v13.1968>
- *Flaten, T., & Vollan, M. (2018). Ukeprøva – en sentral, men utforska teksttype i skolen. *Sakprosa*, 10(2). <https://doi.org/10.5617/sakprosa.5968>
- Forskrift til opplæringslova. (2006). Forskrift til opplæringslova (FOR-2006-06-23-724). Kapittel 3, del III, § 3-15. <https://lovdata.no/forskrift/2006-06-23-724/§3-15>

⁷ Artikler som inngår i litteraturundersøkelsen er merkt med «*»

- Forskrift til friskolelova. (2006). Forskrift til friskolelova (FOR-2006-07-14-932). Kapittel 3, del III, § 3-15. <https://lovdata.no/forskrift/2006-07-14-932/§3-15>
- *Føinum, M. (2016). Når skolens behov blir viktigere enn formell kompetanse: Hvordan veilede nyutdannede lærere til å undervise i skolefag de ikke har? *Acta Didactica Norge*, 10(2), 290–310. <https://doi.org/10.5617/adno.2446>
- Gamlem, S.M. & K. Smith (2013) Student Perceptions of Classroom Feedback, *Assessment in Education, Principles, Policy & Practice*, 20(2), 150–169.
- Gardner, J. 2007. Is teaching a 'partial' profession? Making the Grade, Summer: 18–21.
- *Gillespie, A., & Burner, T. (2019). Principals' views on the implementation of grade-free middle schools in Norway: justifications, challenges and opportunities. *Nordisk tidsskrift for utdanning og praksis*, 13(2), 21–38. <https://doi.org/10.23865/up.v13.1969>
- Harlen, W. 2004. A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. Research Evidence in Education Library. London: EPPI-Centre, Social Science Research Unit, Institute of Education. <http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=61> (accessed March 20, 2010).
- *Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N., & Bernelius, V. (2016). Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data. *Education Inquiry*, 7(4), 463–479. <https://doi.org/10.3402/edui.v7.29413>
- *Harling, M. (2018). Valet, varat och varan. En studie av gymnasievalets logiker och praktiker. *Educare*, 12(1), 115–139. <https://doi.org/10.24834/educare.2018.1.6>
- Hopfenbeck, T. N., Flórez Petour, M. T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: large-scale implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy & Practice*. <http://doi.org/10.1080/0969594X.2014.99652>
- Hopfenbeck, T.N. (2016). *Å lykkes med elevvurdering*, Fagbokforlaget.
- *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 standpunkt vurdering i offentlige og private videregående skoler*. <http://hdl.handle.net/11250/280144>
- *Huhtala, A., Vesalainen, M., Hildén, R., & Rautopuro, J. (2019). Finländska svensklärarens undervisningspraktiker och deras samband med elevers inlärningsresultat. *Nordand*, 3(1), 4–24. <https://doi.org/10.18261/issn.2535-3381-2019-01-01>
- *Jönsson, A., & Klapp, A. (2020). Svenska lärares syn på avvikelser mellan resultat på nationella prov och ämnesbetyg. *Educare*, 14(4), 88–108. <https://doi.org/10.24834/educare.2020.4.4>
- *Jönsson, A., Balan, A., & Hartell, E. (2021). Analytic or holistic? A study about how to increase the agreement in teachers' grading. *Assessment in Education: Principles, Policy & Practice*, 1–16. <https://doi.org/10.1080/0969594X.2021.1884041>
- Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating Measures of Performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. <https://doi.org/10.1111/j.1745-3992.1999.tb00010.x>
- *Knekta, E. (2017). Are all Pupils Equally Motivated to do Their Best on all Tests? Differences in Reported Test-Taking Motivation within and between Tests with Different Stakes. *Scandinavian Journal of Educational Research*, 61(1), 95–111. <https://doi.org/10.1080/00313831.2015.1119723>

- *Lekholm, A. K. (2011). Effects of School Characteristics on Grades in Compulsory School. *Scandinavian Journal of Educational Research*, 55(6), 587–608. <https://doi.org/10.1080/00313831.2011.555923>
- *Löfgren, H., & Aman, R. (2020). A double pressure to perform? Pupils talk about grades and parents in a multicultural school setting in Sweden. *Education Inquiry*, 1–18. <https://doi.org/10.1080/20004508.2020.1831287>
- *Löfgren, R., Löfgren, H., & Lindberg, V. (2019). Pupils' perceptions of grades: a narrative analysis of stories about getting graded for the first time. *Assessment in Education: Principles, Policy & Practice*, 26(3), 259–277. <https://doi.org/10.1080/0969594X.2019.1593104>
- *Lutnæs, E. (2011). *Standpunktvurdering i grunnskolefaget Kunst og håndverk – læreres forhandlingsrepertoar*. Arkitektur- og designhøgskolen i Oslo.
- *Lyrén, P. (2008). Prediction of Academic Performance by Means of the Swedish Scholastic Assessment Test. *Scandinavian Journal of Educational Research*, 52(6), 565–581. <https://doi.org/10.1080/00313830802497158>
- *Lysne, A. (2006). Assessment Theory and Practice of Students' Outcomes in the Nordic Countries. *Scandinavian Journal of Educational Research*, 50(3), 327–359. <https://doi.org/10.1080/00313830600743365>
- *Månsson, N., Ydhag, C. C., & Osman, A. (2021). I skuggan av kulturellt kapital – om konsten att omforma habitus för skolframgång. *Nordic Studies in Education*, 41(2), 130–147. <https://doi.org/10.23865/nse.v41.2353>
- *Mullola, S., Hintsanen, M., Jokela, M., Lipsanen, J., Alatupa, S., Ravaja, N., & Keltikangas-Järvinen, L. (2014). Associations Between Teacher-Rated Versus Self-Rated Student Temperament and School Achievement. *Scandinavian Journal of Educational Research*, 58(2), 147–172. <https://doi.org/10.1080/00313831.2012.725094>
- *Olovsson, T. G. (2014). The assessment process in two different year-five classrooms in Sweden. *Education Inquiry*, 5(4), 561–581. <https://doi.org/10.3402/edui.v5.24619>
- *Olovsson, T. G. (2020). Ämnesövergripande undervisning i SO-ämnena i svensk grundskola - vad säger lärarna? *Acta Didactica Norden*, 14(3), 1–24. <https://doi.org/10.5617/adno.7920>
- *Olovsson, T. G., & Näsström, G. (2018). Ämnesövergripande undervisning och betyg i årskurs 4-6 i svensk grundskola i SO- och NO- ämnena. *Nordidactica – Journal of Humanities and Social Science Education*, 7(4), 88–117.
- *Oltedal, E., Gamlem, S. M., Kleivenes, O. M., Ryslett, K., & Vasset, T. (2016). Teachers' assessment experiences and perceptions in the practical-aesthetic subjects. *Scandinavian Journal of Educational Research*, 60(6), 649–662. <https://doi.org/10.1080/00313831.2015.1066431>
- *Prøitz, T. S., & Borgen, J. S. (2010). *Rettferdig standpunktvurdering – det (u)muliges kunst? Læreres setting av standpunkt karakter i fem fag i grunnopplæringen*. <http://hdl.handle.net/11250/279131>
- Sandvik, L.V. & Buland, T. (2014). *Vurdering i skolen. Utvikling av kompetanse og fellesskap. Sluttrapport fra prosjektet Forskning på individuell vurdering i skolen (FIVIS)*. NTNU Program for lærerutdanning i samarbeid med SINTEF Teknologi og samfunn.
- *Sivenbring, J. (2019). Making Sense and Use of Assessments. *Scandinavian Journal of Educational Research*, 63(5), 759–770. <https://doi.org/10.1080/00313831.2018.1434827>

- Stanley, G., R. MacCann, J. Gardner, L. Reynolds, and I. Wild. 2009. Review of teacher assessment: Evidence of what works best and issues for development. Oxford: Oxford University Centre for Educational Development; report commissioned by the QCA. <http://www.education.ox.ac.uk/assessment/publications.php> (accessed January 12, 2011).
- *Strandler, O. (2016). Equity Through Assessment? Teachers' Mediation of Outcome-Focused Reforms in Socioeconomically Different Schools. *Scandinavian Journal of Educational Research*, 60(5), 538–553. <https://doi.org/10.1080/00313831.2015.1062414>
- *Svensson, A. (2004). Gymnasiebetyg eller högskoleprov som urvalsinstrument? Fallet civilingenjörsutbildningarna. *Pedagogisk forskning i Sverige*, 9(1), 15–36.
- *Tanner, M. (2016). Att visa vad man kan. Elever som medspelare i förändrade policypraktiker om bedömning i år sex. *Kapet*, 12(1), 1–20. <http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akau%3Adiva-43705>
- *Tengberg, M., & Skar, G. B. (2016). Samstämmighet i lärares bedömning av nationella prov i läsförståelse. *Nordic Journal of Literacy Research*, 2(1), 1–18. <https://doi.org/10.17585/njlr.v2.230>
- *Tholin, J. (2010). Bilden av eleven i medias rapportering av betygsfrågor 1932–1990 och 2009. *Didaktisk Tidskrift*, 19(1), 37–53. <http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Ahgo%3Adiva-392>
- *Thorsen, C. (2014). Dimensions of Norm-Referenced Compulsory School Grades and their Relative Importance for the Prediction of Upper Secondary School Grades. *Scandinavian Journal of Educational Research*, 58(2), 127–146. <https://doi.org/10.1080/00313831.2012.705322>
- *Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation*, 18(2), 153–172. <https://doi.org/10.1080/13803611.2012.659929>
- Thronsdén, I., Hopfenbeck, T.N., Lie, S. & E.L. Dale (2009) *Bedre vurdering for læring. Rapport fra "Evaluering av modeller for kjennetegn på måloppnåelse i fag"*. Universitetet i Oslo.
- *Torkildsen, L. G., & Erickson, G. (2016). 'If they'd written more...' – On students' perceptions of assessment and assessment practices. *Education Inquiry*, 7(2), 137–157. <https://doi.org/10.3402/edui.v7.27416>
- Tveit, S., & Olsen, R. V. (2018). Eksamens mange roller i sertifisering, styring og støtte av læring og undervisning i norsk grunnopplæring. *Acta Didactica Norge*, 12(4), Art. 18, 27 sider. <https://doi.org/10.5617/adno.6381>
- Utdanningsdirektoratet (2019). *Erfaringer fra nasjonal satsing på vurdering for læring (2010–2018)*. Sluttrapport til KD.
- Utdanningsdirektoratet (2015). *Rapport om utviklingen av klager på standpunktkarakterer fra 2010 til 2015. Kvantitative og kvalitative data fra fylkesmannsembeter og fylkeskommuner*.
- *Wetterstrand, S., Sundhäll, M., & Lundahl, C. (2017). Att få sina bedömningar granskade: Lärares syn på de nationella provens syfte att främja likvärdig bedömning. *Utbildning & Demokrati*, 26(2), 75–94. <http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Aoru%3Adiva-62682%0A>

Appendix

Forfatter	År	trinn	Fag	N	Lærere	Elever	Rektorer	Karakterer	Dokument
Alm, F., & Colnerud, G.	2015	b.trinn, u.trinn, vgs,	faguavhengig	355					
Andersson, S., & Lunneblad, J.	2020	u.trinn	faguavhengig			14			
Arensmeier, C.	2020	u.trinn	faguavhengig					2.565,000	
Birkelund, I., & Midthaugen, P.	2019	u.trinn, vgs	kroppsoving	8					
Blomqvist, P.	2018	vgs	Morsmål	16					
Brante, E. W., & Lund, E. S.	2017	vgs	morsmål, samfunnsfag	18					
Cliffordson, C.	2006	vgs	faguavhengig					2.954	
Cliffordson, C., & Askling, B.	2008	vgs	faguavhengig					164.106	
Eriksen, H., & Elstad, E.	2019	vgs	morsmål	8					
Flaten, T., & Vollan, M.	2018	b.trinn	faguavhengig						42
Føinum, M.	2016	u.trinn	faguavhengig	3					
Gillespie, A., & Burner, T.	2019	u.trinn	faguavhengig				4		
Harju-Luukkainen, H., Vetteranta, J., Ouakrim-Soivio, N., & Bernelius, V.	2016	u.trinn	morsmål					6.415	
Harling, M.	2018	u.trinn	faguavhengig			54			
Hovdhaugen, E., Seland, I., Lødding, B., Prøitz, T. S., & Vibe, N.	2014	vgs	morsmål, matematikk	32		31	8	172.852	
Huhtala, A., Vesalainen, M., Hildén, R., & Rautopuro, J.	2019	u.trinn	morsmål	98					
Jönsson, A., & Klapp, A.	2021	b.trinn	engelsk, matematikk	74					
Jönsson, A., Balan, A., & Hartell, E.	2020	u.trinn	fysikk, engelsk	21					
Knekta, E.	2017	u.trinn	naturfag,			375			
Lekholm, A. K.	2011	u.trinn	faguavhengig					99.07	
Löfgren, H., & Aman, R.	2020	b.trinn	naturfag, samfunnsfag			19			
Löfgren, R., Löfgren, H., & Lindberg, V.	2019	b.trinn	faguavhengig			127			
Lutnæs, E.	2011	u.trinn	kunst og håndverk	7					
Lyrén, P.	2008	vgs	faguavhengig						10
Lysne, A.	2006	t.uavgengig	faguavhengig						
Månsson, N., Ydhag, C. C., & Osman, A.	2021	vgs	faguavhengig			23			
Mullola, S., Hintsanen, M., Jokela, M., Lipsanen, J., Alatupa, S., Ravaja, N., & Keltikangas-Järvinen, L.	2014	u.trinn	morsmål, matematikk	72		1.063			
Olovsson, T. G., & Näsström, G.	2018	b.trinn	naturfag, samfunnsfag	113					
Olovsson, T. G.	2020	b.trinn	samfunnsfag	6					
Olovsson, T. G.	2014	b.trinn	faguavhengig	4		18			

Forfatter	År	trinn	Fag	N				
				Lærere	Elever	Rekto- rer	Karak- terer	Doku- ment
Oltedal, E., Gamlem, S. M., Kleivenes, O. M., Ryslett, K., & Vasset, T.	2016	u.trinn	praktisk-ests- tiske fag	28				
Prøitz, T. S., & Borgen, J. S.	2010	u.trinn, vgs,	faguavhengig	41		23		
Sivenbring, J.	2019	u.trinn	faguavhengig		28			
Strandler, O.	2016	b.trinn	faguavhengig	6				
Svensson, A.	2004	vgs	faguavhengig				35.495	
Tanner, M.	2016	b.trinn	faguavhengig		12			
Tengberg, M., & Skar, G. B.	2016	u.trinn	morsmål	6				
Tholin, J.	2010	t.uav- gengig	faguavhengig					253
Thorsen, C.	2014	u.trinn	faguavhengig				3.855	
Thorsen, C., & Cliffordson, C.	2012	vgs	faguavhengig				308.053	
Torkildsen, L. G., & Erickson, G.	2016	b.trinn, u.trinn	faguavhengig		54			
Wetterstrand, S., Sundhäll, M., & Lundahl, C.	2017	b.trinn, u.trinn	matematikk	17				

NB. b.trinn = barneskolen, u.trinn = ungdomstrinn, vgs = videregående skole, t.uavgengig = trinnuavhengig.